ISSN 1756-7904

INTERNATIONAL DEVELOPMENT
University of East Anglia

Working Paper Series

The Microfinance of Reproduction and the Reproduction of Microfinance: Understanding the Connections between Microfinance, Empowerment, Contraception and Fertility in Bangladesh in the 1990s

Maren Duvendack and Richard Palmer-Jones

2012

University of East Anglia

Working Paper 40

The School of International Development, University of East Anglia
Norwich, NR4 7TJ, United Kingdom

DEV Working Paper 40

The Microfinance of Reproduction and the Reproduction of Microfinance: Understanding the Connections between Microfinance, Empowerment, Contraception and Fertility in Bangladesh in the 1990s.

Maren Duvendack and Richard Palmer-Jones

First published by the School of International Development in July 2012.

This publication may be cited as:
Duvendack, M. & Palmer-Jones, R., 2012, The Microfinance of Reproduction and the Reproduction of Microfinance: Understanding the Connections between Microfinance, Empowerment, Contraception and Fertility in Bangladesh in the 1990s, Working Paper 40, DEV Working Paper Series, The School of International Development, University of East Anglia, UK.

About the Authors
Maren Duvendack is a Research Fellow at the Overseas Development Institute, London, UK and an Honorary Research Fellow in the School of International Development at the University of East Anglia, Norwich, UK.

Richard Palmer-Jones is a Senior Research Fellow in the School of International Development at the University of East Anglia, Norwich, UK.

Contact:
Email m.duvendack@odi.org.uk
School of International Development
University of East Anglia
Norwich, NR4 7TJ
United Kingdom
Tel: +44(0)1603 593383
Fax: +44(0)1603 451999

ISSN 1756-7904

About the DEV Working Paper Series

The Working Paper Series profiles research and policy work conducted by the School of International Development and International Development UEA (see below). Launched in 2007, it provides an opportunity for staff, associated researchers and fellows to disseminate original research and policy advice on a wide range of subjects. All papers are peer reviewed within the School.

About the School of International Development

The School of International Development (DEV) applies economic, social and natural science disciplines to the study of international development, with special emphasis on social and environmental change and poverty alleviation. DEV has a strong commitment to an interdisciplinary research and teaching approach to Development Studies and the study of poverty.

International Development UEA (formerly Overseas Development Group)

Founded in 1967, International Development UEA is a charitable company wholly owned by the University of East Anglia, which handles the consultancy, research, and training undertaken by the faculty members in DEV and approximately 200 external consultants. Since its foundation it has provided training for professionals from more than 70 countries and completed over 1,000 consultancy and research assignments. International Development UEA provides DEV staff with opportunities to participate in on-going development work, practical and policy related engagement which add a unique and valuable element to the School's teaching programmes.

For further information on DEV and the International Development UEA, please contact:

School of International Development
University of East Anglia, Norwich NR4 7TJ, United Kingdom
Tel: +44 (0)1603 592329
Fax: +44 (0)1603 451999
Email: dev.general@uea.ac.uk
Webpage: www.uea.ac.uk/dev

Abstract

Microfinance (MF) and family planning (FP) are important interventions in the promotion of human development. Several authors (e.g. Amin, Hill and Li, 1995; Schuler, Hashemi and Riley, 1997) using naive methods argue that MF in Bangladesh increases contraceptive use and reduces fertility, largely because MF empowers women. Pitt et al (1999 – henceforth PKML), however, using instrumental variables (IV) estimation find that MF is associated with decreases in contraceptive use especially when females borrow, but male borrowing decreases fertility, perhaps because fertility increasing income effects of MF are outweighed by substitution effects. In this paper we apply matching methods to our reconstruction of the PKML data to test whether these other methods reproduce their results. In addition we build on the analysis of PKML with panel data to examine the long-term effects of MF on contraceptive use and fertility. We find that female borrowing robustly increases contraceptive use but has mainly no effects on fertility, while male borrowing has no effect on contraceptive use or on fertility. Our results are vulnerable to unobservables, but there is no reason to believe that IV based methods are more reliable. Together, these results disagree with some of PKML's headline findings.

Introduction

Microfinance (MF) and family planning (FP) are important interventions in the promotion of human development and in the fight against poverty (Daley-Harris, 2002; Littlefield, Morduch and Hashemi, 2003; UNCDF, 2005; Cleland et al, 2006; Cleland, 2009). MF is not just about credit; it encompasses other financial services (Armendáriz de Aghion and Morduch, 2005), and it is now often combined with other interventions, including, for example, information and advice about contraception and fertility (Leatherman et al, 2011).

It is often argued that access to credit affects FP by increasing the value of time (Desai and Tarozzi, 2011; Pitt et al, 1999 – henceforth PKML; Buttenheim, 2006). However, it is unclear whether this has positive or negative effects on fertility because while making reproduction more costly, any such substitution effect may be offset by an income effect associated with a concomitant rise in income if children are normal goods (PKML, p. 2). In other words, the direction of the impact of MF on fertility is unclear (Desai and Tarozzi, 2011) and few studies (discussed below) have tried to test these links between MF and FP outcomes.

Literature review

MF may have beneficent impacts on a range of socio-economic outcomes but the empirical evidence so far is mixed and unconvincing. There have been four unsystematic reviews of microfinance impact (Sebstad and Chen, 1996; Gaile and Foster, 1996; Goldberg, 2005; Odell, 2010) indicating that, although anecdotes and other inspiring stories (Todd, 1996) show that microfinance can make a real difference in the lives of those served, rigorous quantitative evidence on the nature, magnitude and balance of microfinance impact is still scarce and inconclusive (Armendáriz de Aghion and Morduch, 2005 and 2010). This is corroborated by two recent systematic reviews on the impact of MF (Stewart et al, 2011; Duvendack et al, 2011) which argue that most MF impact evaluations suffer from weak methodologies which fail to adequately control for self-selection and non-random programme placement bias[1] (particularly argued by Duvendack et al, 2011), adversely affecting the reliability of impact estimates; this in turn may have contributed to misconceptions of the actual effects of MF programmes (Roy, 2010; Bateman, 2010; Dichter and Harper, 2007).

Few studies have investigated the causal link between microfinance, contraceptive use, and fertility; until recently the ones that do focus on the case of Bangladesh (Buttenheim, 2006), where, it has been suggested that MF increases contraceptive use and reduces fertility at the individual level, putatively because of the effects MF lent to women has on empowering them (Amin, Hill and Li, 1995; Amin et al, 1994 and 2001; Schuler, Hashemi and Riley, 1997; Hashemi, Schuler and Riley, 1996; Schuler and Hashemi, 1994). It is assumed that women prefer contraceptive use and fewer children than men in this patriarchal society. PKML, however, find that MF is not associated with an increase in contraceptive use or decrease in fertility, in particular for female participants in MF (PKML, p. 1). PKML use a complex two-stage instrumental variables (IV) estimation, arguing that other studies, such as those referred to above, do not control for self-selection and programme placement biases; PKML differentiate by gender of borrower, finding significant negative effects on contraceptive use and mainly no effects on fertility when females borrow and no effects on contraceptive use and significantly negative effects on fertility from male borrowing, strikingly contrary to the usual expectations.

Steele et al (2001), using panel data from Bangladesh from a pipeline research design (Coleman, 1999) produced around the same time as those analysed by PKML, employ fixed and random effects panel models to control for self-selection and programme placement bias. Steele et al (2001, p. 280) conclude that MF has a positive impact on contraceptive use; they rationalise their results by arguing that the membership of a MF group, which is the (dichotomous) variable they use, is more appropriate than the amount borrowed, the variable used by PKML, to capture the empowering effect of MF. In their data, Steele et al (2001) have cases of women who are members of the MF group but have not borrowed; such women may be

[1] MF participants commonly self-select into microfinance, i.e. the assignment process in non-random, and thus they differ from non-participants in observable and unobservable characteristics. The locations of programmes are also chosen in a non-random way and therefore differ from other places that could be used as controls (Coleman, 1999; Pitt and Khandker, 1998).

empowered by group meetings and solidarity. The amount borrowed may only proxy income changes and miss these wider effects[2].

Buttenheim (2006) supports the view of Steele et al (2001), that membership (or participation) is the more appropriate indicator, but extends this, arguing that the level of MF participation in the community or the availability of the MF programme at the community level is the more appropriate measure to assess the impact of MF on contraceptive use, especially when network and spill-over effects on the local community are present (Buttenheim, 2006, p. 10). Moreover, the microfinance institutions (MFIs) in the two Bangladesh data sets are different[3]; in Steele et al (2001) women (only) are members of groups facilitated by Save the Children USA and the Bangladeshi non-governmental organisation (NGO) ASA (Rutherford, 2009) while in the PKML data males and females can be members of the three NGOs represented (Grameen Bank (GB), the Bangladesh Rural Advancement Committee (BRAC), and the Bangladesh Rural Development Board (BRDB)). Save the Children USA had quite intensive interactions of a putatively empowering nature with their members, while ASA were largely focused on microcredit alone, with likely different implications for female empowerment[4]. The PKML NGOs[5] espoused rather different interactions with group members, although in each case some might be considered empowering (c.f. the 16 GB affirmations). Nevertheless, they are unlikely to have had such powerful empowering effects as Save the Children USA. Moreover, both indicators (i.e. membership and amount borrowed) are only indirect evidence of empowerment and income respectively[6].

Desai and Tarozzi (2011) discuss this literature and report a randomised control trial (RCT) conducted in Ethiopia, with data from before and after the intervention with

---

[2] We investigated membership as an indicator of the effects of MF participation using the (our reconstruction of the variables in the) PKML data. However, whether we use MF membership or amount borrowed as a treatment variable did not make much difference in our case, perhaps because all MF members are also borrowers in this data set. In the case of Steele et al (2001) there is a slight discrepancy in this regard, they report more members than borrowers, i.e. women can be members of a credit group regardless of whether they currently borrow or not (Steele et al, 2001, p. 268).

[3] The sample of the World Bank data used in PKML is drawn from 87 villages from 29 thanas across rural Bangladesh while the Save the Children USA/ASA data used by Steele et al (2001) comes from 15 villages from Nasirnagar thana in Brahmanbaria in Eastern Bangladesh.

[4] However, the number of Save the Children USA women in the sample was relatively small ('the estimates for contrasts of SC-ASA membership or non-membership with SC membership should be treated with caution' (Steele et al, 2001, p. 273)).

[5] As well as borrowers from other sources, which are neglected in PKML. Some MF borrowers also borrow from these other sources (Duvendack and Palmer-Jones, 2011b).

[6] Pitt et al (2006) analyse more direct indicators of empowerment, but we do not discuss these data here due to lack of space. Their conclusion fits better the orthodoxy, that MF empowers women, using an extended and perhaps tendentious argument to reconcile the PKML findings with regard to contraception and fertility with their later findings on female empowerment.

three different treatment groups and a control group[7]. The authors find that none of the interventions - alone or in combination - had any impact on increased contraceptive use compared to the control group. However, Leatherman et al (2011), argue that the control group could have been contaminated by spill-over effects from the treatment groups or by the availability of other microcredit or family planning services. In addition, this was a panel of villages rather than households, which differed between panel waves.

Hence, attributing the changes in contraceptive use and fertility to impacts of MF is a complex and challenging task, since many social, economic and cultural factors are likely to influence FP decisions[8] (Livi-Bacci and de Santis, 1998). In this paper, we seek to assess the robustness of the results found by PKML using another estimation method - propensity score matching (PSM) – both because establishing causality with the data used by PKML has been contested (Roodman and Morduch, 2009 – henceforth RnM), and to explore the contrast with Steele et al (2001). PSM may have advantages over random coefficients IV methods produce, which rely on largely untestable assumptions and model dependence[9], by balancing the covariates in the samples of treatment and control groups (Rosenbaum, 2002; DiPrete and Gangl, 2004, p. 276). In addition, we use a second wave of the PKML data allowing panel and differences-in-differences (DID) analyses of the longer term effects of MF on contraceptive use and fertility[10].

---

[7] RCTs are sometimes taken as the 'gold standard' for impact evaluation (Banerjee and Duflo, 2011; Karlan and Appel, 2011); this is contested (Deaton, 2010; Ravallion, 2011; Duvendack et al, 2011).

[8] Thus, the relationship between MF, contraceptive use and fertility is unclear, but of continuing importance (Buttenheim, 2006), warranting further exploration of these issues. In addition to the empirical contradictions, there are potential conflicts within households with regard to FP decisions. Commonly it is believed that men prefer more children and thus might discourage their wives from using contraception, and women often have to hide contraceptives from their husbands (Ashraf, Field and Lee, 2010). Angeles, Guilkey and Mroz (2005) and Gertler and Molyneaux (1994) argue that improved education as well as the development of better economic opportunities increase contraceptive use and decrease fertility. Buttenheim (2006) is more critical of the idea of links between education and contraceptive use. She finds that older women are more likely to use contraceptives, as well as women living in urban areas. The desire to have children also appears to be driven by economic factors. For example, in Buttenheim's (2006) sample (from Indonesia) the desire to have more children in 2000 is higher than in 1993 and 1997, possibly due to Indonesia's slow recovery from the economic crisis in 1998 (Buttenheim, 2006, p. 15).

[9] The assumption that the estimation model captures entirely the effects of all potentially confounding variables (e.g. DiPrete and Gangl, 2004, p. 275).

[10] Khandker (2005) has used the panel version of these data to analyse other effects.

The PKML dataset and estimation strategy is largely the same as that used by Pitt and Khandker (1998 – henceforth PnK). RnM replicated the key PnK studies[11], using different software, and come to the same results as PnK, but conclude that:

> 'decisive statistical evidence in favor of [the idea that microcredit alleviates poverty, smoothes household expenditure and lessens the pinch of hunger especially when women are involved in borrowing] is absent from these studies' (RnM, p. 40)[12].

Duvendack (2010) and Duvendack and Palmer-Jones (2011b - henceforth DPJ) using PSM and sensitivity analysis conclude that the very modest and mixed impacts of MF on the outcome variables used in PnK, are highly vulnerable to confounding by unobservables – such as entrepreneurial ability, and so on. These differences in inference suggest that it is important to replicate the results of the more recent papers by Pitt and co-authors (1999, 2003, and 2006 (which uses the 1998/99 follow-up data)), which use broadly the same data and estimation methods[13]. In this paper we restrict ourselves to replication[14] of the study by PKML on contraceptive use and fertility, a process which is finding increasing support in economics[15].

Thus, the objective of this paper is to re-investigate the findings of PKML who use data first presented in PnK. We follow the approach by DPJ and apply PSM and sensitivity analysis to the data to triangulate these findings and analyse the data as a panel using a random effects model as well as PSM along with DID, to obtain more refined impact estimates.

[11] RnM do not replicate Chemin (2008) or a few other studies that used the PnK data (Khandker, 1996, 2000; Pitt et al, 1999; Pitt, 2000; McKernan, 2002; Pitt and Khandker, 2002; Pitt et al, 2003; Menon, 2006; Pitt, Khandker and Cartwright, 2006).

[12] See http://blogs.cgdev.org/open_book/ where Roodman asserts that PnK methods do not establish causality.

[13] RnM and DPJ also replicate Khandker (2005) who uses the 1998/99 data, but also find replication unsatisfactory, and cannot fully support the claims of either PnK or Khandker (2005).

[14] Replication and reproduction are an important part of scientific practice, especially when there are contradictory or controversial findings, without which results cannot be taken as robust (Hamermesh, 2007; Dewald et al, 1986; McCullough et al, 2006; McCullough et al, 2008). While used in various ways in this literature (McCullough et al, 2006) replication covers checking of the original study (strict or pure replication – Collins (1991)), application of different statistical methods to the same data set, or application of the same or different methods to a different data set which is arguably equivalent to the original study (reproduction); and extension of these methods to other data (scientific replication). In this paper we use the term replication for both checking and reproduction.

[15] The American Economic Review (AER), for example, requires its authors to make their data sets and code available which are then uploaded onto a website maintained by the AER especially for this purpose (see Hamermesh, 2007, p. 717; Burman et al, 2010).

The impact of microfinance in Bangladesh: the case of PnK

PnK use data from a World Bank funded survey in three waves in 1991-1992[16] on three leading microfinance group-lending programmes in Bangladesh: GB, BRAC and BRDB (PnK, p. 959). A quasi-experimental design was used which sampled target (having a choice to participate/being eligible) and non-target households (having no choice to participate/not being eligible) from villages with microfinance programme (treatment villages) and non-programme villages (control villages).

The survey was conducted in 87 villages from 29 thanas[17]; the treatment villages were randomly selected from a list of villages provided by the MFIs' local offices and the control villages were randomly selected from the governments' village census; 1,798 households were selected. Within the treatment villages eligibility criteria are supposedly imposed on membership of the NGOs (see below). 1,538 of the sampled households were labelled target households, putatively cultivating less than 0.5 acres at the time of joining the MFI[18], and 260 were non-target households (PnK, p. 974). Of the 1,538 households, 905 (59%) effectively participated in microfinance. The three survey waves (henceforth R1-3) were timed to account for seasonal variations, (Pitt, 2000, p. 28-29) [19]. PnK find that microcredit has significant positive impacts on many indicators of well-being and find larger positive impacts for women borrowers. For example,

> 'annual household consumption expenditure, […], increased 18 taka for every 100 additional taka borrowed by women from these credit programs [*GB, BRAC, BRDB*], compared with 11 taka for men' (PnK, p. 988).

PnK adopt an estimation strategy for assessing the impact of microfinance participation involving comparisons of 'treated' and 'non-treated' households in 'treated' villages, and 'non-treated' households in 'non-treated' (control) villages. Treatment refers to participating in the loan programme of one of the selected MFIs; at the household level this varies according to the gender of the borrower, and at the village level according to the presence of the MFI in the village. However, comparing households in treatment and control villages is not sufficient for obtaining impact estimates because the villages differ (there is programme placement bias[20]) and households commonly select into microfinance. In this type of group-based lending

[16] In areas not affected by the cyclone of April 1991.

[17] A thana (literally police station, also known as upazila) is a unit of administration in Bangladesh; in 1985 there were 495 upazilas (Bangladesh Bureau of Statistics, 1985) and 507 upazilas in 2001 (Bangladesh Bureau of Statistics, 2004).

[18] See below for discussion of the fuzzy nature of the eligibility criterion applied in practice.

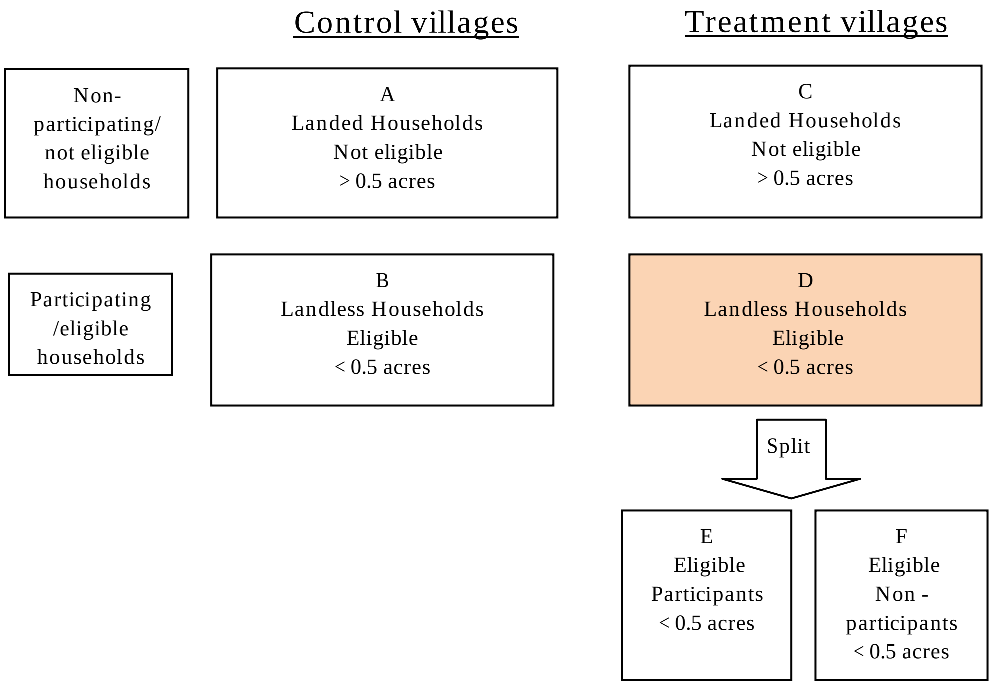
[19] A follow-up data set was collected in 1998-1999 re-surveying the same households that were already interviewed in R1-3; we discuss and use these data below.

[20] The assumption was that MFIs choose more remote and backward villages (PnK; Coleman, 1999). Hence, microfinance impact may vary according to village type.

individuals select themselves, can be selected (or excluded) by their peers and/or by microfinance loan officers, giving rise to selection bias.

In principle all the MFIs operate the eligibility criterion that participating households should be cultivating[21] less than 0.5 acres of land at the time of recruitment into the MFI programme. PnK's (ideal) identification strategy can be understood graphically by looking at Figure 1.

Figure 1: Intended identification strategy



Source: Authors illustration based on Morduch (1998) and Chemin (2008).
Notes: This diagram ignores that the eligibility criterion was not strictly (literally) enforced. Thus the actual strategy used (de facto) participation.

PnK suggest that their estimation strategy is comparing outcomes across the discontinuity between participant (eligible) and non-participant (not eligible) households in treatment and control villages; that is, at the boundary between group B and A in control villages, and between group D to C in treatment villages (Figure 1). The difference between these two sets of comparisons is estimated by applying village-level fixed-effects to account for programme placement bias.

The application of an eligibility criterion as an identification strategy is plausible provided it is strictly enforced. However, as Morduch (1998) points out, mistargeting

[21] There is some confusion about whether the eligibility criterion is cultivated (operated) or owned land, and whether this includes homestead land.

occurred[22] (see also Ravallion, 2008, p. 3818; Chemin, 2008, p. 465). Group D contains participants who own considerably more than 0.5 acres of land. Pitt rationalises this by claiming that the value of land of treated households which cultivate/possess more than 0.5 acres is so low that the value of the land of these households is effectively less than the median value of 0.5 acres of average land. However, in control villages (groups A and B) households were categorised as eligible based on the less than 0.5 acres of cultivated land alone[23]. Pitt (1999, 2011a and b), claims that discarding the households whose membership is contested does not affect the results. In an attempt to check PnK RnM were eventually able to replicate the original PnK data, if not exactly[24], and results, as independently did DPJ, but come to different conclusions with regard to the claim of causality.

Chemin (2008) using PSM applied to his construction of the same data came to different conclusions as to the impacts of MF. DPJ could not replicate Chemin's (2008) data closely, or findings, but also come to conclusions different from PnK, adding that their results remain highly vulnerable to unobservables. DPJ though doubt the ability of the PnK data to provide convincing evidence of impact attributable to MFIs.

There are further concerns about PnK's study and their substantive results. In brief, most microfinance impact evaluations are designed on the assumption that other formal and informal credit organisations are absent and would not have entered the financial markets in the absence of MFIs. However, this is not what the data show (or found in other studies conducted around the time of or soon after the PnK survey (Fernando, 1997; Jain and Mansuri, 2003; Zeller et al, 2001)). Households in the PnK data obtain loans not only from MFIs but also from other formal and informal sources and those with different portfolios will have different observable and unobservable characteristics. Thus, a comparison of (eligible) participants with (eligible) non-participants will include among the participants those who also borrow from other sources, and similarly among the control group(s); these groups will be quite heterogeneous, as will any impacts of microfinance borrowing. Comparison among these different sub-groups is constrained by sample sizes in PnK's data set. In this paper we include variables for different sub-groups, but

[22] Pitt (1999) refuted Morduch's (1998) claims and provided evidence supporting PnK's earlier findings. This debate was revisited by RnM and DPJ and taken up by Pitt (2011a and 2011b). It is not central to this paper to elaborate on this debate; instead the interested reader is referred to RnM and DPJ.

[23] This issue is addressed in more depth in DPJ.

[24] Apparently the data sets and code used for PnK were archived on CD-ROMs which are no longer readable (correspondence from Pitt to Roodman on February 28, 2008). Others who have used these data using similar procedures to PnK cannot supply their data or code (see personal communication with McKernan on April 16, 2009). Hence, it remains moot as to whether the differences between PnK and RnM are due to (1) differences in the raw data used; (2) differences in variable construction; or, (3) differences in the statistical estimations. (1) and (2) cannot be assessed, but those with the appropriate skills can assess RnM.

further exploration of this issue is beyond the scope of this paper (see DPJ).

Estimation strategy

The standard approach to solving the evaluation problem with observational data is to use an IV approach which claims to control for selection on observables as well as unobservables (Heckman and Vytlacil, 2007; Basu et al, 2007). The main goal of the IV method is to identify an instrument(s), that influences the decision to participate in a programme but at the same time does not have an effect on the outcome except through its influence on participation. Adequate instruments are required for IV to be an effective strategy (Morgan and Winship, 2007). However, in many cases weak instruments are employed which can have adverse effects on the accuracy of IV estimates (as argued by PKML and Steele et al, 2001). These drawbacks of the IV method suggest replication and reproduction using a different approach to estimating causal effects, in this case PSM. PSM is a method that has found wide use in a variety of disciplines, increasingly in economics. PSM attempts to mimic the methods of randomised experiment by matching treated cases to untreated cases according to a propensity score for participation estimated from a logit or probit estimation of participation (Rosenbaum and Rubin, 1983 and 1984; Caliendo and Kopeinig, 2005 and 2008; Ravallion, 2001). In ideal circumstances PSM controls for observable differences between treatment and control groups, but is vulnerable to unobservable differences (Smith and Todd, 2005; Becker and Caliendo, 2007). The potential impact of unobservables (‘hidden bias’, Rosenbaum, 2002) can be assessed using sensitivity analysis (Rosenbaum, op. cit.; Nannicini, 2007).

First we replicate the variable constructions of PKML[25] (see DPJ for further details) and then apply PSM using MF membership to explain contraceptive use and fertility. For PSM, we first estimate the likelihood of microfinance participation to match control to treatment cases using the propensity score, and then compute the treatment effects for the various comparison groups. Our first logit model specification (Table 1, column 2) follows the model set out by PKML because we are replicating PKML in this paper. The second model (Table 1, column 3) is a variation of PKML’s specification and forms the basis for the PSM analysis presented below[26].

[25] Most of the data, including questionnaires and variable codes are (at the time of writing this paper) available on the World Bank website (http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/0,,contentMDK:21470820~pagePK:64214825~piPK:64214943~theSitePK:469382,00.html) but replication remains a challenge – see RnM and DPJ.

[26] The logit specification can have important effects on the matches and on estimated impacts. We do not go into the implications this has in this paper because our aim is to assess the robustness of the PKML results, and due to constraints of space.

The models can be expressed as follows:

$$(1) \qquad \text{Logit}\,(y_{ij}) = \alpha + \beta C_{ij} + \gamma G_{ij} + \delta Z_{ij}$$

<u>Where:</u>

y_{ij} = participating household

C_{ij} = vector of individual-specific variables

G_{ij} = vector of household-specific variables

Z_{ij} = village-level fixed-effects

The dependent variable (y_{ij}) in the model presented in equation (1) represents participants (*i*) in village (*j*), taking a value of 1 for participants and 0 for others. C_{ij} is a vector of individual-specific variables such as age and marital status, and G_{ij} is a vector of household-specific variables representing variables such as education and wealth. Z_{ij} is a vector of village level variables.

Results

We are able to reproduce to a fair degree of accuracy the main descriptive statistics of PKML (see Appendix 1); where our figures differ from PKML we prefer ours because they triangulate almost exactly with RnM. Remaining differences in the variables are due to differences in interpretation of the variables rather than differences in data manipulations.

Table 1: Logistic regression model for MF participation using PKML's model specification and a variation thereof

	Logit specifications	
Independent variables	**PKML**	**Authors**
Age (years)	0.066*** *0.000*	0.062*** *0.000*
Age household head (years)	-0.031*** *0.000*	-0.031*** *0.000*
Highest education any male household member	-0.080** *0.019*	-0.067** *0.045*
Sex household head	1.048* *0.056*	1.102** *0.044*
Household land (decimals)	-0.002*** *0.003*	-0.002*** *0.004*
Landholdings household head spouse parents	-0.193** *0.042*	-0.213** *0.021*
Price of mustard oil	-0.063*** *0.000*	-0.040*** *0.008*
Price of milk	0.034 *0.348*	0.085*** *0.008*
Price of potato	0.190** *0.011*	0.145** *0.027*
Average female wage	-0.003 *0.793*	-0.018* *0.063*
Average male wage	-0.004 *0.740*	-0.027** *0.019*
Number of observations	1787	1787
Pseudo R-squared	0.112	0.084

Source: Authors calculations.

Notes: p-values in italics. * significant at 10%, ** significant at 5%, *** significant at 1%. The following control variables are used in PKML: maximum education household head, highest education any female household member, landholdings household head parents, landholdings household head brother, landholdings household head sister, landholdings household head spouse brother, landholdings household head spouse sister, no spouse in household, non-target households, access to primary school, access to rural health care, access to family planning, access to midwife, price of rice, price of wheat flour, price of hen egg, dummy for female wage, distance to bank. Some of those variables were dropped in the authors' logit specification since they were not collected in the follow-up round in 1998/99. The following control variables are used in the authors' logit specification: maximum education household head, highest education any female household member, landholdings household head parents, landholdings household head brother, landholdings household head sister, landholdings household head spouse brother, landholdings household head spouse sister, no spouse in household, access to primary school, price of rice, price of wheat flour, price of hen egg, all insignificant. Descriptive statistics for all logit variables can be found in Appendix 1.

In the authors' logit specification (Table 1, column 3) age of respondent, age of household head, household land, price of mustard oil and price of milk are statistically significant at 1%. Highest education of any male household member, sex of household head, landholdings of household head's spouse parents, price of potatoes and average male wage are significant at 5% and average female wage is significant at 10%. These findings are largely supported by PKML's logit specification. However, the pseudo R-squared in the authors' model is rather low at 0.084. A low pseudo R-squared will have implications for the quality of the matches and thus the robustness of the impact estimates, and consequently may have implications for the conclusions we draw.

Table 2: PSM and covariate balancing

Independent variables	Sample	Mean		Bias (%)	% Reduction in |Bias|	t-test
		Treated	Control			p>|t|
Age (years)	Unmatched	32.082	28.985	35.7		0.000
	Matched	32.082	31.876	2.4	93.4	0.723
Age household	Unmatched	41.34	42.299	-8.3		0.144
	Matched	41.34	41.222	1.0	87.7	0.867
Highest education male household	Unmatched	2.456	3.626	-31.2		0.000
	Matched	2.456	2.473	-0.5	98.5	0.938
Sex household head	Unmatched	1.019	1.011	6.5		0.197
	Matched	1.019	1.017	2.1	67.8	0.768
Household land	Unmatched	47.194	124.46	-25.4		0.000
	Matched	47.194	51.394	-1.4	94.6	0.676
Landholdings household head	Unmatched	0.417	0.607	-24.5		0.000
	Matched	0.417	0.435	-2.3	90.6	0.710
Price of mustard oil	Unmatched	52.99	53.893	-20.8		0.000
	Matched	52.99	53.303	-7.2	65.3	0.278
Price of milk	Unmatched	12.451	12.223	8.9		0.089
	Matched	12.451	12.362	3.5	61.0	0.604
Price of potato	Unmatched	6.959	6.935	2.5		0.636
	Matched	6.959	6.957	0.2	93.1	0.979
Average female wage	Unmatched	17.631	18.139	-7.7		0.159
	Matched	17.631	17.66	-0.4	94.4	0.947
Average male wage	Unmatched	35.987	36.944	-13.9		0.010
	Matched	35.987	36.166	-2.6	81.3	0.690

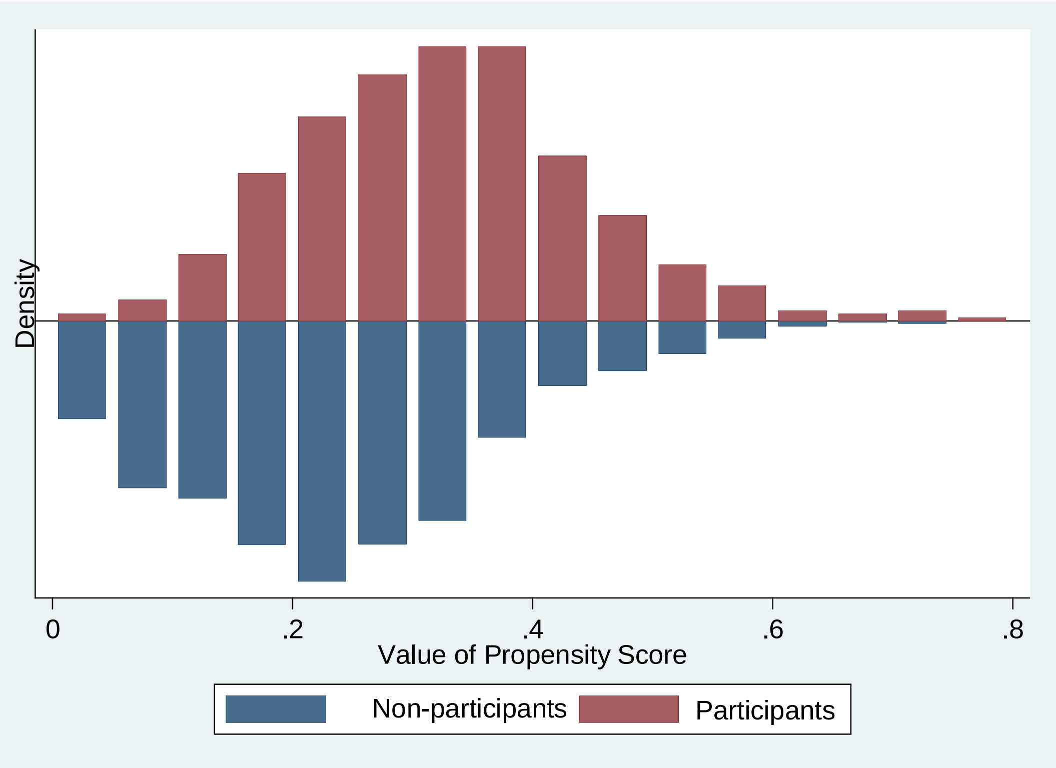
Source: Authors calculations.

The matching process (with replacement) leads to a balancing[27] of the independent variables between the treatment and control samples by restricting the control sample to increase its similarity to the treatment sample. Table 2 presents the results of covariate balancing together with the mean values for treated and controls before and after matching. There are clear differences in the mean values among treated and controls before and after matching, and the results in Table 2 indicate a reduction of bias for most variables that were significant in the logit model outlined in Table 1, in some cases reducing bias by more than 90%.

[27] Balancing in this context means having an acceptable (small) difference between the mean (or other statistic) of the covariates of the treated and untreated sample (DiPrete and Gangl, 2004).

Figure 2 displays the propensity scores of women members, currently married 14-50 year old, and the matched control sample including non-MFI eligible women[28] from both treatment and control villages. This shows considerable common support, although the central tendencies of the two groups is quite different, suggesting that the matching is not entirely successful[29].

Figure 2: Distribution of propensity scores for participants and eligible non-participants across treatment and control villages



Source: Authors calculations.

To get an estimate of the average treatment effect on the treated (ATT), presented in Table 3 and Table 4, we simply take the mean difference of the matched samples.

---

[28] Some households have both male and female borrowers while others have either a male or a female MFI borrower, or none. In principle most NGOs had rules that prohibited more than one NGO MFI member per household, but as with the land eligibility criterion this was imposed with some fuzziness. As noted above, some households, including some who borrow from MFIs, borrow from other formal or informal sources. We found no cases in the data of individuals, or households, borrowing from more than one MFI, although other quantitative data (Zeller et al, 2001) and qualitative studies Fernando (1997) report this to have been common around the same region and time.

[29] We intend to pursue this idea at a later date using "coarse exact matching" (King et al, 2011) which is thought to have considerable advantages over PSM, although at the expense of discarding a greater number of treatment cases that cannot be matched (Blackwell et al, 2009).

Table 3 lists the impact estimates for microcredit participation for all participants (male and female) and Table 4 provides impact estimates for female and male[30] participants separately. We apply nearest neighbour and kernel matching algorithms[31] on the outcome variables as defined by PKML.

Table 3: **Matching estimates of households with female and male borrowers**

Outcome variables	**MF participants vs eligible non-participants**	
	1-Nearest neighbour matching	**Kernel matching, 0.05[32]**
Contraceptive use by currently married women aged 14-30	0.043	0.068***
Contraceptive use by currently married women aged 14-50	0.099***	0.138***
Contraceptive use by currently married women aged 30-50	0.080**	0.092***
Any child born in last 4 years to currently married women aged 14-30 (yes=1; no=0)	-0.015	0.001
Any child born in last 4 years to currently married women aged 14-50 (yes=1; no=0)	-0.013	-0.011

Source: Authors calculations.

Notes: *statistically significant at 10%, **statistically significant at 5%, ***statistically significant at 1%. STATA routine psmatch2[33] using the logit model outlined in Table 1, column 3 is used. Standard errors (not reported) are bootstrapped.

[30] The effect on female contraceptive use of having a male borrower in the household. In both cases we include cases with both male and female MF borrowers in the same household.

[31] The decision for using those algorithms was made in an arbitrary way since the literature in this area is not yet very developed. Morgan and Winship (2007, p. 109) argue that kernel matching which was first introduced by Heckman et al (1998) and Heckman, Ichimura and Todd (1998) appears to be the most efficient and preferred algorithm. In addition, 1-nearest neighbour matching was chosen for its popularity which is probably due to its being easy to understand and comparatively easy to implement. We present only the kernel matching estimates with a bandwidth of 0.05 but also used bandwidths 0.01 and 0.02.

[32] As mentioned earlier, 5-nearest neighbour matching as well as kernel matching with bandwidths 0.01 and 0.02 were applied but the results obtained from the various algorithms and bandwidths did not differ significantly from each other and confirm the results presented in Table 3.

[33] psmatch2 was developed by Leuven and Sianesi (2003), we also used pscore developed Becker and Ichino (2002) as a robustness check. The results obtained did not vary significantly.

The 1-nearest neighbour estimate of the impact of MF borrowing on the probability of contraceptive use is 0.043 (Table 3) indicating that MF participants (pooling across gender of borrowers) aged 14-30 are 4.3% more likely (not significantly so) to use contraceptives than matched non-participants. The kernel matching estimate indicates a 6.8% higher level of contraceptive use for participants than for matched non-participants at a 1% significance level. The impacts of MF on contraceptive use for the age group 14-50 and 30-50 are larger than for the 14-30 group and are consistently significant at mainly 1% (with one exception which is significant at 5%) and vary between 8.0% to 13.8%[34]. The results for fertility variables for both age groups are negative (with one exception) and insignificant, and thus we cannot reach any strong conclusions as to the impact of MF on fertility. Our PSM results cannot confirm the general view of the literature that MF reduces fertility but does support the view that MF appears to increase contraceptive use[35][36].

[34] PKML investigate contraceptive use for the ages 14-30 and 14-50 only. However, Buttenheim (2006) argues that contraceptive use is higher among older women and thus we investigate this claim and add a variable for contraceptive use looking at the ages 30-50.

[35] To test the robustness of our PSM results we also ran the analysis on different subgroups of borrowers, the results broadly confirm our findings presented in Table 3. In addition we applied an IV approach using different instruments and models, i.e. for some models we used eligibility and for others amount borrowed (as done by PKML) as instruments for treatment. The results were very mixed and contrary to the PSM results. The Hansen-Sargan test indicates that our instruments are valid for most of the model specifications we ran, however, as Deaton (2010) notes these tests are not particularly reliable. The difference between the PSM and IV estimates could be explained by selection on unobservables. IV claims to account for selection due to observables as well as unobservables while PSM only accounts for selection on observables and hence one could argue that the unobservables drive the differences in the results. This seems plausible since sensitivity analysis in Table 5 indicates that the unobservables indeed play a role.

[36] We applied PSM to the 1998/99 follow-up data separately and observed a slight change in the results compared to R1-3: none of the results for contraceptive use were significant anymore, fertility for the ages 14-30 turned negative and significant and fertility for ages 14-30 were negative and insignificant.

Table 4: **Matching estimates of impact segregated by gender**

Outcome variables	MF participants vs eligible non-participants			
	1-Nearest neighbour matching		Kernel matching, 0.05[37]	
Contraception	**Women**	**Men**	**Women**	**Men**
Contraceptive use by currently married women aged 14-30	0.092***	-0.021	0.067***	-0.012
Contraceptive use by currently married women aged 14-50	0.148***	-0.046	0.129***	-0.022
Contraceptive use by currently married women aged 30-50	0.077**	-0.046	0.080***	-0.021
Fertility				
Any child born in last 4 years to currently married women aged 14-30 (yes=1; no=0)	0.011	-0.007	0.010	0.010
Any child born in last 4 years to currently married women aged 14-50 (yes=1; no=0)	0.019	0.057	0.002	0.045

Source: Authors calculations.

Notes: *statistically significant at 10%, **statistically significant at 5%, ***statistically significant at 1%. STATA routine psmatch2[38] using the logit model outlined in Table 1, column 3 is used. Standard errors (not reported) are bootstrapped.

Table 4 presents the results by gender of borrower to test the claim that effects on women borrowers are different to those on male borrowers. This table shows that MF membership is significantly positively associated with contraceptive use for female borrowers across all age ranges. The impacts on fertility for both male and female borrowers are predominantly positive, but statistically insignificant for both age ranges. Thus, contrary to PKML but in agreement with the general literature (quoted above), we find that female borrowing has positive and significant effects on contraceptive use, and that male borrowing has largely positive but insignificant effects on fertility. However, we concur with PKML's findings that male borrowing has no effects on contraceptive use; we also concur with PKML, but contrary to the

[37] As in the case of the results presented in Table 3, 5-nearest neighbour matching as well as kernel matching with bandwidth 0.01 and 0.02 were applied in addition to 0.05 but the various algorithms and bandwidths results did not differ significantly and thus only the results using a bandwidth of 0.05 are shown here.

[38] As before, robustness checks were conducted using pscore. The results obtained did not vary significantly.

general literature, that female borrowing has mainly positive effects on fertility outcome variables, although our estimates are statistically insignificant.

It appears that our PSM results support the findings of the general literature on effects of MF on contraceptive use but not necessarily on fertility. We contradict some and weakly support other PKML findings, despite using the same data. Does this allow us to reach any strong conclusions as to the impact of MF on contraceptive use and fertility? As mentioned earlier, there is some controversy over the robustness of IV type estimates which heavily depend on adequate instruments (Morgan and Winship, 2007; Caliendo, 2006). The validity of instruments can be assessed using overidentification tests which, however, should be treated with caution (Deaton, 2010). OLS estimates are often more convincing (Heckman and Vytlacil, 2007)[39]. We conduct sensitivity analysis of our PSM estimates which allows us to explore why our findings differ from those of PKML.

Sensitivity analysis

Although we found statistically significant effects using PSM it is questionable whether these are robust to unobservables. Rosenbaum (2002) developed sensitivity analysis to explore the robustness of matching estimates to selection on unobservables. Ichino, Mealli and Nannicini (2006) argue that 'sensitivity analysis should always accompany the presentation of matching estimates' (p. 19).

As we wrote elsewhere:

> 'Rosenbaum (2002) invites us to imagine a number Γ (gamma) (≥ 1) which captures the degree of association, of an unobserved characteristic with the treatment and outcome, required for it (the unobserved characteristic) to explain the observed impact. Γ is the ratio of the odds that the treated have this unobserved characteristic to the odds that the controls have it; a low odds ratio (near to one) indicates that it is not unlikely that such an unobserved variable exists. Cornfield et al (1959) use the example of the effect of smoking on lung cancer. In this case, which is now surely without doubt, data from the late 1950s gives a gamma > 5 for such an unobserved variable, which is, it is suggested, highly unlikely to have been unobserved because of its strong association between smoking and death' (Duvendack and Palmer-Jones, 2011a).

This approach can be implemented using the **mhbounds** procedure in STATA (Becker and Caliendo, 2007), which is suitable for binary outcome variables[40].

[39] Vinod (2009) has suggested a form of simulation analysis to assess the robustness of IV estimates, but we do not pursue this here.
[40] The **rbounds** procedure in STATA is used for continuous outcomes.

mhbounds uses the matching estimates to calculate the lower and upper bounds of the outcome variable for different values of Γ. If the lowest Γ at which the treatment effect becomes insignificant is relatively small (say < 2) then the likelihood of an unobserved characteristic confounding the treatment effect is relatively high and the estimated impact is rather sensitive to the existence of unobservables (DiPrete and Gangl, 2004).

Table 3 shows that the kernel matching impact estimate with a bandwidth of 0.05 for contraceptive use for the ages 14-50 is 0.138 which is statistically significant at 1%. However, this may not be due to membership *per se* but to unobserved characteristics that account for membership (and/or its impact). Table 5 reports the **mhbounds** results, presenting the minimum and maximum values for the Mantel-Haenszel bounds along with their significance levels. If the value for the maximum significance level is above 0.05, then the result would no longer be significant at the 5% level, if the value is above 0.10, then the result would no longer be significant at 10%. In this case, the results are no longer significant at relatively low levels of Γ. For a Γ of 1.1 the result for contraceptive use aged 14-50 becomes insignificant at 5%, for a Γ of 1.2 they are no longer significant at 10%. This implies that a relatively small increase in the likelihood of being a participant due to an unobservable characteristic which also increases the benefits from borrowing is required to explain the observed impact. It is not unlikely that such an unobserved confounding variable exists; implying caution is required in concluding causality of MF on contraceptive use of fertility from these results.

Table 5: **Sensitivity analysis for contraceptive use ages 14-50 for microfinance participants**

	Mantel-Haenszel bounds		**Significance level**	
Gamma (Γ)	Minimum	Maximum	Minimum	Maximum
1	2.245	2.245	0.012	0.012
1.05	1.911	2.582	0.005	0.028
1.1	1.592	2.902	0.002	0.056
1.15	1.287	3.209	0.001	0.099
1.2	0.996	3.503	0.000	0.160
1.25	0.716	3.785	0.000	0.237
1.3	0.447	4.057	0.000	0.327

Source: Authors calculations.

Similar observations can be made when looking at contraceptive use for the ages 30-50; for a Γ of 1.3 the result become insignificant at 5% and for a Γ of 1.4 they are no longer significant at 10%[41].

Panel data

Given these findings are partially contradictory to PKML, we take advantage of the follow-up round collected in 1998/99 (henceforth R4) to examine the long-term effects of MF on contraceptive use and fertility (Khandker, 2005)[42].

The rate of attrition between survey rounds was 7.4 percent (Khandker, 2005, footnote 10, p. 271). The issue of attrition and the handling of dissolved households posed a challenge for the re-construction of Khandker's R4 data set, and attrition bias is potentially a concern. After formal testing, Khandker (2005) concludes that attrition bias can largely be ignored (ibid.). We also tested for attrition bias and find that it is strongly present, but when corrected using inverse probability weights (Fitzgerald et al, 1998) our results are not substantially altered.

By R4 the already small control group of the original PnK study was further diminished due to the rapid influx of MFIs expanding into the control villages of the 1991-1992 survey. The saturation of the market for microfinance has profound consequences for future studies evaluating the impact of microfinance in Bangladesh since finding suitable control groups, i.e. households that do not participate in microfinance or any other form of finance but are otherwise similar to participating households, has become increasingly difficult.

The panel was first analysed as a full panel and then by a combination of PSM and DID which Khandker, Koolwal and Samad (2010), among others, claim is the way forward to control for observable as well as unobservable time-invariant characteristics. For the latter PSM matches of R1-3 were retained and merged with their successor households in R4. Some treatment households that did not match on

[41] The sensitivity analysis results for the remaining outcome variables can be obtained from the authors upon request. For some outcomes and age groups the treatment effects become significant as Γ increases. As Becker and Caliendo (2007, p. 8-9) point out this is because increasing Γ implies an unobserved variable has an increasingly negative effect on outcome (and selection into treatment) which makes the observed outcome negative and significant at around Γ = 1.35 (p=0.05) for 14-30 year olds.

[42] In addition to the original households (and those that split from them) new households were sampled from the original villages as well as new villages in original and new thanas increasing the overall sample size to 2,599 households (Khandker, 2005, p. 271). We do not analyse the new households; the results replicating Khandker (2005) can be found in RnM and Duvendack (2010). There were several problems reconstructing the R4 variables, but we achieved a data set closely resembling that of RnM's data set for R4.

observable characteristics were dropped, and only matched households were merged with R4. In both analyses the following regression-adjusted model (equation 2) was run with random effects for all outcome variables[43]. A random effects model was chosen because time-invariant variables (such as the membership dummy variable) would be confounded with the fixed effects and could thus not be estimated using a fixed effects model (following Steele et al, 2001):

$$(2) \qquad y_{ijt} = \alpha_i + \delta_t + \beta C_{it} + \theta X_{it} + V_j + \varepsilon_{ijt}$$

Where:

y_{ijt} = outcome on which impact is measured for individual *i*, in village *j*, in period *t*

C_{it} = level of participation in microfinance, i.e. a membership dummy variable constructed based on eligibility criterion (ownership of < 0.5 acres of land), in period *t*

X_{it} = vector of household level characteristics in period *t*

V_j = vector of village level characteristics

α_i = effects unique to household *i*

δ_t = period effect common to all households in period *t*

β, θ = parameters to be estimated

ε_{ijt} = error term representing unmeasured household and village characteristics at period *t*

[43] We do not present the panel data analysis of the gender differentiated results here but they are available upon request.

Table 6: Impact of microcredit participation, comparison of full panel with PSM & DID model

Outcome variables	Full panel estimation	PSM and DID sample estimation
Contraceptive use by currently married women aged 14-30	0.243*	0. 124
Contraceptive use by currently married women aged 14-50	0.538***	0.117
Contraceptive use by currently married women aged 30-50	0.060***	0.180
Any child born in last 4 years to currently married women aged 14-30 (yes=1; no=0)	-0.039	0.095
Any child born in last 4 years to currently married women aged 14-50 (yes=1; no=0)	-0.053	-0.210
Number of observations	2,656	998

Source: Authors calculations.
Notes: *statistically significant at 10%, **statistically significant at 5%, ***statistically significant at 1%. PnK data across R1-3 and R4 downloaded from the World Bank website are used, STATA routine xtlogit is applied.

The random effects model on the full panel that corrects for attrition (see Table 6) indicates significantly positive effects for contraceptive use for women across all age brackets while the PSM/DID random effects model shows positive but insignificant effects. The fertility outcomes for both age ranges are mainly negative (with one exception) but insignificant across both models. The full panel results confirm the cross-section findings presented in Table 3 but these findings are not confirmed by the PSM/DID model which shows no effects for contraceptive use, this is contrary to the cross-section findings.

Conclusion

The literature suggests that MF has positive impacts on contraceptive use and negative impacts of fertility (see references above). The study by PKML using the same data as PnK throws doubts on these findings arguing that most of these studies have not accounted for self-selection and non-random programme placement bias. PKML propose an advanced econometric strategy to control for these biases. They examine the impact of MF by gender of borrower and find that female borrowing has significantly negative effects on contraceptive use and weak positive as well as negative effects on fertility; male borrowing has mainly positive but insignificant effects on contraceptive use and significantly negative effects on fertility. Steele et al (2001), using panel analysis with a sample from a similar domain, supported the orthodoxy that MF enhanced contraceptive use.

The findings of PKML are interesting and challenging; their data and estimation strategy are essentially the same as PnK's which have been the subject of ongoing controversy. We replicated the PKML variables with some difficulty, but triangulate our results successfully with the RnM data. When we apply PSM and follow Steele et al (2001) in using a dichotomous MFI membership variable as the indicator for MF participation, we obtain results which indicate that MF participation has positive and significant impacts on contraceptive use (contrary to PKML at least for females) and positive, albeit insignificant, impacts on fertility for both male and female borrowers. When the gender of the borrower is taken into account, we find that the results for female borrowing are more likely to be significant than those for males.

Overall, our PSM results confirm the findings of the broader MF literature on contraceptive use but not on fertility, and we can contradict some of the most striking PKML findings. However, sensitivity analysis has shown that the PSM estimates presented here are highly vulnerable to selection on unobservables and we cannot be confident about causality between MF membership and FP outcomes.

In the panel data analysis, the full panel random effects model confirms the findings of the cross-section data analysis and supports the orthodoxy. The PSM/DID model fails to show any significant effects of MF on these outcome variables. For contraception, a possible reason is that the effect of MF on contraception and/or fertility occurs before the period to which the baseline data refer, since people became members prior to 1991. Thus this is not a true before/after/with/without data set, and therefore may underestimate early impacts. However, for fertility (since 1988) this is not a plausible explanation. However, an alternative explanation for both types of outcome variables is that PSM and DID cannot account for selection on unobservables. What is compared is the change in outcomes between a group that was already participating in microfinance in R1-3 and a control group surveyed at the same time, with both groups at a later date. This comparison is not adequate for reliably assessing the impact of microcredit and controlling for unobservables because any differences between the treatment and control groups before microfinance cannot be empirically observed in these data.

Overall, the evidence of the impact of MF on contraceptive use and fertility remains contradictory and unreliable. One set of data subjected to alternative estimation methods gives rise to at least partially contradictory results. This raises questions about the key assumption many econometric methods are built on and 'the whimsical character of econometric inference' (Leamer, 1983, p. 38). We can only conclude that the evidence of MF impact on contraceptive use and fertility presented in this paper is partially contradictory to PKML findings, weak, and vulnerable to selection on unobservables. This also implies weaknesses in the underlying research design and data, and the inability of advanced econometric methods to compensate for these lacunae. An important question, perhaps relevant to current controversies over the role of RCTs in assessing development interventions, is why these deficiencies were not grasped earlier. Had this conclusion been reached at an earlier stage more and more rigorous evidence might by now have been available to answer the important question of whether there is any meaningful causal link between MF and these potentially beneficent outcomes.

Appendix 1: Weighted means and standard deviations for R1-3

Variables	PKML[1]			Authors, estimation sample		
	Number of Obs	Mean	Standard deviation	Number of Obs	Mean	Standard deviation
Age of woman	1,733	30.00	9.00	1,787	29.79	9.06
Age of household head (years)	1,757	40.82	12.80	1,787	42.05	12.18
Highest grade completed by HH head	1,757	2.49	3.50	1,787	2.49	3.44
Highest grade completed by any female HH member	1,757	1.61	2.85	1,787	1.67	2.97
Highest grade completed by any male HH member	1,757	3.08	3.80	1,787	3.32	3.97
Sex of household head (male=1)	1,757	0.95	0.22	1,787	1.01	0.12
Household land (decimals)	1,757	76.14	108.54	1,787	104.35	351.14
Parents of HH head own land?	1,725	0.26	0.56	1,787	0.27	0.59
Brothers of HH head own land?	1,725	0.82	1.31	1,787	0.69	1.21
Sisters of HH head own land?	1,725	0.76	1.21	1,787	0.72	1.17
Parents of HH head's spouse own land?	1,735	0.53	0.78	1,787	0.56	0.80
Brothers of HH head's spouse own land?	1,735	0.92	1.43	1,787	0.95	1.46
Sisters of HH head's spouse own land?	1,735	0.75	1.20	1,787	0.80	1.25
No spouse in HH	1,757	0.13	0.33	1,787	0.03	0.16
Nontarget HH	1,757	0.30	0.46	1,787	0.14	0.01
Has any primary school?	1,757	0.69	0.46	1,787	0.69	0.46
Has rural health center?	1,757	0.30	0.46	1,787	0.06	0.24
Has family planning center?	1,757	0.10	0.30	1,787	0.09	0.29
Is dai/midwife available?	1,757	0.67	0.47	1,787	0.68	0.47
Price of rice	1,757	11.15	0.85	1,787	10.54	0.63
Price of wheat flour	1,757	9.59	1.00	1,787	9.09	0.77
Price of mustard oil	1,757	52.65	5.96	1,787	53.65	4.21
Price of hen egg	1,757	2.46	1.81	1,787	2.35	0.69
Price of milk	1,757	12.54	3.04	1,787	12.28	2.49
Price of potato	1,757	3.74	1.60	1,787	6.94	0.93
Average female wage	1,757	16.15	9.61	1,787	18.01	6.68

Dummy variable for no female wage	1,757	0.19	0.40	1,787	0.02	0.15
Average male wage	1,757	37.89	9.40	1,787	36.70	6.91
Distance to bank (km)	1,757	3.49	2.85	1,787	3.48	2.89
Amount borrowed by female from BRAC (Taka)	183	4,678.41	3,561.60	185	4,994.97	3,831.71
Amount borrowed by male from BRAC (Taka)	70	5,685.99	7,091.58	70	7,026.62	9,276.42
Amount borrowed by female from BRDB (Taka)	108	4,094.27	1,931.91	122	3,929.41	2,155.04
Amount borrowed by male from BRDB (Taka)	180	5,996.86	6,202.16	197	5,819.88	5,781.10
Amount borrowed by female from GB (Taka)	233	14,123.59	9,302.40	241	15,567.58	9,737.45
Amount borrowed by male from GB (Taka)	85	16,468.14	10,580.00	90	18,016.63	10,966.17
Outcome variables[2]						
Contraceptive use by currently married women aged 14-30	1,058	0.398	0.490	1,099	0.389	0.488
Contraceptive use by currently married women aged 14-50	1,731	0.378	0.485	1,787	0.388	0.488
Contraceptive use by currently married women aged 30-50	n/a	n/a	n/a	1,787	0.184	0.387
Any child born in last 4 years to currently married women aged 14-30 (yes=1; no=0)	1,056	0.697	0.460	1,099	0.689	0.463
Any child born in last 4 years to currently married women aged 14-50 (yes=1; no=0)	1,729	0.553	0.497	1,787	0.543	0.498

Notes:
1. Source: PKML, table 2, p. 10 and table 3, p. 12.
2. Values for outcome variables are for all individuals across all villages.

PKML descriptive statistics are not on the estimation sample while our descriptive are on our estimation sample. There are slight differences in the number of observations; PKML run the majority of their descriptive statistics on a sample of 1,757 households while our sample is 1,787 households. PKML argue that they restrict their sample to those households with less than 5 acres of land owned and hence excluded 41 additional households from the overall sample of 1,798 (PKML, p. 10, footnote 8). The tabulations for R4 differ for some of the variables presented here, e.g. the education variables have higher values, the landownership ones across relatives are generally lower, etc. Details can be made available upon request.

References

Amin, R., Ahmed, A. U., Chowdhury, J. & Ahmed, M., 1994. Poor women's participation in income-generating projects and their fertility regulation in rural Bangladesh: Evidence from a recent survey. *World Development*, 22 (4), p.555-565.

Amin, R., Hill, R. B. & Li, Y., 1995. Poor Women's Participation in Credit-based Self-employment: The Impact on their Empowerment, Fertility, Contraceptive Use, and Fertility Desire in Rural Bangladesh. *The Pakistan Development Review*, 34 (2), p.93-119.

Amin, R., St. Pierre, M., Ahmed, A. & Haq, R., 2001. Integration of an Essential Services Package (ESP) in Child and Reproductive Health and Family Planning with a Micro-Credit Program for Poor Women: Experience from a Pilot Project in Rural Bangladesh. *World Development*, 29**,** p.1611-1621.

Angeles, G., Guilkey, D. K. & Mroz, T. A., 2005. The Effects of Education and Family Planning Programs on Fertility in Indonesia. *Economic Development and Cultural Change*, 54**,** p.165-201.

Armendáriz de Aghion, B. & Morduch, J., 2005. *The Economics of Microfinance.* Cambridge: MIT Press.

Armendáriz de Aghion, B. & Morduch, J., 2010. *The Economics of Microfinance, 2nd ed.* Cambridge: MIT Press.

Ashraf, N., Field, E. & Lee, J., 2010. Household Bargaining and Excess Fertility: An Experimental Study in Zambia. *Available at: http://ipl.econ.duke.edu/bread/papers/0910conf/Field.pdf.*

Banerjee, A. & Duflo, E., 2011. *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty.* New York: Public Affairs.

Bangladesh Bureau of Statistics, 1985. *Statistical Yearbook of Bangladesh 1984-85.* Dhaka: People's Republic of Bangladesh.

Bangladesh Bureau of Statistics, 2004. *Statistical Yearbook of Bangladesh 2004.* Dhaka: People's Republic of Bangladesh.

Basu, A., Heckman, J. J., Navarro-Lozano, S. & Urzua, S., 2007. Use of Instrumental Variables in the Presence of Heterogeneity and Self-selection: An Application to Treatments of Breast Cancer Patients. *Health Economics*, 16 (11), p.1133-1157.

Bateman, M., 2010. *Why Microfinance Doesn't Work? The Destructive Rise of Local Neoliberalism*. London: Zed Books.

Becker, S. O. & Caliendo, M., 2007. Sensitivity Analysis for Average Treatment Effects. *The STATA Journal*, 7 (1), p.71-83.

Becker, S. O. and Ichino, A., 2002. Estimation of Average Treatment Effects Based on Propensity Scores. *The STATA Journal,* 2 (4), p.358-377.

Blackwell, M., Iacus, S., King, G. & Porro, G., 2009. CEM: Coarsened Exact Matching in Stata. *The STATA Journal,* 9**,** p.524-546.

Burman, L. E., Reed, W. R. et al., 2010. A Call for Replication Studies. *Public Finance Review,* 38(6), p.787-793.

Buttenheim, A., 2006. Microfinance Programs and Contraceptive Use: Evidence from Indonesia. Working Paper Series, California Center for Population Research, UC Los Angeles, November.

Caliendo, M., 2006. *Microeconometric Evaluation of Labour Market Policies.* Berlin: Springer.

Caliendo, M. & Kopeinig, S., 2005. Some Practical Guidance for the Implementation of Propensity Score Matching. Forschungsinstitut zur Zukunft der Arbeit (IZA) Discussion Paper No. 1588, May.

Caliendo, M. & Kopeinig, S., 2008. Some Practical Guidance for the Implementation of Propensity Score Matching. *Journal of Economic Surveys*, 22 (1), p.31-72.

Chemin, M., 2008. The Benefits and Costs of Microfinance: Evidence from Bangladesh. *Journal of Development Studies,* 44 (4), p.463-484.

Cleland, J., Bernstein, S. et al., 2006. Family Planning: the Unfinished Agenda. *The Lancet* 368 (9549), p.1810-1827.

Cleland, J., 2009. Contraception in Historical and Global Perspective. *Best Practice & Research Clinical Obstetrics & Gynaecology*, 23(2), p.165-176.

Coleman, B. E., 1999. The Impact of Group Lending in Northeast Thailand. *Journal of Development Economics,* 60 (1), p.105-141.

Collins, H. M., 1991. The Meaning of Replication and the Science of Economics. *History of Political Economy,* 23(1), p.123-142.

Cornfield, J., Haenszel, W., Hammond, E. & Lilienfeld, A., 1959. Smoking and Lung Cancer: Recent Evidence and a Discussion of Some Questions. *Journal of the National Cancer Institute,* 22, p.173-203.

Daley-Harris, S., 2002. State of the Microcredit Summit Campaign Report 2002. *Available at: http://www.microcreditsummit.org/pubs/reports/socr/2002/socr02_en.pdf.*

Deaton, A., 2010. Instruments, Randomization, and Learning about Development. *Journal of Economic Literature* 48, p.424-456.

Desai, J. & Tarozzi, A., 2011. Microcredit, Family Planning Programs and Contraceptive Behavior: Evidence from a Field Experiment in Ethiopia. *Demography*, 48 (2), p.749-782.

Dewald, W. G., Thursby, J. G. et al., 1986. Replication in Empirical Economics: the Journal of Money, Credit and Banking Project. *American Economic Review*, 76, p.587-603.

Dichter, T. and Harper, M. eds., 2007. *What's Wrong with Microfinance?* Warwickshire: Practical Action Publishing.

DiPrete, T. A. & Gangl, M., 2004. Assessing Bias in the Estimation of Causal Effects: Rosenbaum Bounds on Matching Estimators and Instrumental Variables Estimation with Imperfect Instruments. *Sociological Methodology*, 34 (1), p.271-310.

Duvendack, M., 2010. Smoke and Mirrors: Evidence from Microfinance Impact Evaluations in India and Bangladesh. *Unpublished PhD Thesis. School of International Development.* Norwich: University of East Anglia.

Duvendack, M. & Palmer-Jones, R., 2011a. Comment on: Abou-Ali, H., El-Azony, H., El-Laithy, H., Haughton, J. & Khandker, S., 2010. Evaluating the Impact of Egyptian Social Fund for Development Programmes. *Journal of Development Effectiveness*, 2 (4), p.521 – 555. *Journal of Development Effectiveness 2011*, 3 (2), p.297-299.

Duvendack, M. & Palmer-Jones, R., 2011b. High Noon for Microfinance Impact Evaluations: Re-investigating the Evidence from Bangladesh. Working Paper 27, DEV Working Paper Series, The School of International Development, University of East Anglia, UK.

Duvendack, M., Palmer-Jones, R. et al., 2011. What is the Evidence of the Impact of Microfinance on the Well-being of Poor People? EPPI Centre, Social Science Research Unit, Institute of Education, University of London.

Fernando, J. L., 1997. Nongovernmental Organizations, Micro-Credit, and Empowerment of Women. *The ANNALS of the American Academy of Political and Social Science*, 554 (1), p.150-177.

Fitzgerald, J., Gottschalk, P. & Moffit, R., 1998. An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics. *Journal of Human Resources*, 33(2), p.251-299.

Gaile, G. L. & Foster, J., 1996. Review of Methodological Approaches to the Study of the Impact of Microenterprise Credit Programs. Report submitted to USAID Assessing the Impact of Microenterprise Services (AIMS), June.

Gertler, P. & Molyneaux, J. W., 1994. How Economic Development and Family Planning Programs Combined to Reduce Indonesian Fertility. *Demography*, 31**,** p.33-63.

Goldberg, N., 2005. Measuring the Impact of Microfinance: Taking Stock of What We Know. Grameen Foundation USA Publication Series, December.

Hamermesh, D. S., 2007. Viewpoint: Replication in Economics. *Canadian Journal of Economics*, 40 (3), p.715-733.

Hashemi, S. M., Schuler, S. R. & Riley, A. P., 1996. Rural Credit Programs and Women's Empowerment in Bangladesh. *World Development*, 24 (4), p.635-653.

Heckman, J. J., Ichimura, H., Smith, J. & Todd, P., 1998. Characterizing Selection Bias Using Experimental Data. *Econometrica,* 66 (5), p.1017-1098.

Heckman, J. J., Ichimura, H. & Todd, P., 1998. Matching as an Econometric Evaluation Estimator. *The Review of Economic Studies,* 65 (2), p.261-294.

Heckman, J. J. & Vytlacil, E., 2007. Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast Their Effects in New Environments. In Heckman, J. J. & Leamer, E. E., eds. *Handbook of Econometrics, Volume 6B.* Amsterdam: North-Holland.

Ichino, A., Mealli, F. & Nannicini, T., 2006. From Temporary Help Jobs to Permanent Employment: What Can We Learn from Matching Estimators and their Sensitivity? Forschungsinstitut zur Zukunft der Arbeit (IZA) Discussion Paper No. 2149, May.

Jain, S. & Mansuri, G., 2003. A Little at a Time: The Use of Regularly Scheduled Repayments in Microfinance Programs. *Journal of Development Economics* 72, p.253-279.

Karlan, D. & Appel, J., 2011. *More than Good Intentions: How a New Economics Is Helping to Solve Global Poverty*. London: Penguin.

Khandker, S. R., 1996. Role of Targeted Credit in Rural Non-farm Growth. *Bangladesh Development Studies,* 24 (3 & 4).

Khandker, S. R., 2000. Savings, Informal Borrowing and Microfinance. *Bangladesh Development Studies*, 26 (2 & 3).

Khandker, S. R., 2005. Microfinance and Poverty: Evidence Using Panel Data from Bangladesh. *The World Bank Economic Review,* 19 (2), p.263-286.

Khandker, S. R., Koolwal, G. B. & Samad, H. A., 2010. *Handbook on Impact Evaluation:*

Quantitative Methods and Practices. Washington, D.C.: The World Bank.

King, G., Nielsen, R., Coberley, C., Pope, J. E. & Wells, A., 2011. Comparative Effectiveness of Matching Methods for Causal Inference. *Available at: http://gking.harvard.edu/gking/files/psparadox.pdf.*

Leamer, E. E., 1983. Let's Take the Con Out of Econometrics. *The American Economic Review*, 73 (1), p.31-43.

Leatherman, S., Metcalfe, M., Geissler, K. & Dunford, C., 2011. Integrating Microfinance and Health Strategies: Examining the Evidence to Inform Policy and Practice. *Health Policy and Planning***,** p.1-17.

Leuven, E. and Sianesi, B., 2003. PSMATCH2: STATA Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Matching. *Available at: http://ideas.repec.org/c/boc/bocode/s432001.html.*

Littlefield, E., Morduch, J. & Hashemi, S., 2003. Is Microfinance an Effective Strategy to Reach the Millennium Development Goals? *CGAP Focus Note 24, January.*

Livi-Bacci, M. & de Santis, G., 1998. *Population and Poverty in the Developing World*. Oxford: Clarendon Press.

McCullough, B. D., McGeary, K. A. and Harrison, T. D., 2006. Lessons from the JMCB Archive. *Journal of Money, Credit, and Banking*, 38 (4), p.1093-1107.

McCullough, B., K. A. McGeary, et al., 2008. "Do Economics Journal Archives Promote Replicable Research." *Canadian Journal of Economics*, 41(4), p.1406-1420.

McKernan, S.-M., 2002. The Impact of Microcredit Programs on Self-Employment Profits: Do Noncredit Program Aspects Matter? *Review of Economics and Statistics,* 84 (1), p.93-115.

Menon, N., 2006. Non-linearities in Returns to Participation in Grameen Bank Programs. *Journal of Development Studies,* 42 (8), p.1379 - 1400.

Morduch, J., 1998. Does Microfinance Really Help the Poor? New Evidence from Flagship Programs in Bangladesh. Unpublished mimeo.

Morgan, S. L. & Winship, C., 2007. *Counterfactuals and Causal Inference. Methods and Principles for Social Research.* Cambridge: Cambridge University Press.

Nannicini, T., 2007. Simulation-based Sensitivity Analysis for Matching Estimators. *The STATA Journal*, 7 (3), p.334-350.

Odell, K., 2010. Measuring the Impact of Microfinance: Taking Another Look. Grameen Foundation USA Publication Series, May.

Pitt, M. M., Khandker, S. R. & Cartwright, J., 2006. Empowering Women with Micro-finance: Evidence from Bangladesh. *Economic Development and Cultural Change*, p.791-831.

Pitt, M. M., Khandker, S. R., Chowdhury, O. H. & Millimet, D. L., 2003. Credit Programmes for the Poor and the Health Status of Children in Rural Bangladesh. *International Economic Review,* 44 (1), p.87-118.

Pitt, M. M., 1999. Reply to Jonathan Morduch's "Does Microfinance Really Help the Poor? New Evidence from Flagship Programs in Bangladesh". Unpublished mimeo.

Pitt, M. M., 2000. The Effect of Nonagricultural Self-Employment Credit on Contractual Relations and Employment in Agriculture: The Case of Microcredit Programs in Bangladesh. *Bangladesh Development Studies,* 26 (2 & 3), p.15-48.

Pitt, M. M., 2011a. Overidentification Tests and Causality: A Second Response to Roodman and Morduch. *Available at: http://www.pstc.brown.edu/~mp/papers/Overidentification.pdf.*

Pitt, M. M., 2011b. Response to Roodman and Morduch's "The Impact of Microcredit on the Poor in Bangladesh: Revisiting the Evidence". *Available at: http://www.pstc.brown.edu/~mp/papers/Pitt_response_to_RM.pdf.*

Pitt, M. M. & Khandker, S. R., 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), p.958-996.

Pitt, M. M. & Khandker, S. R., 2002. Credit Programmes for the Poor and Seasonality in Rural Bangladesh. *Journal of Development Studies,* 39 (2), p.1-24.

Pitt, M. M., Khandker, S. R., McKernan, S.-M. & Latif, M. A., 1999. Credit Programs for the Poor and Reproductive Behavior of Low-Income Countries: Are the Reported Causal Relationships the Result of Heterogeneity Bias? *Demography,* 36 (1), p.1-21.

Ravallion, M., 2001. The Mystery of the Vanishing Benefits: An Introduction to Impact Evaluation. *The World Bank Economic Review,* 15 (1), p.115-140.

Ravallion, M., 2008. Evaluating Anti-Poverty Programs. In T. P. Schultz & J. Strauss, eds. *Handbook of Development Economics, Volume 4.* Amsterdam: Elsevier.

Ravallion, M., 2011. On the Implications of Essential Heterogeneity for Estimating Causal Impacts Using Social Experiments. Washington DC; The World Bank.

Roodman, D. & Morduch, J., 2009. The Impact of Microcredit on the Poor in Bangladesh: Revisiting the Evidence. Center for Global Development, Working Paper No. 174, June.

Rosenbaum, P. R., 2002. *Observational Studies.* New York: Springer.

Rosenbaum, P. R. & Rubin, D. B., 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika,* 70 (1), p.41-55.

Rosenbaum, P. R. & Rubin, D. B., 1984. Reducing Bias in Observational Studies Using Subclassification on the Propensity Score. *Journal of the American Statistical Association,* 79 (387), p.516-524.

Roy, A., 2010. *Poverty Capital: Microfinance and the Making of Development*. Routledge: London.

Rutherford, S., 2009. *ASA: Peasant Politics, and Microfinance in the Development of Bangladesh.* Oxford: Oxford University Press.

Schuler, S. R. & Hashemi, S. M., 1994. Credit Programs, Women's Empowerment, and Contraceptive Use in Rural Bangladesh. *Studies in Family Planning*, 25 (2), p.65-76.

Schuler, S. R., Hashemi, S. M. & Riley, A. P., 1997. The Influence of Women's Changing Roles and Status in Bangladesh's Fertility Transition: Evidence from a Study of Credit Programs and Contraceptive Use. *World Development*, 25 (4), p.563-575.

Sebstad, J. & Chen, G., 1996. Overview of Studies on the Impact of Microenterprise Credit. Report submitted to USAID Assessing the Impact of Microenterprise Services (AIMS), June.

Smith, J. A. & Todd, P., 2005. Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators? *Journal of Econometrics*, 125, p.305-353.

Steele, F., Amin, S. & Naved, R. T., 2001. Savings/Credit Group Formation and Change in Contraception. *Demography*, 38 (2), p.267-282.

Stewart, R., van Rooyen, C., Dickson, K., Majoro, M. & de Wet, T., 2010. What Is the Impact of Microfinance on Poor People? A Systematic Review of Evidence from Sub-Saharan Africa. Technical Report, EPPI-Centre, Social Science Research Unit, University of London.

Todd, H., 1996. *Women at the Centre.* Dhaka: University Press Limited.

UNCDF, 2005. International Year of Microcredit 2005. *Available at: http://www.uncdf.org/english/microfinance/uploads/thematic/2005_Final_Report.pdf.*

Vinod, H. D., 2009. Stress Testing of Econometric Results Using Archived Code for Replication. *Journal of Economic and Social Measurement*, 34, p.205-217.

Zeller, M., Sharma, M., Ahmed, A. U. & Rashid, S., 2001. Group-based Financial Institutions for the Rural Poor in Bangladesh: An Institutional- and Household-level Analysis. Research Report of the International Food Policy Research Institute, (120), p.97-100.