The microfinance of reproduction and the reproduction of microfinance: understanding the connections between microfinance, empowerment, contraception and fertility in Bangladesh in the 1990s.

Duvendack, Maren and Palmer-Jones, Richard

University of East Anglia, Norwich, UK

June 2011

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

Authors: Maren Duvendack[1] and Richard Palmer-Jones[2]

DRAFT version, comments welcome. Do not quote without consent of the authors.

Abstract

Microfinance (MF) and family planning (FP) are thought to be very important interventions in the promotion of human development and it has been suggested that MF has significant beneficent impacts on contraceptive adoption and fertility. Thus, several authors, e.g. 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), using naive methods find that MF in Bangladesh increases contraceptive use and reduces fertility at the individual level, 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, and male borrowing decreases fertility, perhaps because fertility increasing income effects of MF outweigh substitution. Steele et al (2001), also using data from Bangladesh from around the same time as the PKML study, come to conclusions closer to the orthodoxy, arguing that PKML use an inappropriate metric for MF programme participation. In this paper we apply matching methods to our reconstruction of the PKML data to test whether other methods reproduce their results. We find that female borrowing substantially increases contraceptive use but has mainly no effects on fertility, while male borrowing has no effect on contraceptive use or on fertility; this contradicts some of the findings of PKML. Our results are shown to be vulnerable to unobservables, but there is no reason to believe that results on IV based methods are more reliable.

[1] Postdoctoral fellow at the International Food Policy Research Institute, Washington DC, USA.
[2] Reader at the School of International Development, University of East Anglia, Norwich, UK.

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, Bernstein, et al, 2006; Cleland, 2009). MF is not just about credit; it encompasses other services such as savings, insurance, and remittances, and non-financial services such as financial literacy training (Armendáriz de Aghion and Morduch, 2005) and it is now often combined with other interventions, for example providing information and advice about contraception and fertility (Leatherman et al, 2011).

It is often argued that access to credit improves women's economic opportunities and affects their FP by increasing the value of their time (Desai and Tarozzi, 2009; Pitt et al, 1999; Buttenheim, 2006). However, it is unclear whether an increase in the value of their time has positive or negative effects on fertility because while it makes reproduction more costly because time consuming, it may be associated with an opposite effect due to concomitant rise in income. A positive effect on fertility is more likely when the additional time women spend on economic activities leads to a significant increase in income which in turn increases the demand for children on the assumption that children are normal goods (Pitt et al, 1999, p. 2). In other words, the direction of the impact of MF on fertility is unclear (Desai and Tarozzi, 2009) and few studies (discussed below) have tried to test this causal link between MF and FP decisions.

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 evidence 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 (particularly argued by Duvendack et al, 2011), thus 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). It has been suggested there are significant beneficent impacts of MF on contraceptive use and fertility. Thus 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), find that in Bangladesh MF increases contraceptive use and reduces fertility at the individual level, putatively because of the effects MF lent to women has on empowering them. It is assumed that women prefer contraceptive use and fewer children than men in this patriarchal society. Pitt et al (1999 – henceforth 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[3]; PKML's study, differentiates 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 produced around the same time as those analysed by PKML, in a pipeline design similar to Coleman (1999), employ fixed and random effects panel estimation methods to control for self-selection and programme placement bias. They obtain results which are contradictory to those of PKML and conclude that MF has a positive impact on contraceptive use. Steele et al (2001, p. 280) 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, which is the variable used by PKML to represent MF participation, because it more appropriately proxies female empowerment. In their data, Steele et al (2001) have cases of women who are members of the MF group but have not borrowed, but who, they argue, nevertheless may be empowered by group meetings and solidarity. Thus using a dummy variable for membership (empowerment) Steele et al (2001) find significant impacts of MF membership on contraceptive use. The amount borrowed, the indicator used by PKML may only proxy income changes, and miss these wider effects.

[3] 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; PnK).

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 data sets are different[4]; in Steele et al (2001) women (only) are members of groups facilitated by Save the Children USA and the Bangladeshi NGO ASA (Rutherford, 2009) while in the PKML data males and females are members of Grameen Bank (GB), the Bangladesh Rural Advancement Committee (BRAC) or the Bangladesh Rural Development Board (BRDB) groups. 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[5]. The PKML sample includes members of GB, BRDB and BRAC[6], which each espoused rather different interactions with group members; although in each case some might be considered empowerment related (c.f. the GB 16 affirmations) 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 at best indirect evidence of empowerment and income respectively.

Desai and Tarozzi (2009) discuss this literature and report an RCT conducted in Ethiopia, with data from before and after the intervention with three different treatment groups and a control group[7]. Randomisation occurred at the village level. The first treatment group had access to microcredit as well as family planning services, the second and third treatment groups had access to either microcredit or family planning and the fourth group did not have access to either of these services. The authors find that none of these interventions - alone or in combination - had any impact on increased contraceptive use compared to the control group. They conclude that the family planning intervention provided the wrong contraceptive methods and hence the

---

[4] 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 Eastern Bangladesh.

[5] However, the number of women in the sample joining Save the Children USA 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)).

[6] 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, 2011).

[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; Duvendack and Palmer-Jones et al, 2011).

lack of impact. Women in the study area preferred injectibles while pills and condoms were offered. There are possible limitations of this study. As argued by Leatherman et al (2011), 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, the households during baseline and follow-up surveys are not the same, new households were sampled in the follow-up survey, thus making the study a panel of villages and not of households.

This leaves the issue of the relationship between MF, contraceptive use and fertility decidedly unclear, yet it is of continuing importance (Buttenheim, 2006); there seems to be evidence both for and against beneficent pathways between MF and contraceptive use and fertility, warranting further exploration of these issues. At the theoretical level it has been argued that 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, while 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 and points towards other factors that can also play a role. 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 the desire to have more children in 2000 is higher than in 1993 and 1997 irrespective of age and education level of the women interviewed. She speculates that this could have been due to Indonesia's slow recovery from the economic crisis in 1998 (Buttenheim, 2006, p. 15). 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 (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) – in part because the ability of establishing causality in the PnK research design has been contested (Roodman and Morduch, 2009 – henceforth RnM), and partly 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[8], by balancing the covariates in the samples of treatment and control groups (DiPrete and Gangl, 2004, p. 276).

It is only partially possible test Steele et al's (2001) assertions that membership is more likely to show positive effects on contraceptive use and negative ones on fertility than amount borrowed because all members are borrowers in the PKML data. The PKML dataset and estimation strategy is largely the same as that used by PnK. The data and analysis has been difficult to replicate; however, RnM replicated the key PnK studies[9], using similar estimation strategies but different software, and come to the same results as PnK, but conclude differently, as noted above.

> '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)[10].

Duvendack (2010) and Duvendack and Palmer-Jones (2011 - henceforth DPJ) using PSM and sensitivity analysis conclude that the very modest and mixed impacts they find are highly vulnerable to unobservables – such as entrepreneurial ability, motivation, risk preferences, and so on - which may account for both MFI membership and the impacts observed. These difficulties 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)), since they use broadly the same data and similar estimation strategies and software[11]. Ideally we would want to re-investigate this set of studies but it is beyond the scope of this paper and hence we restrict ourselves to the study by PKML on contraceptive use and fertility.

Replication and reproduction[12] are an important part of scientific practice, even in economics, without which results cannot be taken as robust (Hamermesh, 2007;

---

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

[9] 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).

[10] However, RnM's results have been refuted (see Pitt, 2011a and 2011b) but the debate is ongoing and the interested reader is referred to Roodman's blog: http://blogs.cgdev.org/open_book/ for updates. Roodman maintains the basic conclusion that PnK do not establish causality due to the unsatisfactory research (sample) design of the PnK study.

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

[12] While used in various ways in this literature (McCullough et al, 2006) we consider the terms replication and reproduction to cover applying the same data construction and statistical methods to a given raw

McCullough et al, 2006), perhaps especially when there appear to be contradictions with much other relevant work. Economics papers applying complex statistical methods can have errors of both variable construction (data manipulation) and statistical estimation (Dewald et al, 1986; McCullough et al, 2006; McCullough et al, 2008). The application of different methods can lead to different conclusions with practical relevance (op. cit.). To allow replication, good documentation of the study design and data are required, and there should be access to the data, and details of their variable construction and analysis[13].

As noted above, the findings of PKML differ from most other findings of the relation between MF and contraception for Bangladesh, or indeed elsewhere; a similar conclusion as to the general evidence of MF impact and contraceptive use and women's empowerment has been drawn by Buttenheim (2006) for Indonesia using an apparently robust set of data and estimation method. Indeed Pitt et al (2006) find MF borrowing empowers females using data on their original sample collected at a later date. The explanation given by Pitt et al (2006), which fits better the orthodoxy - that MF empowers women -, relies on an extended and perhaps somewhat tendentious argument to reconcile the PKML findings with regard to contraception and fertility with their later findings on female empowerment. It is possible, however, given the critique of the data and methods used in the PnK oeuvre that the original empirical findings cannot be replicated, or are found not to be robust, which would go some way to resolving this paradox without recourse to asserting the dominance of the income over substitution effects. This paper furthers the critical discussion of the extensive and influential set of studies using the PnK data set and possibly helps to resolve some of the puzzles regarding links between MF and well-being raised by them.

Thus, to summarise, 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

---

data set, application of different statistical methods to the same data set, and application of the same or different methods to a different data set which is arguably equivalent to the original study. However, in this paper we use reproduction to apply to use of different methods on a given (raw) data set, and replication to mean using the same (or very similar) methods. Data, or variable constructions may differ because of definitional differences using the same concept of the variables (replication), or different concepts of the variables (reproduction).

[13] 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. Authors have been compliant with this policy so far but can opt out in case their data are proprietary and/or confidential (Hamermesh, 2007, p. 717). This is not the common practice in economics, although increasingly advocated (Burman et al, 2010).

and sensitivity analysis to the data to triangulate these findings and obtain more refined impact estimates.

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[14] on three leading microfinance group-lending programmes in Bangladesh, namely 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[15]; 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 out of which 1,538 were labelled target households, putatively cultivating less than 0.5 acres at the time of joining the MFI[16], and 260 were non-target households (PnK, p. 974). According to PnK (p. 974), out of those 1,538 households, 905 effectively participated in microfinance (59%). The three survey waves (henceforth R1-3) were timed to account for seasonal variations, (Pitt, 2000, p. 28-29). The study focuses on measuring the impact of microfinance participation by gender on indicators such as labour supply, school enrolment, expenditure per capita and non-land asset ownership. PnK find that microcredit has significant positive impacts on many of these indicators and find larger positive impacts when women are involved in borrowing. 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)[17].

---

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

[15] 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).

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

[17] A follow-up data set (henceforth R4) was collected in 1998-1999 re-surveying the same households that were already interviewed in R1-3 and some new households increasing the overall sample size to 2,599 households (Khandker, p. 271). Khandker uses standard panel analysis to conclude that microcredit has positive impacts on the poorest and reduces poverty among programme participants, especially when women are involved in borrowing, and thus confirms PnK's headline results. However, RnM's replication of Khandker casts doubts about Khandker's approach and findings (RnM, p. 39). Using our, slightly different data we concur with RnM that panel estimation does not show clear evidence of microfinance impact. We do not further discuss this approach here.

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 to the presence of the MFI in the village. However, comparing households in treatment and control villages is not sufficient for obtaining impact estimates for microfinance programme participation because the villages differ (there is programme placement bias[18]) and households commonly self-select into microfinance. In this type of group-based lending 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 an eligibility criterion that participating households should be cultivating[19] less than 0.5 acres of land at the time of recruitment into the MFI programme, so that only households meeting this criterion are eligible. In fact, the eligibility criterion is not strictly met by quite a few microfinance borrowers (Morduch, 1998), so that there is a gap between participation and eligibility[20] and this raises further doubts about PnK's estimation strategy.

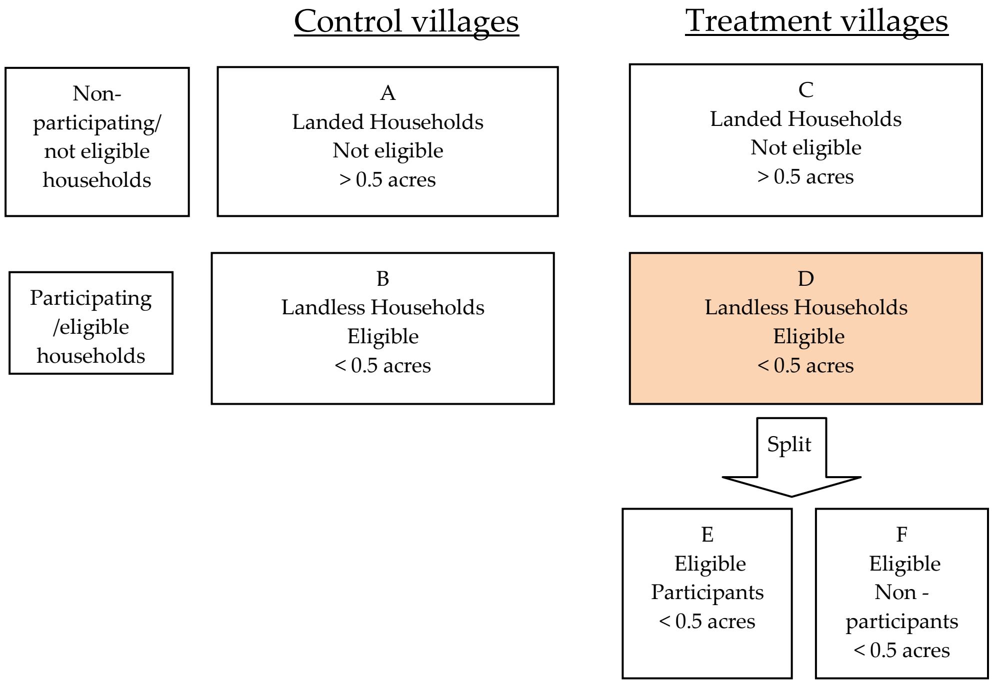
In brief, PnK use the (de facto) participation criterion as their identification strategy, assuming that it is exogenous. They sample treatment and control villages containing non-target/landed and target/landless households. PnK's (ideal) identification strategy can be understood graphically by looking at Figure 1.

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

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

[20] Thus there are de jure (cultivating less than 0.5 acres), and de facto (participating) eligibility categories; this is discussed further below. Some 42% of de facto MFI members are not eligible by the 0.5 acres rule.

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 discontinuity or cut-off point 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 unobserved differences between treatment and control villages.

The application of an eligibility criterion as an identification strategy is plausible provided it is strictly enforced. However, as Morduch (1998) points out, mistargeting occurred[21] (see also Ravallion, 2008, p. 3818; Chemin, 2008, p. 465). Group D contains participants who own 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

[21] 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.

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[22]. RnM were eventually able to replicate the original PnK data if not exactly[23], as independently did DPJ, but come to different conclusions as to the claim of causality.

Chemin (2008) using PSM came to different conclusions as to the impacts of MF from PnK. DPJ could not replicate Chemin's (2008) data closely, or findings, although they to 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 which are elaborated in more detail by DPJ. 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. While it might be desirable to compare more homogenous sub-groups separately so one could distinguish differences in impacts and probably obtain more precise and statistically significant results, this is constrained by sample sizes in existing data sets. DPJ explore these issues in a separate paper and it is beyond the scope of this paper to elaborate this further.

---

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

[23] 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.

Estimation strategy

The standard approach to solving the evaluation problem 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 a variable or a set of variables, i.e. instruments, that influence the decision to participate in a programme but at the same time do not have an effect on the outcome equation. Only when adequate instruments can be identified, then the IV approach is an effective strategy for estimating causal effects (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 Pitt et al, 1999 and Steele et al, 2001). These drawbacks of the IV methods suggest using a different approach to estimating causal effects, in this case PSM.

First we replicate the variable constructions of PKML[24] (see DPJ for further details) and then apply PSM using MF membership, rather than amount borrowed, to explain contraceptive use and fertility. PSM matches participants to non-participants based on a "propensity score" computed generally from a logit or probit model (Caliendo and Kopeinig, 2005 and 2008) to predict the probability (the "propensity score") of programme participation (Ravallion, 2001). The samples of participants and the matched non-participants are pooled to estimate programme participation. PSM accounts for selection on observables but fails to control for selection on unobservables (Smith and Todd, 2005). Selection on unobservables or 'hidden bias' as Rosenbaum (2002) calls them, are driven by unobserved variables that influence treatment decisions as well as potential outcomes (Becker and Caliendo, 2007). Matching estimators are commonly not robust to selection on unobservables. Hence, to test the robustness of our matching estimates, we also apply sensitivity analysis.

---

[24] 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. Firstly, the survey forms and variable descriptions are problematic; secondly certain data necessary for replication were (and others are) missing. It is not possible to be sure that the data posted are indeed exactly the same as those analysed by PnK, but the main problems probably lie not in variations in the raw data but in subsequent manipulations, variable constructions, and analytical procedures. Some of these data such as data on consumer price indices, sampling weights and landholding details were obtained after contacting the authors (either by Roodman, or ourselves). The replication exercise reported here was greatly facilitated by RnM who have made all their data and codes available: http://www.cgdev.org/content/publications/detail/1422302. The data and variable construction are mainly in SQL, although statistical analysis is in STATA; our data manipulation and analysis is all in STATA.

For PSM, we first estimate of the likelihood of microfinance participation (propensity score) to match control to treatment cases using the propensity score, and then compute the treatment effects for the various comparison groups. Our logit model specification follows the model set out by Pitt et al (1999)[25]. The model can be expressed as follows:

$$(1) \qquad \text{Logit}\ (\mathrm{y}_{ij}) = \alpha + \beta \mathrm{C}_{ij} + \gamma \mathrm{G}_{ij} + \delta Z_{ij}$$

<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 eligible participants (i) in village (j); a value of 1 is assumed when an individual participates and a value of 0 if not. 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. All estimations use village-level fixed-effects.

Results

As noted above replicating the PnK data is not a trivial task, however 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 to a very high degree of accuracy with RnM although using different software. Remaining differences in the variables are due to differences in interpretation of the raw data rather than differences in data manipulations (see Appendix 3 for details on the remaining discrepancies).

The logit specification follows that of PKML as illustrated in Table 1, the descriptive statistics can be found in Appendix 1.

[25] 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, so we use their model, and due to constraints of space.

Table 1: Logistic regression model for MF participation using PKML's model specification

Independent variables	PKML's specifications
Age (years)	0.084*** *0.000*
Age household head (years)	-0.033*** *0.000*
Highest education any male household member	-0.117*** *0.004*
Household land (decimals)	-0.004*** *0.000*
Has any primary school?	-0.495** *0.024*
Midwife available?	-0.561* *0.096*
Price of rice	1.113*** *0.000*
Price of wheat flour	-0.548** *0.050*
Price of mustard oil	0.092** *0.041*
Average female wage	-0.060** *0.031*
Average male wage	-0.122*** *0.000*
Village dummies	Yes
Number of observations	1841
Pseudo R-squared	0.393

Source: Authors calculations.

Notes: p-values in italics. * significant at 10%, ** significant at 5%, *** significant at 1%. Data round 1 is used. The following control variables are used in PKML's model specification as well: maximum education household head, highest education any female household member, sex of households head, landholdings household head parents, landholdings household head brother, landholdings household head sister, landholdings household head spouse parents, landholdings household head spouse brother, landholdings household head spouse sister, no spouse in household, nontarget household, access to rural health care, access to family planning centre, price of hen egg, price of milk, price of potato, dummy for no female wage, distance to bank (km), all insignificant. Descriptive statistics for all logit variables can be found in Appendix 1.

In this specification age of respondent, age of household head, highest education of any male household member, household land, price of rice and average male wage are statistically significant at 1%. Has any primary school, price of mustard oil, average female wage and price of wheat flour are significant at 5% and access to midwife is significant at 10%. The pseudo R-squared is rather low at 0.393. 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 Treated	Mean Control	Bias (%)	% Reduction in |Bias|	t-test p>|t|
Age (years)	Unmatched	31.936	28.799	35.8		0.000
	Matched	31.936	31.840	1.1	96.9	0.872
Age household head	Unmatched	41.143	41.543	-3.4		0.545
	Matched	41.143	41.200	-0.5	85.8	0.937
Highest education male household	Unmatched	2.470	3.417	-25.3		0.000
	Matched	2.470	2.455	0.4	98.4	0.947
Household land	Unmatched	39.745	113.88	-25.4		0.000
	Matched	39.745	43.372	-1.2	95.1	0.579
Has any primary	Unmatched	0.697	0.691	1.4		0.803
	Matched	0.697	0.694	0.8	44.4	0.908
Midwife available?	Unmatched	0.751	0.641	24.1		0.000
	Matched	0.751	0.745	1.4	94.2	0.821
Price of rice	Unmatched	10.552	10.532	3.1		0.572
	Matched	10.552	10.542	1.6	48.9	0.810
Price of wheat flour	Unmatched	9.107	9.094	1.6		0.763
	Matched	9.107	9.084	2.8	-73.7	0.671
Price of mustard oil	Unmatched	53.026	53.771	-17.4		0.001
	Matched	53.026	53.279	-5.9	66.0	0.390
Average female wage	Unmatched	17.584	18.199	-9.2		0.097
	Matched	17.584	17.523	0.9	90.0	0.887
Average male wage	Unmatched	35.998	36.906	-13.2		0.015
	Matched	35.998	36.108	-1.6	87.9	0.806

Source: Authors calculations.

The matching process leads to a balancing[26] 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 and

[26] 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).

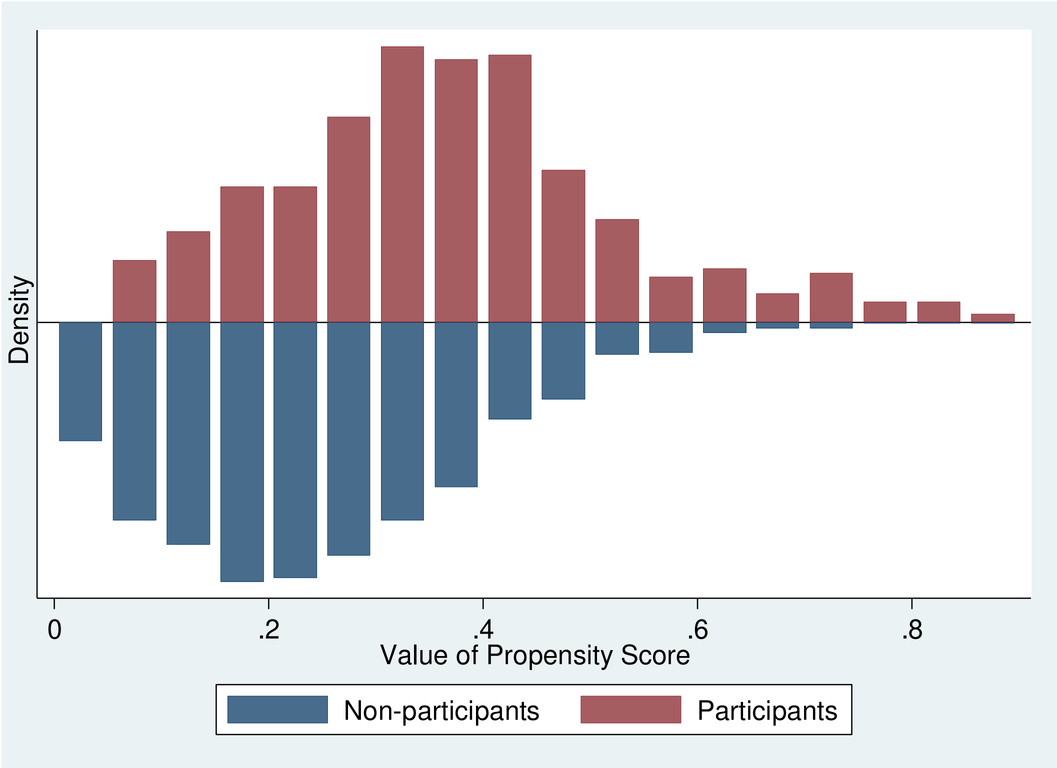
presents the mean values for treated and controls before and after the matching process. There are clear differences in the mean values among treated and controls before and after matching but the results in Table 2 indicate a reduction of bias for most variables that were significant in the logit model outlined in Table 1. In other words, the matching process has led to a sample with more balanced covariates between treatment and control groups, in some cases reducing bias by more than 90%. One way to assess the suitability of the matched data set for estimating impact is to look at the distribution of propensity scores to examine the degree of overlap; Figure 2 displays the propensity scores of de facto eligible women (currently married 14-50 year old) and the matched control sample including non-MFI eligible women[27] from both treatment and control villages and de facto eligible persons in treatment villages[28]. 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].

[27] Some households have both male and female borrowers while others have either a male or a female MFI borrower, or none. 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 be common.

[28] As noted by Morduch (1998), we cannot know who would be de facto eligible in control villages.

[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).

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. **Error! Reference source not found.**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 two different matching algorithms, i.e. nearest neighbour matching and kernel matching[31].

---

[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.

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	0.071**	0.056**
Contraceptive use by currently married women aged 14-50	0.135***	0.115***
Contraceptive use by currently married women aged 30-50	0.084***	0.075***
Any child born in last 4 years to currently married women aged 14-30 (yes=1; no=0)	0.028	0.020
Any child born in last 4 years to currently married women aged 14-50 (yes=1; no=0)	0.024	0.019

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 is used. Standard errors (not reported) are bootstrapped.

The 1-nearest neighbour estimate of the impact of MF borrowing on the probability of contraceptive use is 0.071 (Table 3) indicating that MF participants (pooling across gender of borrowers) aged 14-30 are 7.1% significantly (at 5%) more likely to use contraceptives than matched non-participants. The kernel matching estimate indicates a 5.6% higher level of contraceptive use for participants than for matched non-participants at a 5% 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 1% and vary between 7.5% to 13.5%[34]. The results for fertility variables for both age groups are negative 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

[32] As mentioned in the previous footnote, 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.

[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.

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**[35]
Contraceptive use by currently married women aged 14-30	Women	0.047	0.055**
	Men	-0.017	-0.007
Contraceptive use by currently married women aged 14-50	Women	0.094**	0.108***
	Men	-0.028	-0.018
Contraceptive use by currently married women aged 30-50	Women	0.067**	0.069***
	Men	-0.017	-0.023
Any child born in last 4 years to currently married women aged 14-30 (yes=1; no=0)	Women	0.036	0.030
	Men	-0.034	0.011
Any child born in last 4 years to currently married women aged 14-50 (yes=1; no=0)	Women	0.034	0.036
	Men	0.024	0.047

Source: Authors calculations.

Notes: *statistically significant at 10%, **statistically significant at 5%, ***statistically significant at 1%. STATA routine psmatch2[36] using the logit model outlined in Table 1 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, but more statistically significant for the older 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 we find that female borrowing does have significantly positive effects on contraceptive use and that male borrowing has largely positive but insignificant effects on fertility. However, we concur with PKML's finding that male borrowing has no effects on contraceptive use across all ages and that female borrowing has mainly positive but insignificant effects across both fertility outcome variables.

[35] 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.

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

Overall, our findings suggest that the estimates for males and females separately are not particularly different from the combined estimates presented in Table 3 for fertility outcomes, i.e. the estimates are mainly positive but statistically insignificant across age groups. However, there are positive and significant effects of MFI membership on contraceptive use when females borrow while male borrowing has no effects and across gender these estimates are mainly positive and statistically significant.

It appears that our PSM results robustly[37] support the findings of the general literature on MF and contraceptive use but not necessarily on fertility and we can only partly support the PKML findings. Thus, using the same data but different methods we obtain somewhat different results; but does this allow us to reach any strong conclusions as to the impact of MF on contraceptive use and fertility? In the next section we subject our results to sensitivity analysis shedding light on this issue by examining the robustness of our matching estimates.

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 (Rosenbaum, 2002). Ichino, Mealli and Nannicini (2006) argue that 'sensitivity analysis should always accompany the presentation of matching estimates' (19).

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.

This approach can be implemented using the **mhbounds** procedure in STATA (Becker and Caliendo, 2007); this procedure is suitable for binary outcome variables and uses the matching estimates to calculate the lower and upper bounds of the outcome

[37] As mentioned before, robustness checks were conducted. E.g. pscore was applied across the various matching algorithms. The various results obtained did not vary dramatically which suggests some degree of robustness.

variable for different values of Γ. If the lowest Γ is relatively small (say < 2) then one may assert that the likelihood of an unobserved characteristic is relatively high and therefore that the estimated impact is rather sensitive to the existence of unobservables (DiPrete and Gangl, 2004).

Sensitivity analysis can be illustrated by calculating the Γ at which the estimated impact of microfinance participation on the outcome variables presented here is no longer statistically significant. E.g. Table 3 shows that the kernel matching impact estimate with a bandwidth of 0.05 for contraceptive use for the ages 14-30 is 0.056 which is statistically significant at 5%. However, this may not be due to membership *per se* but to unobserved characteristics that account for membership (and or its impact). Sensitivity analysis explores the vulnerability of this impact estimate to selection on unobservables.

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, i.e. for a Γ of 1.05; the result for contraceptive use aged 14-30 already becomes insignificant at 5% and for a Γ of 1.15 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. Consequently, we suggest, the observed impact of microfinance membership on contraceptive use may well be confounded by one or more unobserved variables associated with both MFI borrowing and this impact and caution is required interpreting these results.

Table 5: Sensitivity analysis for contraceptive use ages 14-30 for microfinance participants

	Mantel-Haenszel bounds		**Significance level**	
Gamma (Γ)	Minimum	Maximum	Minimum	Maximum
1	1.852	1.852	0.032	0.032
1.05	1.579	2.128	0.017	0.057
1.1	1.317	2.391	0.008	0.094
1.15	1.068	2.643	0.004	0.143
1.2	0.830	2.884	0.002	0.203
1.25	0.601	3.117	0.001	0.274
1.3	0.382	3.341	0.000	0.351

Source: Authors calculations.

Similar observations can be made when looking at contraceptive use for the ages 14-50; for a Γ of 1.3 the result become insignificant at 5% and for a Γ of 1.35 they are no longer significant at 10%. For contraceptive use for the ages 30-50, the result becomes insignificant at 5% when Γ is 1.25 and insignificant at 10% when Γ is 1.35. Similar observations can be made for the fertility outcome variables[38].

Further research

To enrich this paper we initially wanted to further investigate the claims made by Steele et al (2001) that membership is a better indicator for MF participation than amount borrowed. However, whether we use MF membership or amount borrowed as a treatment variable does not make much of a difference in our case since all MF borrowers are also members. In the case of Steele et al (2001) there is a slight discrepancy in this regard, they report more members than borrowers. But given a bit more time we will investigate this claim further and try various approaches in STATA to match with continuous treatment variables.

Furthermore, we would like to investigate the various matching methods in more depth. King et al (2011) argue that PSM is unreliable and Mahalanobis matching or coarsened exact matching (CEM) generally outperform PSM and should be preferred. It

[38] The sensitivity analysis results for the remaining outcome variables can be obtained from the authors upon request.

is worth investigating this claim further since PSM is frequently heralded as a magic bullet but often not examined critically enough.

Also, we would like to analyse the data as a panel and examine the effects of MF on contraceptive use and fertility over time using the follow-up round of the PnK data collected in 1998/99. In addition, we obtained the empowerment data set used by Pitt et al (2006) and would like to add this into the panel to further investigate the link between women's empowerment, family planning and MF.

Conclusion

The literature suggests that MF has positive impacts on contraceptive use and negative impacts of fertility. The study by PKML 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 that controls 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.

The findings of PKML are interesting and challenging given the difficulties of replicating PnK's data and the questionable estimation strategy that has been the subject of ongoing controversy. We replicate the PKML variables with some difficulty and triangulate our results successfully with RnM. When we apply PSM to the data 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 across gender. We find few differences when the gender of the borrower is taken into account. Hence, our PSM results confirm the findings of the broader MF literature on contraceptive use but not on fertility, and we can only partially confirm the PKML findings with (our variable constructions of) their data (the remaining differences in our data reconstruction compared to PKML are outlined in Appendix 3). Sensitivity analysis has shown that the PSM estimates presented here are highly vulnerable to selection on unobservables and so we cannot be confident about causality between MF membership and FP outcomes.

The evidence of the impact of MF on contraceptive use and fertility remains contradictory and unreliable. One set of data subjected to different estimation methods leads to different results. This raises doubts about the econometric techniques employed and we concur with Leamer (1983) who criticises the key assumptions many econometric methods are built on and complains about "the whimsical character of econometric inference" (p. 38). We can only conclude that the evidence of MF impact on contraceptive use and fertility presented in this paper is weak, i.e. vulnerability of estimates to selection on unobservables and partially contradictory to PKML findings, this also implies weaknesses in the underlying data and the research design and the inability of advanced econometric methods to compensate for unsatisfactory data and thus the evidence presented in this paper does not allow us to draw any strong conclusions.

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. & 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.

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., 2009. Microcredit, Family Planning Programs and Contraceptive Behavior: Evidence from a Field Experiment in Ethiopia. *Available at: http://ipl.econ.duke.edu/bread/papers/working/247.pdf.*

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., 2011. 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? Forthcoming draft report for DFID, The School of International Development, University of East Anglia, UK (available on request from authors).

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.

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.

Goetz, A. M. & Sen Gupta, R., 1996. Who Takes the Credit? Gender, Power, and Control Over Loan Use in Rural Credit Programs in Bangladesh. *World Development,* 24 (1), p.45-63.

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.

Kabeer, N., 2001. Conflicts over Credit: RE-evaluating the Empowerment Potential of Loans to Women in Rural Bangladesh, *World Development*, 29(1), p.63-84.

Kabeer, N., 2005. Is Microfinance, the 'Magic Bullet' for Women's Empowerment: Analysis of Findings from South Asia. *Economic and Political Weekly*, 29/20/2005, p.4709-4718.

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. & Latif, M. A., 1996. The Role of Family Planning and Targeted Credit Programs in Demographic Change in Bangladesh. World Bank Discussion Papers, 337.

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.*

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

Latif, M. A., 1994. Programme Impact on Current Contraception in Bangladesh. *Bangladesh Development Studies*, 22 (1), p.27-61.

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.

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.

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.

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.*

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.

Appendix 1: Weighted means and standard deviations

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	1,841	29.65	9.14
Age of household head (years)	1,757	40.82	12.80	1,841	41.78	12.24
Highest grade completed by HH head	1,757	2.49	3.50	1,841	2.47	3.44
Highest grade completed by any female HH member	1,757	1.61	2.85	1,841	1.65	2.95
Highest grade completed by any male HH member	1,757	3.08	3.80	1,841	3.28	3.96
Sex of household head (male=1)	1,757	0.95	0.22	1,841	1.01	0.12
Household land (decimals)	1,757	76.14	108.54	1,841	100.16	337.33
Parents of HH head own land?	1,725	0.26	0.56	1,841	0.27	0.59
Brothers of HH head own land?	1,725	0.82	1.31	1,841	0.68	1.20
Sisters of HH head own land?	1,725	0.76	1.21	1,841	0.71	1.16
Parents of HH head's spouse own land?	1,735	0.53	0.78	1,841	0.56	0.80
Brothers of HH head's spouse own land?	1,735	0.92	1.43	1,841	0.95	1.46
Sisters of HH head's spouse own land?	1,735	0.75	1.20	1,841	0.79	1.24
No spouse in HH	1,757	0.13	0.33	1,841	0.03	0.16
Nontarget HH	1,757	0.30	0.46	1,841	0.14	0.01
Has any primary school?	1,757	0.69	0.46	1,841	0.69	0.46
Has rural health center?	1,757	0.30	0.46	1,841	0.06	0.24
Has family planning center?	1,757	0.10	0.30	1,841	0.09	0.29
Is dai/midwife available?	1,757	0.67	0.47	1,841	0.68	0.47
Price of rice	1,757	11.15	0.85	1,841	10.54	0.63
Price of wheat flour	1,757	9.59	1.00	1,841	9.09	0.77
Price of mustard oil	1,757	52.65	5.96	1,841	53.65	4.19
Price of hen egg	1,757	2.46	1.81	1,841	2.35	0.68
Price of milk	1,757	12.54	3.04	1,841	12.27	2.49

Price of potato	1,757	3.74	1.60	1,841	6.94	0.93
Average female wage	1,757	16.15	9.61	1,841	17.97	6.73
Dummy variable for no female wage	1,757	0.19	0.40	1,841	0.02	0.16
Average male wage	1,757	37.89	9.40	1,841	36.66	6.93
Distance to bank (km)	1,757	3.49	2.85	1,841	3.50	2.90
Amount borrowed by female from BRAC (Taka)	183	4,678.41	3,561.60	190	4,908.72	3,817.90
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	124	3,943.59	2,146.84
Amount borrowed by male from BRDB (Taka)	180	5,996.86	6,202.16	202	5,943.38	5,886.96
Amount borrowed by female from GB (Taka)	233	14,123.59	9,302.40	242	15,622.64	9,754.91
Amount borrowed by male from GB (Taka)	85	16,468.14	10,580.00	93	18,017.39	10,976.07
Outcome variables[2]						
Contraceptive use by currently married women aged 14-30	1,058	0.398	0.490	1,142	0.377	0.485
Contraceptive use by currently married women aged 14-50	1,731	0.378	0.485	1,841	0.379	0.485
Contraceptive use by currently married women aged 14-50	n/a	n/a	n/a	1,841	0.179	0.383
Any child born in last 4 years to currently married women aged 14-30 (yes=1; no=0)	1,056	0.697	0.460	1,142	0.669	0.471
Any child born in last 4 years to currently married women aged 14-50 (yes=1; no=0)	1,729	0.553	0.497	1,841	0.532	0.499

Notes:

1. Source: PKML, table 2, p. 10.
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 1,841 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).

Appendix 2: Headline findings of PKML

Outcome variable	Headline findings
Contraceptive use 14-50 (Table 5)	probit/wesml-probit/wesml-liml/wesml-fe/**wesml-liml-fe**
Female	
BRAC	+ive (ns)/+ive (ns)/+ive (sig)/+ive (ns)/**-ive (sig)**
BRDB	-ive (ns)/-ive (ns)/+ive (ns)/-ive (ns)/**-ive (sig)**
GB	+ive (sig)/+ive (ns)/+ive (sig)/-ive (ns)/**-ive (sig)**
Male	
BRAC	+ive (ns)/+ive (ns)/-ive (ns)/+ive (ns)/**+ive (ns)**
BRDB	+ive (sig)/+ive (ns)/-ive (ns)/+ive (sig)/**+ive (ns)**
GB	-ive (sig)/-ive (sig)/-ive (ns)/-ive (ns)/**-ive (ns)**
Contraceptive use 14-30 (Table 5)	**wesml-liml-fe**
Female	
BRAC	**-ive (sig)**
BRDB	**-ive (sig)**
GB	**-ive (sig)**
Male	
BRAC	**+ive (ns)**
BRDB	**+ive (ns)**
GB	**-ive (ns)**
Fertility 14-50 (Table 7)	probit/wesml-probit/wesml-liml/wesml-fe/**wesml-liml-fe**
Female	
BRAC	+ive (ns)/-ive (ns)/+ive (ns)/-ive (ns)/**+ive (sig)**
BRDB	-ive (ns)/-ive (ns)/+ive (ns)/-ive (ns)/**+ive (ns)**
GB	-ive (sig)/-ive (sig)/-ive (ns)/ +ive (sig)/**-ive (ns)**
Male	
BRAC	-ive (ns)/-ive (ns)/+ive (ns)/-ive (ns)/**+ive (ns)**
BRDB	+ive (ns)/+ive (ns)/-ive (ns)/-ive (ns)/**-ive (sig)**
GB	+ive (ns)/+ive (ns)/-ive (ns)/-ive (ns)/**-ive (sig)**
Fertility 14-30 (Table 7)	**wesml-liml-fe**
Female	
BRAC	**+ive (ns)**
BRDB	**+ive (ns)**
GB	**-ive (sig)**
Male	
BRAC	**+ive (ns)**
BRDB	**-ive (sig)**
GB	**-ive (sig)**

Source: PKML.

Note: The results in bold obtained from WESML-LIML-FE are according to PKML the most reliable since they control for all major sources of bias (p. 14).

Appendix 3: Main differences between RnM and authors' findings for round 1- 3:

Variables	RnM variable names	Authors variable names	Explanation, R 1-3	RnM SQL File
Non-land assets	nlasset fnlasset	nonlandasset nonlandwomen	We used the same variables as RnM – see their SQL file. RnMs average values for both variables are lower than ours, though this is misleading since RnM do not have any data for round 3 as a round by round comparison shows. RnMs respective round 1 and 2 values for nlasset are in fact higher than ours and the opposite applies for their fnlasset values. We follow our interpretation.	dbo. HHassets
Landed assets	flandvala flandvalb landbef landaft landvala landvalb	landawomenval landbwomenval halab halaa halaa_tot halab_tot	We used the same variables as RnM – see their SQL file. There is no difference in landaft and halaa. The remaining variables still have a few discrepancies. RnM assumes that landbef is equal to halaa if halab is equal to 0, hence the differences. We follow our interpretation.	dbo.individual land dbo. HH land
Eligibility	eligible eligbrac eligbrdb eliggram q r bracvill brdbvill gramvill villprog	elig_defacto elig_defacto_brac elig_defacto_brdb elig_defacto_gb elig_dejure_brac elig_dejure_brdb elig_dejure_gb elig_defacto_treat/elig_defacto elig_dejure_treat/elig_dejure bracvill_MD brdbvill_MD gbvill_MD vill_prog_rpj	Differences for 183 hh " " 79 hh " " 24 hh " " 61 hh " " 30hh " " 29hh " " 42hh " " 1386 hh / 569 hh " " 1245 hh / 1599 hh Spot on. Spot on. Spot on. Spot on. A few differences, mainly because RnM assume landbef is equal to halaa if halab is equal to 0. This does not seem justifiable and we follow our own interpretation.	dbo. HH program status