Is Replication Worthless Work? Sub-group Heterogeneity among Microfinance and Informal Sector Borrowers in Bangladesh

Maren Duvendack[1, 2]

Richard Palmer-Jones[1]

EADI/DSA

19-22 September 2011

1. UEA, 2. ODI

Introduction : Replication

- Replication is important to assess the robustness of findings
  - Part of quality assurance/peer review process
- Necessary part of scientific process
- The robustness of findings may be weaker than they appear
  - Not guaranteed by peer review process
  - Nor does peer review guarantee authorial probity and integrity
  - Nor is robustness guaranteed by academic status of authors

Replication continued ….

- Many good reasons to conduct replications
  - To fully understand what the authors have done
  - To check for errors
    - Of data cleaning and variable construction
    - Of estimation processes (specification or data mining….)
  - To use other operationalisation of concepts into variables to see if this makes a difference to findings
  - To apply different estimation methods including different software
  - To check against different comparable data sets
    - Including different but relevant contexts for external validity

Challenges of Replication

- Problems in replication may arise for a number of reasons
  - Original authors not just interested – or (passively) resistant
  - Question why should you be interested in replication
- Desirable to have good documentation and access of study design and data – questionnaires, code books...
  - To communicate or not to communicate with original authors
    - Desirable to be helpful but non-directive
- Reproduction vs replication:
  - Reproduction: different methods applied to the same raw data
  - Replication: same (or very similar) methods to the same raw data

Introduction: Microfinance Evaluations

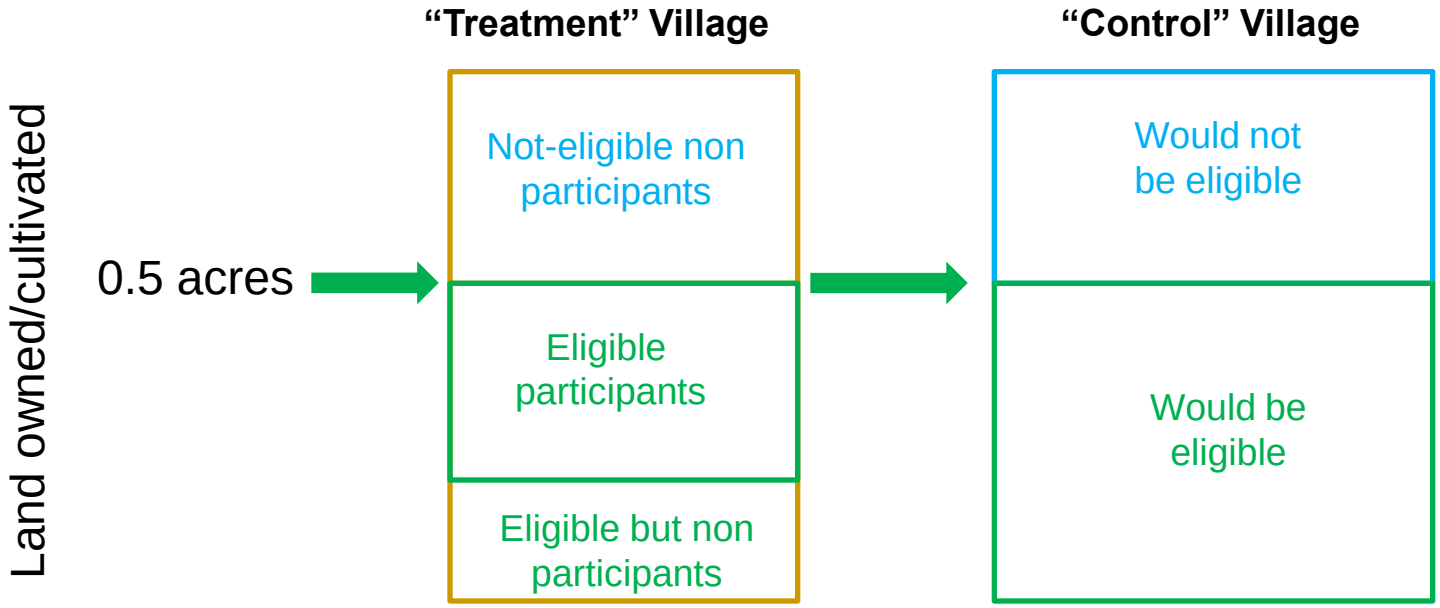
- Microfinance an important strategy in the fight against poverty
  - Financial and non-financial services to low-income households
  - Key features of microfinance: Group liability (initially) & targeting of women
- Studies suggesting social and economic benefits:
  - Hulme and Mosley (1996), Coleman (1999), Pitt and Khandker (1998), Khandker (1998 and 2005), Rutherford (2001) and Morduch and Haley (2002)
- Critical voices:
  - Roodman and Morduch (2009), Bateman (2010) and Dichter and Harper (2007), Roy, 2010)
  - First two RCTs in the sector (Banerjee et al, 2009; Karlan and Zinman, 2009) raising doubts about the causal link between MF and poverty alleviation.
- Most influential MF IE to date: Pitt and Khandker (1998)

Importance of Pitt and Khandker

- Methodologically innovative
  - Large original World Bank survey in 1991-2
    - With follow up panel in 1998-9
  - Complex and sophisticated analysis (WESML-LIML)
  - Most rigorous impact evaluation of microfinance
- Key work of main academic author(s)
  - Key paper in top economics journal & 2 books
  - More than 21 papers mainly in peer reviewed journals sole or co-authored with collaborators and or PhD students
  - Second most highly cited paper in Pitt's career

Introduction to Pitt and Khandker (1998)

- Iconic study finding positive impacts of MF especially when lending to women (male and female groups)
- Quasi-experimental design & eligibility condition used to identify impact
  - Primary eligibility criterion: landownership (0.5 acres = 50 decimals)
  - Overall sample:1,798; 1,538 households from treatment villages, 260 from controls



Outline of Replication Challenges

Overview of methodological approaches of studies replicating PnK:

Study	Method
PnK 1998	Weighted Exogenous Sampling Maximum Likelihood-Limited Information Maximum Likelihood-Fixed Effects (WESML-LIML-FE)
Morduch 1998	Differences-in-Differences (DiD)
Pitt 1999	Expansion of PnK model and comparison to Morduch using a simulation-based approach
Chemin 2008	Propensity Score Matching (PSM)
RnM 2009	2-stage (Weighted, Truncated) Maximum Likelihood Estimation using cmp[1]
Duvendack 2010, Duvendack & Palmer-Jones, 2011	PSM, DID, cmp, & Panel Random Effects

[1]cmp developed by Roodman (2009) is a STATA routine that contains a range of STATA commands that calculate recursive mixed process estimators, i.e. multi-equation models with different types of dependent variables which can also appear on the right hand side of any of the other equations in the model. Uses STATA routines for probit & tobit, and user written (Roodman) xtabond2

Morduch 1998 & Pitt 1999

- Morduch, 1998, challenges PnK
  - Reconstructs data – some mistakes
  - Difference in Difference regression estimate
- Pitt, 1999, dismisses Morduch
  - More heavy duty econometrics
  - “misunderstands” ...
- Chemin applies PSM but does not engage in controversy between Morduch and Pitt
- Pitt,1999, effectively puts an end to criticism until Roodman & Morduch and Duvendack & Palmer-Jones

Roodman & Morduch

- Reconstruct data
  - Lot of dialogue with Pitt and Khandker
    - Get additional data, and some clarification
    - Make data and variable construction publicly available
  - Largely agrees with our data set reconstruction
- Re-estimated main results
  - PnK, 1998, Morduch, 1998, Pitt 1999, & Khandker 2005
  - Initially could not obtain same results but later did reproduce main results.

RnM conclude that “decisive statistical evidence in favor of them [the ideas that microcredit is effective in reducing poverty and especially so when women do the borrowing] is absent from these studies.” Get same statistical results but doubt that it establishes causality.

- Leaves Chemin as only credible critique of PnK

Duvendack

- Reconstructs PnK data & replicates analysis
  - Beaten to it by Roodman – but makes data triangulation possible
  - Duvendack and Roodman data largely agrees
  - Replicates and agrees with IV analysis (using cmp for WESML-LIML)
  - Reconstructs Chemin data and applies PSM
    - Cannot replicate Chemin data or results
- **Conducts sub-group analysis**
  - Shows impact of MFI no different to other borrowings
- Applies sensitivity analysis to PSM
  - Finds PSM results highly vulnerable to unobservables

Sub-groups & Microfinance

- PnK imply ignore alternative sources of finance
  - Should have been in the design because one should test against next best alternative
  - Research design neglected alternatives
  - But other sources of borrowing appear in their data
  - Little analysed and not vs MF

Sub-Group Comparisons

- Lack of sub-group heterogeneity undermines claim that MFIs make unique contribution to poverty reduction.

1. Y^{MF} versus Y^{None} - MF participants versus all other individuals across treatment and control villages that do not have other sources of borrowing.

2. $Y^{MF} + Y^{Multiple} + Y^{Borr}$ versus Y^{None} - individuals that participate in either MF, MF and other non-MF borrowing and other non-MF borrowing only versus all other individuals across treatment and control villages that do not have other sources of borrowing.

3. Y^{MF} versus Y^{Borr} - MF participants versus individuals that have other non-MF borrowing across treatment and control villages.

PSM Results

Outcome variables	Y^{MF} vs Y^{None}	Y^{MF}+ $Y^{Multiple}$+Y^{Borr} vs Y^{None}	Y^{MF} vs Y^{Borr}
Comparison	**1**	**2**	**3**
	Kernel matching, 0.05		
Log per capita expenditure (Taka)	-0.011	0.019	-0.089**
Log women non-landed assets (Taka)	0.498***	0.349**	-0.022
Girl school enrolment, aged 5-17 years	0.060*	0.061**	0.077
Boy school enrolment, aged 5-17 years	0.035	0.060**	-0.011

- Results by gender: No obvious advantage of female borrowing
- Sensitivity analysis suggests that impact estimates of all comparisons are sensitive to selection on unobservables.
- Summary: PnK overestimated the impact of microcredit participation

Problems of Replication

- Gaining access to (all) data
  - Clarification of raw variables
  - Construction of aggregate variables
- Final data used by PnK not available
  - Highly controlled – but (effectively) re-appears in 2011
- Implementation of econometric analysis
  - PnK use custom built software
  - Roodman claims replication with cmp
- Difference in results between Duvendack and Chemin could be either in data construction or analytical method
  - Unresolvable dispute due to lack of Chemin's data constructions. (Further replications will not help)

Conclusions

- Replication is highly desirable
  - 15 wasted years believing beyond reason in MF → ***no clear evidence*** for MF which is supported by systematic reviews
  - Award prestige only if public deposit of original data and code;
  - and test using alternative data construction, software, etc.
  - Repeat studies in different locations, data sets, etc.
- Poor research design with low quality data cannot be compensated by sophisticated econometric techniques
  - Doubts about capacity of PSM to solve attribution problem:
    - Sensitivity analysis indicates ***vulnerability to unobservables***
- Challenges to status of e**con**om(ics)etrics
- Continuing need for high quality observational studies with ethical reporting and publication practices (enabling replication)

Q & A Session

For further questions or comments please email:

m.duvendack@odi.org.uk; or

r.palmer-jones@uea.ac.uk

Appendix: PSM Results – By Gender

Outcome variables		Y^{MF} vs Y^{None}	$Y^{MF+Multiple}$ vs Y^{None}	Y^{MF} vs Y^{Borr}
Comparison		**1**	**2**	**3**
	Kernel matching, 0.05			
Log per capita expenditure (Taka) 1 2	Women	-0.013	0.012	-0.126*
	Men	-0.046**	0.015	-0.079**
Log women non-landed assets 3 4	Women	0.754***	0.561***	-0.848
	Men	-0.000	0.244	0.249
Girl enrolment, aged 5-17 years 5 6	Women	0.067*	0.061**	-0.216
	Men	0.032	0.060**	0.133
Boy enrolment, aged 5-17 years 7 8	Women	0.045	0.050	-0.128
	Men	-0.001	0.054**	-0.017

Source: Authors calculations.

Notes: *statistically significant at 10%, **statistically significant at 5%, ***statistically significant at 1%. Results refer to the differences in the mean values between matched samples. t-tests before and after matching employed to investigate the differences in the mean values for each covariate X across matched samples; the test provided conclusive results.

Appendix: Sensitivity Analysis

- PSM estimate for log of **women non-landed assets** for Y^{MF}: 0.498*** (comparison 1) - sensitive to selection on unobservables?

	Significance levels		**Hodges-Lehmann point estimates**		**95% Confidence intervals**	
Gamma (Γ)	Minimum	Maximum	Minimum	Maximum	Minimum	Maximum
1	< 0.0001	< 0.0001	0.886	0.886	0.315	1.317
1.2	< 0.0001	< 0.0867	0.465	1.218	-0.245	1.570
1.3	< 0.0001	< 0.2329	0.274	1.341	-0.532	1.694
1.4	< 0.0001	< 0.4422	0.065	1.439	-0.710	1.796
1.5	< 0.0001	< 0.6547	-0.159	1.533	-0.886	1.891

Source: Authors calculations.