The Journal of Development Studies

ISSN: 0022-0388 (Print) 1743-9140 (Online) Journal homepage: https://www.tandfonline.com/loi/fjds20

Response to 'High Noon for Microfinance Impact Evaluations'

Matthieu Chemin

To cite this article: Matthieu Chemin (2012) Response to 'High Noon for Microfinance Impact Evaluations', The Journal of Development Studies, 48:12, 1881-1885, DOI: 10.1080/00220388.2012.727561

To link to this article: https://doi.org/10.1080/00220388.2012.727561

Journal of Development Studies,
Vol. 48, No. 12, 1881–1885, December 2012

Routledge
Taylor & Francis Group

Response to 'High Noon for Microfinance Impact Evaluations'

MATTHIEU CHEMIN
McGill University, Montreal, Canada

Final version received August 2012

ABSTRACT *Duvendack and Palmer-Jones are critical of analysis and conclusions in Chemin (2008) because they are unable to replicate my results. This response identifies key differences between the two papers, especially regarding the sample and measurement of variables, which imply that Duvendack and Paler-Jones should not be considered as either a replication or a criticism of my work.*

Duvendack and Palmer-Jones (2012), hereafter DPJ, cast doubt on results in Chemin (2008) because they obtain different results and experienced problems in replicating my data constructions.[1] This response identifies 25 differences between DPJ and Chemin (2008) that account for our dissimilar findings; the fundamental problem with DPJ lies in the fact that they did not use the same sample data or Stata do code as in Chemin (2008), a problem that could easily have been avoided had they simply asked. To avoid confusion or misunderstanding, it should be noted that in response to requests[2] in 2009 and later I provided Duvendack and Palmer Jones with data and Stata files related to my paper. At that time there was no suggestion that replication for publication purposes was the aim or that they were encountering problems in replication so although not all files were provided I was left with the impression that the files required for their purposes had been provided.[3] Differences in the sample selected and the way in which certain variables are measured give rise to inconsistencies and errors in their analysis. Consequently, DPJ should not be considered as either a replication or a criticism of my work.

The response details the differences between the two papers in four key areas: 1) the sample used; 2) the construction of the variables; 3) the propensity score estimation and matching techniques; and 4) the datasets. I conclude with some comments on future directions for replication studies. Note that differentiating between male and female borrowers is a contribution by DPJ that expands upon my original work in an insightful way that can further advance policy in the microfinance field.

I. Differences in the Sample Used

As microfinance in the sample was limited to individuals owning less than 0.5 acres of land, I restricted my sample to only those individuals owning less than 0.5 acres of land (as described in

Correspondence Address: Matthieu Chemin, Department of Economics, Room 419, Leacock Building, 855 Sherbrooke Street West, Montreal, Quebec, H3A 2T7, Canada. Email: matthieu.chemin@mcgill.ca

ISSN 0022-0388 Print/1743-9140 Online/12/121881-05
http://dx.doi.org/10.1080/00220388.2012.727561

Chemin, 2008: 472). As a result, my sample included less than the total 9397 observations (only 5037 observations). However, in their propensity score specification, DPJ (Table 1) do not restrict their sample to include only those who are benefiting from microfinance but include individuals who are both benefitting from and technically excluded from microfinance, that is the entire population of 9397 observations. Interestingly, when performing matching analysis, DPJ do restrict their sample to individuals with less than 0.5 acres of land (the propensity scores corresponding to this sample selection could be compared to my results, but this was not done). Authors should be consistent with sample selection throughout the entire analyses.

Chemin (2008) omitted control villages (without access to microfinance) from the sample used as otherwise there would be perfect multicollinearity between the set of village dummies for control villages and the participation decision of individuals in those villages (always zero by definition). As the control village dummies perfectly predict the participation outcome, these observations are dropped from the propensity score estimation. DPJ do not omit control villages as they include the full sample, but then exclude the control village dummies from their propensity score estimation. Their propensity score is thus not estimated correctly as village dummies have to be included for control as well as treated villages. These differences (restricting the sample to individuals under 0.5 acres of land, and living in treated villages) explain why the sample size in Chemin (2008) is lower than in DPJ and probably also explains significant differences in results.

II. Differences in the Definition of the Variables

There are a number of key differences in the definitions or measures of the dependent and independent variables used that contribute to the different findings of the two papers.[4]

- In Chemin (2008), the dependent variable (participation in microfinance) was determined by the take-up of any loans through BRAC, BRDB, or Grameen Bank in round 1. Moreover, I oversampled the participants by including individuals that participated in Rounds 2 and 3, but not Round 1, of the dataset. DPJ follow the same approach but exclude participants with more than 0.5 acres. Therefore, when they do not restrict the sample to people under 0.5 acres of land, the control group contains individuals who participate in microfinance. This is obviously the wrong control group since microfinance participants are then compared to other participants. This could significantly bias downward any positive impact of microfinance. When they restrict their sample to people under 0.5 acres, we have the same variable.
- I used the variable *flopt* (Total Operational Land) to measure land owned. However, DPJ at times use *flopt* but at other times use the variable *halaa* (Landed Assets: After); this is both inconsistent and raises replication difficulties.
- Whereas I used a measure of the exact age (age in years plus age in months divided by 12) to represent the age of individuals, DPJ use years only (except in using exact age when calculating the age of the household head).
- Regarding gender, I followed Pitt and Khandker (1998) and calculated a dichotomous variable equal to 1 if the household does not contain any adult males (no adult male in household) and 0 otherwise. DPJ calculate a variable representing the total number of adult males in the household (number adult male in household). These two entirely different concepts will generate different (non-replication) results.
- In Chemin (2008) a variable representing education level was computed from the highest grade completed for the individual who was taking up the microfinance loan. In the preferred specification, a dichotomous variable equal to 1 was calculated if the individual has some schooling at all, or 0 if the individual has none. DPJ use the highest grade completed of the most educated person in the household.
- Minor differences are:[5] I used household level savings, while DPJ use individual savings; I used a dichotomous variable equal to 1 if the household owns a non-farming enterprise (0 otherwise), while DPJ use an individual-level dichotomous variable; to calculate agricultural

wages, I consider the agricultural permanent wage rate per day, and the seasonal/casual wage if a permanent wage is missing, whereas DPJ use the sum of permanent and seasonal/casual wage; to estimate non-agricultural wages, I considered only the primary occupation, while DPJ sum together the primary and secondary occupations.

III. Differences in the Propensity Score Estimation and Matching

DPJ omit many of the variables I used in the propensity score estimation,[6] including: the cube of age, agricultural income, the amount of land owned by the household, father's education, mother's education, marital status, and mother still alive. Moreover, I included 72 village dummies, while DPJ use only 24 *thana* (a larger administrative division) dummies. My use of additional control variables and more disaggregated geographical dummies contributed to more rigorous results, and it is unclear why DPJ did not choose to include these variables if the aim was replication.

Chemin (2008: 473) used a 'corrected' propensity score to estimate the propensity score for individuals in control villages (without access to microfinance): as they do not participate in microfinance by definition the dummies of control villages perfectly predict the participation decision. This perfect multicollinearity between explanatory variables and dependent variables causes Stata to drop the control villages from the analysis. However, one still needs to estimate the propensity score of individuals in control villages when comparing them to participants in treated villages. Chemin (2008) addresses this by first estimating the coefficients of the propensity score on the sample of villages with access to microfinance and then predicting the propensity score after equating all village dummies to 0 for individuals with access to microfinance. This essentially makes the propensity score estimation for individuals with and without access to microfinance comparable. DPJ choose to follow an entirely different procedure to address the same issue. Although they present in Table 1 a propensity score estimated on all villages (but without control village dummies, which will likely bias the results), they do not use this when performing matching. Instead they regress participation on control variables, with no village dummies, and restrict the sample to only one of the five control districts (*thana* 25). This procedure is incorrect for several reasons: omitting village dummies is likely to bias the results; there were five control districts, not only one; and the results of this propensity score are not presented. The difference in the two methodologies used is likely to explain the differences, and represents a serious flaw in their analysis.[7]

IV. Differences in the Two Datasets

The dataset used in Chemin (2008) was made available to me in 2003 by Mark Pitt, the co-author of the seminal paper (Pitt and Khandker, 1998). There are four differences between the dataset I used and the dataset that was later posted on the World Bank website:[8] in Rounds 1, 2, and 3, seven individuals have different education levels; only in Round 1, the amount of land owned is 0 in the original data, but missing in the World Bank dataset; in Rounds 1, 2, and 3, three households have a different age of the household head; in Rounds 2 and 3, the information pertaining to loans from microfinance institutions varies quite substantially. This does not affect the participation variable but affects the amount of land owned, since I use the three rounds to construct this variable. As the amount of land owned varies through time, the same individual may have different land owned in the original data than in the World Bank data.

V. Results

To reassure myself that the issues raised in DPJ were not a substantive criticism I replicated the original results of Chemin (2008).[9] The propensity score estimation is exactly the same as my original paper. In addition, the matching results are the same, but marginally differ when using the control villages as a control group, due to an improvement in my code. All results are similar to the original results, and even stronger, with a 7 per cent increase in expenditure for individuals participating in microfinance compared to individuals in control villages (not significantly different from previous results at 3%). These results support the main conclusion of a 'positive, but lower than previously thought, effect of microfinance' (Chemin, 2008: 463).

The results are also replicated with the World Bank data and are again very similar. Therefore, contrary to the DPJ claim that my results are not replicable, carefully following the procedure outlined in Chemin (2008) by using the same sample, variable construction, propensity score estimation, matching technique, and dataset, demonstrates that the results are robust.

VI. Concluding Comments

Replication studies are an important and integral part of research by ensuring that research findings that can influence theory and policy are both valid and reliable. However, replication research can also be considered a double-edged sword. On the one hand, employing rigorous methodology to replicate previous studies' findings can help protect the integrity of the research community, as well as help expand the field by generalising to other populations and improving upon the original paper's results. On the other hand, replication studies that do not follow the rigorous protocol required for replication can seriously undermine the work of the original researchers who are attempting to progress research in new directions. Thus, to ensure the continued integrity of replication research, authors who replicate previous papers need to ensure that the procedures pertaining to replication are rigorous, consistent, and done correctly. In order to do this, authors should be explicit about their intentions when asking for data and do-files, thus acquiring the needed cooperation from the original authors in order to complete their study. As noted above, when communicating with me DPJ did not clarify that they were engaged in replication for publication purposes; had they done so, confusion and errors could have been avoided.

On a final note, DPJ appear to criticise the matching technique in my paper as 'a sophisticated analytical method to compensate for weak research design' (Duvendack and Palmer-Jones, 2012: 1876). I completely agree that matching methods are imperfect, and have noted the specific and serious problem that matching is based 'on the assumption that selection is based on observables' (Chemin, 2008: 464). Unless one identifies a specific improvement in methodology, there is limited benefit in replicating these methods. Economists have developed other approaches with better research design and less identification assumptions than matching, such as randomised experiments, regression discontinuity designs, and difference-in-differences. I urge researchers of replication studies to focus on the replication of these techniques rather than matching.

Notes

1. A longer version of this response with more details and all relevant data and Stata files are available on my home page at http://matthieuchemin-research.mcgill.ca/.
2. These were provided for the purposes of practicing matching technique (for PhD student Duvendack) and organising a 'hands-on workshop' (Palmer-Jones).
3. The complete email correspondence is available on my website to let the interested reader draw their own conclusions.
4. These are surprising as DPJ could have replicated my variables from the file containing the construction of the variables that was among those provided (databaseR1.do).
5. The construction of the following variables is in a file (database.do) that DPJ did not request or access.
6. This is surprising as I had provided the code necessary for replication (microfinancefinal.do).
7. I did not use the Stata command psmatch2 as it did not exist in 2002, when I started working on my article, but developed my own matching programs (these are available online). In the online do-files, I replicate with psmatch2 and find very similar results.
8. Details on these differences are available on my website. To replicate my results with the World Bank data, one needs to download the World Bank data, and include the four amendments to make the datasets comparable. These four amendments are all incorporated in the folder 'final replication with World Bank data'.

9. The data and do-files are in the folders 'original data' and 'do files\final original data' on my website. To replicate my results with the World Bank data, one needs to download the World Bank data, but include the four amendments above to first make the datasets comparable. These four amendments are all incorporated in the folder 'final replication with World Bank data'. In addition, use 11AR2.DTA and 11AR3.DTA provided on my website since csbwol is different in Rounds 2 and 3. This dataset was originally provided to me by the co-author of the seminal study. The replication with the World Bank data is in the folder 'final World Bank'.

References

Chemin, M. (2008) The benefits and costs of microfinance: Evidence from Bangladesh. *Journal of Development Studies*, 44(4), pp. 463–484.

Duvendack, M. and Palmer-Jones, R. (2012) High noon for microfinance impact evaluations: Re-investigating the evidence from Bangladesh. *Journal of Development Studies*, 48(12), pp. 1864–1880.

Pitt, M. and Khandker, S. (1998) The impact of group-based credit programs on poor households in Bangladesh: Does the gender of participants matter? *Journal of Political Economy*, 106(5), pp. 958–996.