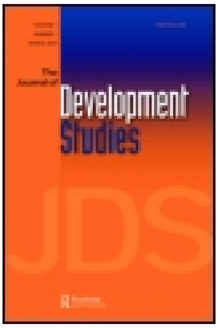


The Journal of Development Studies
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/fjds20

# Response to 'The Impact of Microcredit on the Poor in Bangladesh: Revisiting the Evidence'

Mark M. Pitt[a]

[a] Brown University, Providence, Rhode Island, USA

Published online: 14 Jan 2014.

*The Journal of Development Studies*, 2014
http://dx.doi.org/10.1080/00220388.2013.868141

Routledge
Taylor & Francis Group

# Response to 'The Impact of Microcredit on the Poor in Bangladesh: Revisiting the Evidence'

MARK M. PITT*
Brown University, Providence, Rhode Island, USA



'The Impact of Microcredit on the Poor in Bangladesh: Revisiting the Evidence,' by David Roodman and Jonathan Morduch (2009; henceforth RM) is the most recent of a sequence of papers and web postings that seeks to refute the findings of Pitt and Khandker (1998; henceforth PK) that microcredit for women had significant, favourable effects on household consumption and other outcomes. In this paper, RM have backed off many of their prior claims and methods after earlier replies noted their faults (Pitt 1999, 2011a, 2011b; Pitt and Khandker 2012). Nonetheless, RM make a number of criticisms of and claims against PK. The strong wording of RM may give the impression that the criticisms are equally strong and valid. This brief response shows that all of the criticisms are readily refuted. The principal claims are addressed in turn below; others are addressed in Pitt and Khandker (2012) and Pitt (2013) providing a more extensive discussion of all issues raised by RM.

## 1. The Econometrics of RM and PK

The essential nature of the model as formulated in PK is quite simple. Some individuals have the choice of participating in a credit programme (receiving treatment), and the choice to participate is exogenous conditional on a set of exogenous covariates that include village fixed-effects. Actual participation and the level of treatment (the amount borrowed) are endogenous. Using the RM notation but abstracting from gender-specific choice and other details, such as censoring and weighting, for brevity, the log-likelihood $L_{PK}$ that is maximised is the sum of the standard two-stage least squares (2SLS) likelihood ($L_{2SLS}$) and the standard ordinary least squares (OLS) regression likelihood ($L_{OLS}$).

$$\begin{aligned} L_{PK} =& L_{2SLS}\left(y_0 - y_{fm}\delta - \mathbf{x}\beta_0,\ y_{fm} - \mathbf{x}\beta_{fm}\right) \text{ for those with the choice to participate} \\ &+ L_{OLS}(y_0 - \mathbf{x}\beta_0) \text{ for those without the choice to participate} \end{aligned} \tag{1}$$

The log-likelihood $L_{PK}$ requires only one set of exogenous variables, **x**. There are no exclusion restrictions, and thus there is no separate set of instruments **z**. Identification in PK arises from having observations on the dependent variable $y_0$ and the exogenous variables **x** for individuals that exogenously have no choice to obtain (credit) treatment. It is the number of observations without choice of treatment that critically affects identification – at the limit, if there are too few such observations in the data, the effect of treatment (credit) $\delta$ is not identified (but this is not the case with the data). The two additive parts of $L_{PK}$ are log-likelihoods that, contrary to the claim of RM, are not complex; *Stata* very easily estimates both $L_{2SLS}$ and $L_{OLS}$ with a single simple command and can estimate the sum using Roodman's *cmp* command. That limitation in *Stata* does not make this model complex.

**Correspondence Address*: Mark M. Pitt, Department of Economics, Box B, Brown University, Providence, Rhode Island, 02912 USA.

The RM likelihood is

$$L_{RM} = L_{2SLS}\left(y_0 - y_{fm}\delta - \mathbf{x}\beta_0, y_{fm} - \mathbf{x}\beta_{fm} - \mathbf{z}\pi\right) \tag{2}$$

for everyone in the sample including those without choice, where **z** is a set of identifying instruments having parameters $\pi$. The **z** instruments, as RM explain, are interactions of **x** with a dummy variable for having the choice to participate. Both PK and RM estimate their models by maximising their respective likelihoods. Although the assumption of normality is made in constructing the OLS log-likelihood and the 2SLS likelihood, linear regression provides *exactly* the same parameter estimates as OLS by maximum likelihood ($L_{OLS}$) no matter what the distribution of the errors, and for $L_{2SLS}$, the assumption that the error terms are multivariate normal also need not be true to insure that the parameter estimates are consistent and asymptotically normal (Davidson & MacKinnon, 1993, p. 641).

There are two differences between the PK log-likelihood (Equation 1) and the RM log-likelihood (Equation 2). First, for those without credit programme choice, RM estimate the determinants of the outcome $y_0$ (household expenditure) as part of the 2SLS likelihood, even though credit is deterministically and uniformly zero for every observation so that ordinary least squares is unbiased and fully efficient. Since credit is deterministically zero (with a sample and theoretical variance of zero) for those without choice, these observations cannot provide any information to estimate the first-stage parameters. In contrast, PK treat the determinants of $y_0$ for those exogenously excluded from credit as exogenous – in the OLS part of the likelihood set out in Equation (1). Second, RM add instruments **z** based upon the exogeneity of choice. Given that they have treated exogenous observations as endogenous, they must do this in order for the impact of credit to be identified. In contrast, PK do not need to add instruments **z** because they treat exogenous observations as exogenous; the log-likelihood is much simpler than RM because it does not need the artifice of constructing instruments **z,** and because it uses OLS rather than 2SLS for the exogenous observations.

In what follows, it is useful to keep in mind that it is the specification of the first-stage equation that is the key difference between RM and PK, and that fully efficient estimates of the first-stage can be obtained by linear regression in both cases. Consequently, the comments below focus on the first-stage and the 'predicted' credit estimated from the first-stage in a model is which there is gender-specific treatment, and abstracts from issues of censored regression and sample weighting. These comments are not specific to the PK data only but apply in general to *any* dataset having multiple endogenous treatment variables in which some observations have been exogenously (conditional on **x**) denied the choice to be treated.

- The RM first-stage regression predicts non-zero credit with positive sample variance for those who have deterministically zero credit. This is akin to predicting a non-zero probability that an infant will have graduated university in a model of schooling achievement that includes all ages. In comparison, the PK first-stage 'predicts' zero credit with zero variance for those who are deterministically without choice.
- The RM first-stage will, in general, estimate different marginal effects than the PK first-stage for all of the exogenous variables, hence generates different predicted values for credit than PK even for observations with choice.
- If additional non-choice observations are added to the dataset, the first-stage estimates of RM will change as if these observations added new information even though observations with deterministically zero credit necessarily contain no information on the determinants of credit behaviour. The PK first-stage does not suffer from this anomaly.
- The first-stage equations of RM differ from those of PK not only because RM includes non-choice observations but also because they add the 'instruments' **z** – the interactions of choice with the exogenous variables **x**. If there are two (or more) endogenous variables requiring first-stage equations (as in PK), RM and PK estimates can be very different, as the RM approach adds many extra first-stage parameters compared to PK. This proliferation of instruments generates

weak instrument bias (Pitt 2012). For example, the reported RM first-stage for the case of three credit programme types times two sexes yields a data matrix of independent variables that is 58 times larger than in PK. In the case of male participation in the Grameen Bank, for example, the first-stage equation used by RM (used to estimate column (1) of their Table 5) contains a matrix of independent variables for which fully 87.5 per cent (931,872 out of 1,064,472) of the elements correspond to households in which males have (deterministically) no choice of participation in the Grameen Bank.

In addition, RM estimate their models using linear limited information maximum likelihood (LIML), which is particularly unstable compared to plain 2SLS when the first-stage inappropriately models the determinants of choice with observations that do not have choice. Large differences between standard 2SLS and linear LIML are a consequence of the questionable RM setup and not the PK dataset. Pitt and Khandker (2012) demonstrate that linear LIML generates outlandish estimates, that are completely unlike 2SLS estimates, with simulated data for the RM model. This instability is apparent in the estimates of credit effects (the $\delta$'s) that RM present using the PK data. For example, examination of the complete computer output for the estimates presented by RM in column (1) of their Table 5 shows that 105 out of 110 slope coefficients in the second-stage output as reported by Stata have a reported t-ratio of *exactly* 0.01 in absolute value, and a (centred) R-squared of –381.4. The linear LIML t-ratio for the female borrowing parameters is reported by RM to be t = 0.10 in column (2) of Table 5, as compared to the t-ratio of t = 10.63 when the t-ratios are re-computed using the asymptotically consistent 2SLS method. Based upon the female credit effect estimated by RM in this regression, one might wrongly conclude that the problem with Pitt and Khandker (1998) is that they *underestimate* the strongly statically significant and positive effect of women's credit on household consumption, as well as the t-ratio. The point estimate of the female credit effect found by RM is much larger than that found by PK by a factor of 10.

In summary, the RM first-stage equations have a very different specification than the PK first-stage equations, although both models have the same second-stage. The RM model including non-choice observations in the first-stage and many times the number of independent variables (weak instruments) cannot be used to test for 'instrument weakness' of the PK model, and compounds the bias by using linear LIML rather than 2SLS.

## 2. Instability and Outliers

The discussion of outliers in the RM paper provides the reader with only a selective view of how the distribution of the data affects the PK results (their use of the word 'destroying' is misleading at best). RM also ignore the estimation and testing strategy of PK, in which the exogeneity of credit is formally tested and then, if not rejected, is imposed by estimating the outcome regression by ordinary least squares. Indeed RM arbitrarily select one way to trim the data – dropping the 16 largest values for consumption – with no consideration of how other trimming strategies affect the results, including the effect of trimming on tests of exogeneity. Pitt (2013) presents an analysis of results using differing levels of symmetric trimming and all confirm positive estimated female credit effects. Symmetric trimming of 16 observations from each tail of residuals results in non-rejection of exogeneity and OLS estimates that have statistically significant and positive credit effects for both females and males. Trimming more observations symmetrically results in larger estimated female credit effects and higher t-ratios. The RM approach of deleting a single selected set of observations and then obtaining estimates that do not follow the protocol clearly described and implemented in PK cannot be seen to 'destroy' the PK results (and does not seem in the spirit of proper replication).

Also, consider the interpretation of the sensitivity analysis of RM in their Table 4, carried out with maximum likelihood, in which they re-estimate the PK model over subsets of the data. Re-estimating the PK model with a subset of the data is not like, for example, breaking up a 2SLS wage equation sample into white and non-white sub-samples. Parameter identification arises from the deterministic zero credit observations (no choice to join) in the PK model – the second part of the log-likelihood given by Equation (1) – so that dropping large shares of non-choice observations, as RM do, eliminates a large share of the source of parameter identification. Essentially, RM demonstrate that when they drop out observations that identify the model, the model is less well identified. Why should that be surprising and on what basis can they draw the conclusion that the model is not well specified?

## 3. 'A missing discontinuity'

Although RM claim to 'follow the advice of Imbens and Lemieux (2008) on preliminaries to discontinuity-based regression', they omit the vital first step of graphing the average value of the *outcome* variable (household consumption expenditure) around the discontinuity in the forcing variable (RM include a graph of the value of credit by gender against landholdings, but credit is not the outcome variable and landholdings is not the forcing variable). Because this is a fuzzy regression discontinuity design setup (contrary to the claim of RM), landholdings affect not only whether a household *can* join a credit programme, but also whether it actually does and, conditional on joining, how much it borrows. Consequently, in a region around the land eligibility rule, the local relationship of land to borrowing can be obscured by land's effect on whether a household participates and how much they borrow, not just if it can borrow. In addition, the other determinants of participation and borrowing in PK (a total of 80 variables that include education, age, household composition and the village fixed effects) are correlated with land. By not controlling for these confounding covariates, this bivariate Lowess graph of borrowing versus land tells us little.

Pitt (2013) documents various other failings in the way RM test for the existence of a discontinuity, and (i) finds a discontinuity for a graph of the *outcome* (consumption) on landholding using Lowess regression; (ii) finds a discontinuity in a regression that controls for the actual forcing variable, all other independent variables, and a sixth degree polynomial for landholding; and (iii) reiterates the result of Pitt and Khandker (2012) that shows that whether a land discontinuity exists in the data is irrelevant. Statistically significant female credit effects are found even when all households in villages with credit programmes are treated as having endogenous choice to participate; that is, when no use is made of a discontinuity in landholdings at all.

## 4. Consistency, Sources of Identification, Simulation and the 'Logarithm of Zero'

Pitt (2013) and Pitt and Khandker (2012) refute the technical criticisms made by RM, such as the consistency of their linear LIML model as compared to the PK model, the sources of identification, and the role of the 'logarithm of zero.' RM's claim about the consistency of their linear LIML model does not hold up under closer scrutiny under *any* distribution of the data. There are *no* exogenous slope ($\mathbf{x}$) variables in any of the models that RM simulate. The *only* variables in the first-stage are intercepts, and as argued in Pitt (2013), those intercepts should *not* be identifying instruments. The RM model is inconsistent and the PK model is consistent if there are any exogenous slope variables and the errors are normal (Pitt, 2013; Pitt & Khandker, 2012). It is the existence of exogenous regressors that generates the plethora of interaction terms in the 2SLS setup of RM that make the RM instrument set weak and the linear LIML estimates invalid. As Pitt and Khandker (1998) estimate a model with 254 free parameters, the RM simulation without any exogenous $\mathbf{x}$ slope variables is a seriously deficient model from which to make claims about the PK and RM approaches, particularly without making any mention of the crucial importance of slope variables.

The 'logarithm of zero' issue is concerned with how to assign a value of the 'treatment', call it log ($\alpha$), to those who do not receive treatment (borrow) because they do not have choice. RM fail to realise that by controlling for credit choice in the second stage, PK simulations have made the estimates of programme effects invariant to linear translation – that is, the estimates are invariant to the choice of $\alpha$. Invariance to linear translation means that if any constant term is added or subtracted to credit (that is, to the value of the treatment of the treated), the estimated programme effects are unaffected. (Since the model is also invariant to scale, invariance to linear translation makes it invariant to any linear transformation of credit.) Consequently, RM are in error to say that the PK simulation is deficient by choosing zero mean borrowing, since the PK specification would have reported *exactly* the same results had the mean been any value.

Although the Pitt and Khandker (2012) simulation model is invariant to linear translation of the credit variable, the estimates presented in Pitt and Khandker (1998) are not fully invariant for reasons noted in Pitt (2013). However, the results are barely changed when credit is rescaled by dividing all credit by 10 (or 20), which is equivalent to increasing the credit assigned to non-participators from log ($\alpha$ = 1) to log($\alpha$ = 10) , a change in scale responsive to the untested view of RM that log($\alpha$ = 1) is (implausibly low). Doing so modestly *increases* the estimated female credit effect without changing the t-ratio, even though with $\alpha$=10 the gap between the smallest credit treatment and no treatment falls by 90 per cent and the variance of female and male credit falls by nearly half

## 5. The Process of Replication

In their first version, Roodman and Morduch (2009) erroneously claimed statistically significant and negative female credit effects, the opposite of the PK findings, a finding that they advertised widely. This finding came about because of what can only be called a typographical error on their part. This error, introduced by Roodman into part of his code, led everyone astray for some time because what they actually did was not what they say they did. This error was not the result of being confused by some complicated bit of mathematics, econometric modelling and programming, or by us not sharing computer code, as they clearly understood what the correct variables are or they would not have listed them correctly in the tables of all their papers or in their initial email to me. Before that error was discovered, they made many public pronouncements declaring that Pitt and Khandker (1998) is faulty; testified to a committee of the US Congress (and posted the video of that testimony on Youtube at http://www.youtube.com/watch?v=5Y10XbCLys8) on the supposed faults of PK; wrote a book for the popular market (Roodman 2011) with 'Due Diligence' in the title that has as a central claim that PK and those who funded microfinance programmes did not perform 'due diligence' in their work. They have made very self-assured proclamations that PK are flawed, claiming (in David Roodman's Microfinance Open Book Blog: http://www.cgdev.org//open_book) that 'academia has some explaining to do: first the most prestigious study says microcredit reduces poverty, then it is overturned [by RM] (posted 1 March 2010)' and that Roodman and Morduch (2009) is 'the academic equivalent not of a citation but an indictment … It is a long document packed with logic and evidence that the flaws are not merely possible but provable in academic court and important enough to generate wrong results (posted 1 March 2010)' and that '[the] message that a lot of research published in prestigious journals is wrong *does* carry over to economics in general and microfinance in particular. Cases in point are the papers that Jonathan and I replicated (posted 6 January 2011)'. Finally, in a blog entry title titled 'Taking the Con Out of Econometrics,' Roodman writes:

> how could the economics profession have gone so wrong for so long? … the old research is fundamentally suspect and the new much better (though hardly perfect). The fancy math in what was once the leading study of microcredit's impacts is, though beautiful, typical of the old generation in its propensity to obscure rather than resolve the fundamental barriers to identifying cause and effect (posted 28 March 2010).

Replicators have every right to make the results of their research known and to challenge academic conventions. But perhaps statements as definitive as these should wait until after an external review process has been completed. Otherwise the replicators appear to be wedded to an outcome via their public pronouncements, and their objectivity as the process proceeds appears somewhat diminished. The serious issue of specification searches ('data mining') in empirical economics applies as well to mis-specification searches in replications (and apparent ‘data selection’ and ‘model selection’). The work of replicators is important and it is their responsibility to take extra care in the quality and accuracy of their methods and the claims they make (Pitt 2012).

## Acknowledgements

I thank Shahidur Khandker, Tiemen Woutersen and an editor of *JDS* for comments and suggestions. The longer version of this paper is available as Pitt (2013) at http://www.brown.edu/research/projects/pitt/.

## References

Davidson, Russell & James MacKinnon (1993). Estimation and Inference in Econometrics. New York: Oxford University Press.

Imbens, G.W., & Lemieux, T. (2008). “Regression discontinuity designs,” *Journal of Econometrics*, *142*, 615–635.

Morduch, J. (1998). Does microfinance really help the poor? New evidence from flagship programs in Bangladesh. Retrieved from http://nyu.edu/projects/morduch/documents/microfinance/Does_Microfinance_Really_Help.pdf

Pitt, M. M. (1999). Reply to Jonathan Morduch’s ‘Does microfinance really help the poor? New evidence from Flagship Programs in Bangladesh’. Retrieved from http://www.brown.edu/research/projects/pitt/

Pitt, M. M. (2011a). Response to Roodman and Morduch’s ‘The impact of microcredit on the poor in Bangladesh: Revisiting the evidence’. Retrieved from http://www.brown.edu/research/projects/pitt/

Pitt, M. M. (2011b). Overidentification tests and causality: A second response to Roodman and Morduch. Retrieved from http://www.brown.edu/research/projects/pitt/

Pitt, M. M. (2012). Gunfight at the NOT OK Corral: Reply to ‘High noon for microfinance’. *Journal of Development Studies, 48*, 1886–1891. [Expanded version available for retrieval at http://www.brown.edu/research/projects/pitt/]

Pitt, M. M. (2013). “Re-Re-Reply to “The Impact of Microcredit on the Poor in Bangladesh: Revisiting the Evidence” Retrieved from http://www.brown.edu/research/projects/pitt/

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

Pitt, M. M., & Khandker, S. R. (2012). *Replicating replication: Due diligence in Roodman and Morduch’s replication of Pitt and Khandker (1998)* (Working Paper 6273). Washington, DC: World Bank.

Roodman, D. (2011). *Due diligence: An impertinent inquiry into microfinance*. Washington, DC: CGD Books.

Roodman, D., & Morduch, J. (2009). *The impact of microcredit on the poor in Bangladesh: Revisiting the evidence* (Working Paper 174). Washington, DC: Center for Global Development.

Roodman, D., & Morduch, J. (2014). The impact of microcredit on the poor in Bangladesh: Revisiting the evidence. *Journal of Development Studies* (this issue).