ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Smits, Joeri; Günther, Isabel

Article

Do financial diaries affect financial outcomes? Evidence from a randomized experiment in Uganda

Development Engineering

Provided in Cooperation with:
Elsevier

Suggested Citation: Smits, Joeri; Günther, Isabel (2018) : Do financial diaries affect financial outcomes? Evidence from a randomized experiment in Uganda, Development Engineering, ISSN 2352-7285, Elsevier, Amsterdam, Vol. 3, pp. 72-82, https://doi.org/10.1016/j.deveng.2018.02.001

This Version is available at:
https://hdl.handle.net/10419/187792

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

https://creativecommons.org/licenses/by-nc-nd/4.0/

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Development Engineering 3 (2018) 72–82

Contents lists available at ScienceDirect

Development Engineering

journal homepage: www.elsevier.com/locate/deveng

Do financial diaries affect financial outcomes? Evidence from a randomized experiment in Uganda☆

Joeri Smits [a,b,*], Isabel Günther [c]

[a] Yale School of Management, New Haven, CT 06520, USA
[b] Harvard Kennedy School, Harvard University, Cambridge, MA 02138, USA
[c] Chair of Development Economics, NADEL, ETH Zurich, Switzerland

ARTICLE INFO

Keywords:
Household income and expenditure
Measurement error
Financial diaries

ABSTRACT

Survey data on income and expenditure is often of low quality and does not capture the volatile and irregular nature of cash flows of poor households. Financial diaries are increasingly used to improve the precision and accuracy of consumption and income estimates. In this paper we analyze whether keeping track of income and expenditures changes financial behavior and outcomes, which could reduce the validity of diaries as a measurement instrument. Members of urban Ugandan microcredit groups were, through random assignment, offered financial diaries to keep a record of their daily cash flows for more than a year. We find no evidence that financial diaries change numeracy skills, loan repayment, reported income, or food consumption. We only found a difference in savings, but this is unlikely to represent any impact of the financial diaries, as it does not exceed the amount provided as an incentive to the respondent for participation.

1. Introduction

The day-to-day cash flows among poor households in general, and microentrepreneurs in particular, tend to be highly temporally variable, an aspect often not captured by one-time recall household surveys eliciting household income or expenditures. Another measurement problem with standard recall data on income and expenditure, apart from not capturing temporal volatility, is that such data is typically subject to recall error. Household consumption expenditure, and especially household income, are typically under-reported, with the error being relatively larger at lower incomes (Azzarri et al., 2010; Hoderlein and Winter, 2007). Such differential measurement error contrasts with non-differential, or classical, measurement error, which is statistically independent of the 'true' variable itself, and of the covariates. In ordinary least squares (OLS) regression, non-differential measurement error in the dependent variable does not generate bias, but only inflates standard errors. However, differential measurement error in the dependent variable generates substantial biases in estimates of treatment effects across a range of estimators (Gutknecht, 2011; Millimet, 2011).

Researchers have used techniques like financial diaries to take high frequency recordings of individuals' or households' cash flows for prolonged periods of time in order to get a more insightful picture of their financial lives and/or income and profits. More intensive measurement techniques such as financial diaries may also allow for data that is to a lesser extent subject to differential measurement error (and could therefore potentially serve as validation data for standard recall data). Perhaps the most well-known example is the series of financial diary studies in South Africa, Bangladesh and India reported in Collins et al. (2009), along with those in the US (Hannagan and Morduch, 2015).

The approach we took in this study was to have respondents record their own cash inflows and outflows in cashbooks designed specifically for this purpose; example entries are depicted in Appendix A. This approach is similar to Kamath and Ramanathan (2015), who asked ninety respondents in Ramanagaram (India) to record their own cash flows. We follow their procedure and asked the participants' children to write the diary if they are illiterate or unable to write. For this approach to data collection, it is still largely unknown whether having

☆ Some of the data in this paper were collected as part of a research project that was supported financially by the Independent Evaluation Unit of the KfW Development Bank. We thank Martin Brown and Lore Vandewalle, as well as participants at the 2016 SEEDEC conference and the 2016 EPFL Tech4Dev conference for comments on earlier drafts of this paper. George Rugaba and Praise Asiimwe provided excellent research assistance.
* Corresponding author. Yale School of Management, New Haven, CT 06520, USA.
E-mail address: joeri.smits@yale.edu (J. Smits).

https://doi.org/10.1016/j.deveng.2018.02.001
Received 8 September 2016; Received in revised form 11 August 2017; Accepted 1 February 2018
Available online 15 February 2018
2352-7285/© 2018 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

individuals track their cash flows - or having an enumerator track them - alters financial behavior and outcomes. Such behavioral responses would potentially jeopardize the validity of such diaries as a measurement instrument, and could reduce or offset any accuracy gains over standard, recall-based approaches. The book "Portfolios of the poor", by Collins et al. (2009), in which 300 individuals from three countries are interviewed on a biweekly basis for over a year (i.e., a slightly different approach than ours), notes:

> "There is […] doubt as to whether participating in the diaries changed the behavior of some respondents. In some cases it may have done so. […] However, without a different type of study design, it is difficult to tease out exactly how much of an influence we might have had." (pp. 209–210).

This could happen through various mechanisms. First, being interviewed on a biweekly basis for over a year - as in the case of Collins et al. (2009), or keeping record of one's cash flows in a cashbook for an extended period of time (as in the current study), may increase financial awareness, perhaps reducing temptation expenditures, saving some business costs, or changing business strategies more generally. Second, being repeatedly interviewed or keeping track of one's financial data could alter people's effort levels, in the sense of observer or Hawthorne effects. Third, in the case of our study, some participants wrote the names of people who bought goods from them on credit in the book and made a 'check' sign when the money has been repaid. This may help those using the cashbooks to keep track of their lending relationships and increase repayment, or lend differently, than they would without the use of the cashbook. Disentangling changes in reported expenses (e.g., reported food consumption) from changes in the underlying behavior (e.g., actual food consumption) is, however, difficult.

In a review of the literature on the measurement of consumption expenditures, Crossley and Winter (2013) identified only two studies, both from the UK, that investigate effects of expenditure diaries on behavior. In the first, Kemsley et al. (1980) report on a small scale experiment on a subsample of the UK Family Expenditure Survey, in which household members aged over 15 were asked to keep a diary of their expenses. Making a payment to record-keepers improved the response rate substantially, with the response rate increasing in the payment. In the second, McKenzie (1983) concludes that there is no evidence that keeping a diary affects telephone usage (when comparing diary records to metered usage). Of course, the latter result does not necessarily generalize to other categories of expenditures. In an experiment in Tanzania by Beegle et al. (2012), who randomized 4000 households into different methods of data collection, recall modules also lead to lower (reported) consumption than a personal diary, with the gap being larger among poorer households.

Whereas the aforementioned studies looked at expenditure diaries, De Mel et al. (2009) randomly handed ledger books to microenterprises in Sri Lanka so they could record business revenues and profits. They found that the use of diaries led to significantly higher (reported) business expenses and to higher (reported) revenues of similar magnitudes (hence having no effects on profits), suggesting that standard recall surveys lead firms to under-report both revenue and expenses.

None of the studies we reviewed here, however, explicitly analyzed the impact of diaries on financial behavior and outcomes and contributes to the literature on improved measurement of income and expenditures. It is to our knowledge the first study to randomize access to cashbooks for the self-recording of cash flows to gauge its effect on financial behavior and outcomes. Our overall findings are reassuring for financial diary-type approaches to the measurement of cash flows, and can be summarized in one sentence: we do not find any evidence that diaries affect financial outcomes. The lack of effects detected in this study of diary-logging on financial outcomes indicates that diary studies may be a way to obtain more credible cash flow measurements compared to standard recall surveys.

There have been other approaches to measuring income and expenditures, with varying success. An alternative approach to measuring microenterprise profits, which are a main source of income for the substantial share of self-employed people in many developing countries, was attempted by De Mel et al. (2016) in Sri Lanka. They used radio frequency identification (RFID) tags as a means of objectively measuring stock levels and stock flow in small retail firms, but found that RFID was difficult to use and time-consuming for the business owners, and that they performed very poorly for most products sold by microenterprises. Yet another approach is being piloted by Carletto (2013), who conducted a randomized experiment comparing standard recall with enumerator visits twice a week (which he considers a gold standard) and telephone calls twice a week. Although Carletto (2013) measures crop output, their approach should be applicable to other development indicators. The question that remains is whether being interviewed repeatedly over the telephone affects an individual's and a household's effort levels and economic and financial outcomes.

The rest of this paper is organized as follows. Section 2 describes the experimental design and the data. Section 3 describes the econometric approach, section 4 reports the results, and section 5 discusses and interprets the results and offers recommendations for further research.

2. Data and design

To test the null hypothesis that the use of cashbooks has no effect on financial outcomes, a clustered randomized encouragement design study was carried out.[1] We collaborated with a major microfinance institution in Uganda in order to sample microcredit borrowing groups for an unrelated research project on the impact of microcredits. Specifically, three types of groups were interviewed during the baseline study from September 2013 until March 2014:

- Village Groups: individuals who have a 'Village Group Loan', a non-collateralized loan product with joint liability[2] from the microfinance institution. These groups have ten or more members.
- Applicant Village Groups: these individuals have formed a group and have applied for their first Village Group Loan, but have not yet received their loan at the time of the baseline survey.
- Small Groups: these individuals have a 'Small Group Loan', another loan product from the same microfinance institution. Some of the larger loans for this loan type are collateralized. The group size varies between five and ten members.

Within each of the three group types, the groups were alternately assigned to either treatment (T) or control (C) based on the order of being interviewed. Within each group type, the alternating sequence starts with a treated group as follows:

- Village Group Loan: T,T,C,C,T,T,C,C,T,T,C,C,T,T
- Applicants to Village Group Loan: T,T,C,C,T,T,C,C,T,T,C,C,T,T
- Small Group Loan: T,C,T,C,T,C,T,C,T,C,T,C

The members of groups (i.e. clusters) assigned to treatment (22 groups, n = 207) were offered a cashbook free of cost, while members of groups assigned to control were not (18 groups, n = 110).

[1] When the study was initiated, only a dozen RCTs had been registered at the American Economic Association's (AEA) RCT registry - the current study was not among them. We recognize the value of public pre-registration of RCTs and hope future, larger-scale pre-registered RCTs will aim to replicate our study design.

[2] Each member receives his or her own loan, and has to repay it individually. Joint liability means that if one member is late on repayment or refuses to repay the loan, the other members have to make up for it (possibly after a grace period), or the lender can withdraw from their compulsory savings that are part of the repayment installments.

Since the groups were initially interviewed as part of a research project whose primary goal was not the current experiment, the order in which they were interviewed is statistically independent of treatment assignment: our research design is, hence a cluster randomized encouragement design.

For the purpose of this study, this sampling strategy has several advantages, along with some disadvantages. A first practical advantage is that the disbursement of the diaries, the training in their usage, and the tracking compliance of individuals are all made less costly and time-consuming by the fact that individuals meet at least monthly in groups for repayment in the first months following their loan disbursement (the loan cycle). A second advantage is that virtually all (95%) of these borrowers are microentrepreneurs, i.e., they have a business of some sort which is generating income. This is the type of respondent for whom it is hardest to obtain accurate income and expenditure figures through standard one-time recall surveys. Previous studies have shown that under-reporting is much more prevalent and severe for self-employment income data than it is for wage income data (Grosh and Glewwe, 2000; De Mel et al., 2009; Joshi et al., 2011; Akee, 2011). Entrepreneurs are also the type of individual for whom the books are likely to be of most use, since wage earners experience less income volatility and uncertainty. Borrowers taking collateralized individual liability loans are typically the ones with larger businesses, and many will already use their own cashbooks.

A downside of this sampling strategy is that for some of the results we obtain, care should be taken with extrapolation to the general (non-group borrowing) population. In order to save on costs and eliminate enumerator bias, we worked with one (very experienced) enumerator to visit all households on a monthly basis to encourage continued (proper) use of the books. Participants who were offered diaries were trained to use them: a filled-in example was provided and the monthly visits were also used to see if participants had difficulty using the diaries as intended.

Since groups assigned to control did not have access to the cashbooks, non-compliance is one-sided. However, since the treated and control clusters (borrowing groups) are sampled from the same divisions of Kampala ,[3] we cannot fully rule out spillovers in terms of the spread of the more general idea of book-keeping as a result of interaction between treated and control loan groups. We deem such spillovers relatively unlikely given that only 203 individuals were offered a cashbook in a city of 1.5 million residents. There does not seem to be a systematic increase in record-keeping behavior post-treatment for the controls: among the individuals assigned to control, 40% reported keeping some form of record at baseline and 36.7% reported doing so at follow-up.

The size of the borrowing groups (i.e., the cluster size) varied between five and 24 individuals (average: 11.41). Of the 205 individuals offered a book at baseline, 135 (67.5%) still used it 13 months later, when the experiment ended with a follow-up survey.[4] 11 individuals (5.4%) rejected the offer of treatment (either because they had no time to use the book or no interest in it), and 55 (27.1%) started using the book, but stopped at some point before the follow-up survey took place — most of them within six months of receiving the book. Four individuals could not be found and re-interviewed during the follow-up survey. Since individuals in the control group (112 individuals) did not have access to the diaries, none of the individuals in the control group used the diaries. Fig. 1 depicts the sampling scheme and treatment compliance.

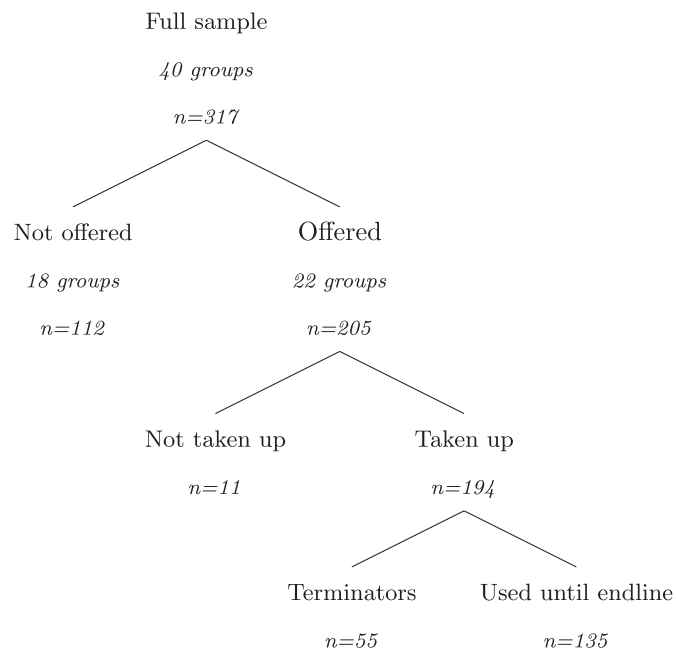


Fig. 1. Sampling scheme and treatment compliance.

A small monetary incentive was offered four months (UgSh[5] 25,000), eight months (UgSh 30,000), and 12 months (UgSh 35,000) into the experiment to those who were still using the cashbooks. The first incentive was not announced, but after receiving the first incentive, respondents were told they would receive further incentives four and eight months later. These one-off amounts, although small, may have affected some of the estimated treatment effects (more on this in the results section).

The outcomes elicited at baseline and follow-up are:

- Current personal savings: Personal savings (in 1000's UgSh.) the respondent has at the time of interview.
- Food consumption index: A food frequency questionnaire (FFQ) was administered to infer the frequency of intake of a range of 20 food categories (Kikafunda and Lukwago, 2005). For more details, see Appendix B. A continuously distributed food consumption index was constructed from these consumption frequencies using polychoric principal component analysis (Kolenikov and Angeles, 2009). By assuming that the first principal component represents food intake quantity and quality, we expect this index to be monotonically increasing in total food expenditure.[6]
- Self-reported repayment delays. The answer (yes/no) to the question 'Have you ever repaid too late (i.e., later than agreed beforehand) or not at all for this loan?' for the largest loan the household has outstanding.
- Wealth index. From the amount of household assets (bicycle, radio, TV, etc.), livestock (poultry, goats) and categorical indicators of housing conditions (roofing, type of water supply), a wealth index was constructed, again using polychoric principle component analysis (PCA). The full list of assets is reported in Appendix B.
- Personal income. Answer to the question 'What is on average your personal monthly income?'.

[3] Kampala city has five divisions: Central, Kawempe, Makindye, Nakawa and Lubaga. We sampled from all five divisions as well as from the Wakiso, Luweero and Mukono districts surrounding Kampala.

[4] As a result the sum of the forth row of Fig. 1 does not add up to 194. It is unclear whether individuals that could not be found during the second wave used the books until end-line or terminated the use of the books beforehand. In addition one individual of the control group could not be interviewed during follow up. We consider the attrition of 5 out of 317 individuals as very low.

[5] 1 USD ≈ 3000 UgSh as of 08/07/2016.

[6] We did not elicit food expenditures in the baseline and follow-up survey, the measurement error in such a variable being one of the raisons d'être behind the initiation of the current study in the first place.

Table 1
Summary statistics at baseline and balance tests.

Variable	N	Mean	SD	Treated avg	Control avg	Diff. (t-stat)[b]
Current personal savings[a]	234	29.70	56.79	26.46	35.93	1.21
Food cons. index	317	0.10	0.97	0.05	0.19	1.25
Late repayment on loan	197	0.20	0.40	0.22	0.17	0.90
Log (personal income)	257	12.55	0.83	12.54	12.59	0.41
Wealth index	317	−0.00	0.99	0.01	−0.02	0.17
Covariates						
Female (=1)	317	0.78	0.42	0.83	0.68	3.23***
Head of household (=1)	317	0.51	0.50	0.48	0.57	1.51
Age in years	317	37.63	10.71	37.42	38.03	0.48
Household size	317	5.54	3.07	5.40	5.80	1.12
Nr. of children in household	317	2.63	2.13	2.54	2.79	1.03
Completed primary educ.	317	0.32	0.47	0.32	0.32	0.08
Compl. secondary educ.	317	0.26	0.44	0.22	0.31	1.72
Compl. tertiary educ.	317	0.05	0.22	0.05	0.05	1.19
Muslim (=1)	317	0.22	0.42	0.21	0.24	0.64
Financial literacy score	317	6.40	2.48	6.39	6.44	0.18
Applicant Village Group	317	0.40	0.49	0.41	0.39	0.29
Non-Applicant Village Group	317	0.43	0.50	0.46	0.37	1.59
Small Group	317	0.16	0.37	0.13	0.21	1.92**
Member of ROSCA (=1)	317	0.74	0.44	0.71	0.80	1.88**
Shock took place (=1)	317	0.58	0.49	0.62	0.49	2.31**
Income stable[c]	317	0.34	0.47	0.32	0.37	0.79
Income unstable	317	0.55	0.50	0.57	0.51	1.06
Income very unstable	317	0.09	0.28	0.09	0.09	0.04
Recorded income c)	317	0.17	0.37	0.17	0.16	0.23
Recorded both	317	0.22	0.42	0.21	0.24	0.64
No idea of cash flows	317	0.32	0.47	0.35	0.28	1.35*

[a] In 1000s UgSh. A substantial share of respondents was not able to come up with (or did not want to give) an estimate of their savings.
[b] T-statistic from two-sample *t*-test of [mean(control) - mean(treated) = 0]; *p<0.1, **p<0.05, ***p<0.001.
[c] These three variables capture the responses to the question 'which of the following statements applies most to your household regarding keeping record of income and expenditures?' The answer options are 'we keep records of everything, recording all revenues and all expenditures', 'we only keep records of revenues', 'we do not keep records', and 'we do not know how much money we receive and spent per month'.

- Numeracy score: the sum of correct answers to a quiz with 30 numeracy questions for which the respondent has five minutes to answer (see Appendix B).

Many respondents found the 'current personal savings' question rather sensitive, and we have many missing values for this variable. For the food consumption index, it is important to note that this variable is not based on consumption expenditures (which are recorded by the respondent in the diaries). An observed effect on reported food expenditure could reflect just the effect of being more aware of the expenditures due to the daily self-recording - and not effects on the individual's *actual* food consumption. For this reason, the index is constructed from the intake *frequency* of consumption of a range of food categories. Also note that the variable numeracy score was only elicited during the follow-up survey.

We also observed a range of other variables at baseline and at follow-up. These include the standard demographics (gender, age, household size, household composition, religion of the participant, and whether the respondent is the head of the household), as well as variables more specific to the current study, such as an indicator for the participant being a member of a saving group, such as Rotating Savings and Credit Association (ROSCA), financial literacy score, and an indicator variable for adverse economic shocks (see Appendix B for details on the construction of these variables). The control variables also include indicators for the type of group loan: Small Group loan (5–9 members), Village Group loan (10 + members), or first-time applicant groups for Village Group loans. In addition, respondents were asked at baseline whether they already (i.e., prior to the experiment) kept somehow track of either their cash inflows and/or outflows. Since outcomes such as food consumption may be subject to seasonality even in urban areas, indicator variables for the calendar month of the baseline interview are included (but only needed for the baseline month, as follow-up interviews took place 13 months after the baseline interview).

3. Econometric approach

We estimate the intention-to-treat (ITT) effect: the average effect (on an outcome) of being offered a cashbook for all the households

Table 2
Subjective evaluations of the diaries by the respondents.

	Totally disagree	Disagree	Neither nor	Agree	Totally agree
The book has made me better in working with numbers.	0%	0%	2%	50%	48%
The book has given me a clearer picture of my cash flows and my business.	0%	2%	2%	48%	48%
Filling in the book took me a lot of time.	10%	17%	11%	42%	20%
Using the books was a waste of time.	12%	29%	7%	30%	22%
The book has helped me in recovering debts	1%	8%	6%	60%	25%
The book has helped me limit my consumption expenditure	1%	25%	11%	53%	10%
I will continue to use the book on my own from now on	2%	25%	4%	61%	10%

Table 3
ITT estimates of financial diary access.

Outcome	DID (1)	DID with covariates (2)	DID with FE (3)	FD (4)	Obs.
Current personal savings (1000s of UgSh.)	16.342	15.532	22.092	13.556	466
	(7.465)**[a]	(7.692)**	(8.902)**	(6.938)*	
Food cons. index	0.245	0.104	0.245	0.245	624
	(0.205)	(0.210)	(0.205)	(0.205)	
Self-reported late repayment[b]	−0.056	−0.012	−0.074	−0.074	410
	(0.152)	(0.130)	(0.141)	(0.142)	
Wealth index	−0.161	−0.224	−0.145	−0.145	564
	(0.163)	(0.155)	(0.162)	(0.162)	
log (Monthly personal income)	0.122	0.168	0.024	−0.066	456
	(0.197)	(0.200)	(0.213)	(0.478)	
Numeracy quiz score[c]	0.013	0.003			312
	(0.605)	(0.038)			
Joint test (p-value)	0.612	0.170		0.148	

[a] Robust standard errors in parentheses, clustered at the level of the borrowing group. *p < 0.1, **p < 0.05, ***p < 0.01.
[b] Observations for individuals who were loan applicants at baseline are excluded from this estimation (i.e., the members of Applicant Village Groups). As they did not have a loan at baseline, they could not be too late with repayment at baseline.
[c] Since only follow-up data are available for this outcome, column (1) reports the results from a simple *t*-test, and column (2) reports results from a Poisson regression.

in the borrowing groups assigned to treatment compared to all the households in the control groups, irrespective of whether the household actually took up the treatment offer. The pooled difference-in-difference (DID) estimation (Angrist and Pischke, 2008) takes the form

$$y_{ijt} = \alpha + \gamma treated_j + \delta post_t + \beta^{DID}(treated_j \cdot post_t) + \epsilon_{ijt} \quad (1)$$

where y_{ijt} is an outcome variable for individual i in group j at survey round t ($t = 0$ for the baseline, $t = 1$ for the follow-up), $treated_j$ is an indicator for the group being assigned to treatment, $post_t$ is an indicator variable for the follow-up survey, and ϵ_{ijt} is an idiosyncratic error term, clustered at the borrowing group level. The coefficient β^{DID} is the difference-in-difference estimator of the ITT effect. Bruhn and McKenzie (2009) argue for controlling for variables that are strongly correlated with the outcome even if they are balanced at baseline, as this may increase statistical power. Hence, our second estimation method includes individual-level controls X_{ijt}:

$$y_{ijt} = \alpha + \gamma treated_j + \delta post_t + \beta^{DID}(treated_j \cdot post_t) + \eta X_{it} + \epsilon_{ijt} \quad (2)$$

In addition, we report estimates where we include household fixed effects α_i:

$$y_{ijt} = \alpha_i + \delta post_t + \beta^{DID}(treated_j \cdot post_t) + \epsilon_{ijt} \quad (3)$$

Finally, we report estimates where any time-invariant unobserved heterogeneity at the respondent level is taken out through first-differencing:

$$\Delta y_{ij} = \alpha + \beta^{FD} \Delta treated_j + \Delta\epsilon_{ij} \quad (4)$$

4. Results

Table 1 reports summary statistics on the key variables. Over 70% of the participants are female, as the group loans of many microfinance institutions target women in particular. Table 1 also gives the means of the variables for subsamples assigned to treatment and control, and t-statistics of t-tests that the differences in means are statistically significantly different from zero, i.e., a baseline balance test. For most of the variables and for all the financial outcomes we study, differences between the treatment and control groups at baseline are not statistically significant.

There is one notable exception. There are notably more females in groups assigned to treatment than in groups assigned to control. This can, however, be explained by the following features of the experimental design: (i) recall that we interviewed three types of groups, and randomized alternately within each group type: (a) Village Groups, (b) Applicant Village Groups, (c) Small Groups; (ii) for Village Groups and Applicant Village Groups, the randomization was T,T,C,C,T,T, … (T denoting assignment to treatment, C assignment to control), while for Small Groups the randomization was T,C,T, …; and (iii) the final number of Village Groups sampled was even (14 each) as was the number of Small Groups sampled (12). Hence, for Village Groups we had four more groups assigned to treatment than assigned to control. In the traditional microcredit model, microfinance institutions target women for Village Group recruitment and they indeed have a higher share of women: 53.6% of Village Group borrowers are female, whereas only 48.8% of Small Group borrowers are female.

As a first step in the analysis, we report the subjective evaluations of the financial diaries (by those individuals who reported using them) in Table 2. The reported figures are from the follow-up survey. While the answers to the first few evaluation questions may be somewhat subject to social desirability bias, a majority of diary users reported that the books improved their recovery of debts (85%) and increased their capacity to accumulate savings (63%). The number reporting these positive effects was slightly greater than those who felt the diaries were a waste of time (52%).

To quantitatively analyze the effects of the diaries, we report estimated effects according to specifications (1)–(4) in Table 3. These correspond to the ITT effect, or the average effect of having access to financial diaries. Column (1) reports estimates from the basic difference-in-difference (specification 1); column (2) reports the estimates of the specification with the full set of covariates (specification 2); column (3) corresponds to the estimation with fixed effects (specification 3); and column (4) reports the estimates in first differences (specification 4).

We only find a statistically significant ITT effect for current personal savings, with a point estimate of increased personal savings amounting to 3.2% of the median household income. We address the issue of multiple inference in two ways. First, we run seemingly unrelated regressions to test for the joint significance of the ITT coefficients across outcomes. The joint p-value, reported at the bottom of Table 3, does not reach statistical significance. Second, to control for the Family-Wise-Error-Rate (FWER), the p-value would need to be below a Bonferroni-corrected level. Using Bonferroni corrections adjusted for the average correlation

between the outcomes ($\overline{\rho} = 0.150$ in our setting for the continuously distributed outcomes) following Sankoh et al. (1997), the corrected p-value would need to be below 0.0191 for a single-hypothesis testing level of $\alpha = 0.1$, below 0.0096 for $\alpha = 0.05$, and below 0.0019 for $\alpha = 0.01$.

The statistically significant ITT effect estimate for savings may be (at least partly) due to the incentives we offered book users. The ITT effect for current personal savings is UgSh 15,000–24,000. Suppose, in the extreme case, that individuals accumulate the incentive payments received for their continued use of the diaries, then these cumulative savings would amount to UgSh 56,000[7] which would explain the estimated treatment effect. A more likely scenario is that only the last incentive, received one month before the follow-up study, affects the response to the savings question. The last incentive of roughly UgSh 23,000[8] could in itself still explain the ITT estimate. Thus, we find no effect of the diaries on financial outcomes, and for the single effect which is statistical significant, we can advance several plausible scenarios in which the measured effect does not represent the actual impact of the diaries.

The change over time of the means of treated and control groups in the outcome variables is displayed graphically in Appendix C. The increase over time in self-reported late repayment outcome is explained simply by the fact that a subsample of subjects were loan applicants at baseline and have obtained a loan at the time of the follow-up survey. The associated repayment obligations also explain the reduction in liquidity and therefore reported savings.

5. Discussion

This study analyzed an experiment involving the distribution of financial diaries and training respondents to use them. The aim was to see if keeping record of one's cash flows affects behavior and outcomes, which would undermine the validity of financial diaries as a measurement tool. The main message of this paper is reassuring in this regard: we do not find evidence that individuals 'treated' with access to a free financial diary changed their financial behavior in significant ways.

The lack of effects found in this study does not imply that the issue is settled. Larger samples may be able detect statistically significant effects of diary usage. This possibility is revealed by power calculations. For instance, to attain 80% statistical power to detect a statistically significant difference of 0.2 standard deviations at the 5% level for the food consumption index, a sample size of 724 respondents would be needed, more than twice our sample size. Similarly, a sample size of 740 respondents would be needed for the wealth index.[9] We should, however, take into account that we do not even find any (insignificant) discernible effects on outcomes other than those related to savings, *despite* the monetary incentives we offered to encourage continued use of the diaries. Hence, sizable effects of using the diaries appear unlikely, but we cannot rule out minor effects.

It is also possible that any effects the diaries may have had is diluted by the fact that some respondents may not have benefited a lot from them as they were already keeping some form of records on their own; others were not literate and their children assisted them; and some respondents seemed to not be using the diaries seriously, but rather seemed to be keeping records mainly to be eligible for the periodic monetary incentives.[10] These hypotheses offer avenues for future research.

Our null findings leave open the possibility that financial diaries could be a valuable tool to improve the measurement of cash flows in some settings. Improved measurement is important in impact evaluations that aim to investigate the impact of policies on household income or expenditures, given the volatile nature of cash flows in lower income countries and the differential measurement error in income and expenditure. In addition, for more conclusive evidence regarding effects of financial diary keeping or intensive interview methods on respondent behavior and outcomes, outcome measures that are less noisy and do not rely on respondents' recall would be needed.

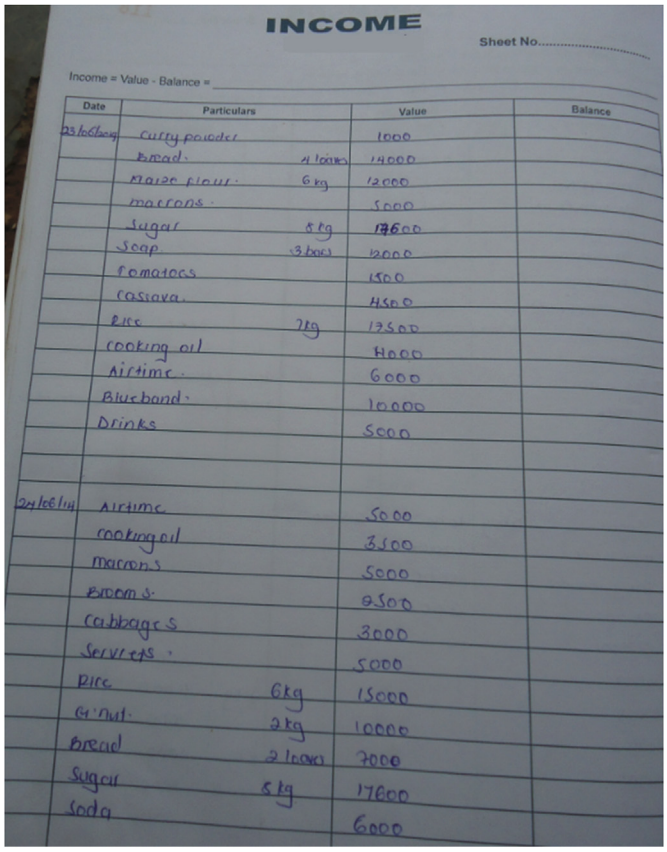
[7] A rough estimate of 25,000 + 30,000 + 35,000 times 135/207, the fraction of individuals in the assigned-to-treatment group using the books until the endline survey.

[8] A rough estimate of 35,000 times 135/207, the fraction of individuals in the assigned-to- treatment group using the books until the endline survey.

[9] These estimates are obtained using the sampsi STATA command with the empirical standard deviations of treated and control groups at baseline.

[10] In future work, we will quantify the quality of the cash flow data in the books using our data on cash holdings by the respondent at the time of the enumerator's visits, using the accounting formula as in the Appendix of Collins et al. (2009).

Appendix A. Excerpts from financial diaries



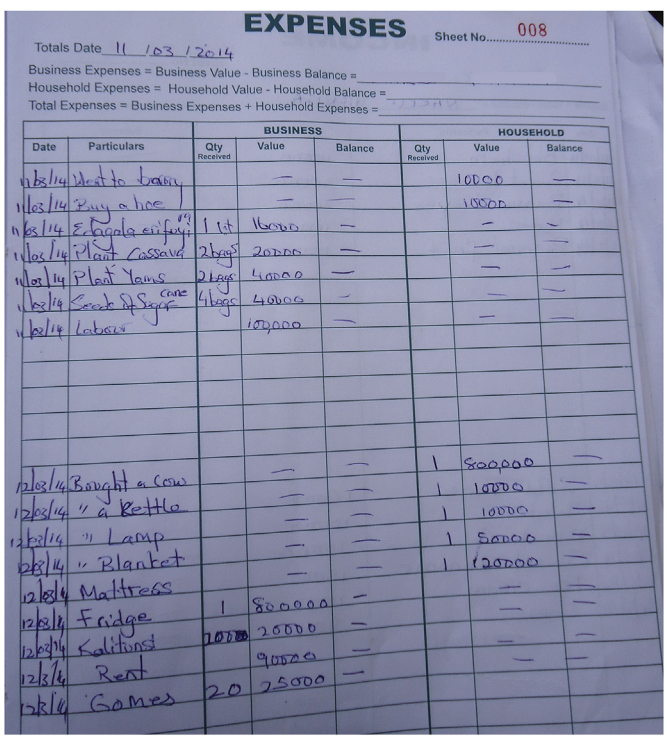
INCOME

Sheet No.....

Income = Value - Balance = _____

Date	Particulars		Value	Balance
25/06/2014	curry powder		1000	
	Bread.	4 loaves	14000	
	Maize flour.	6 kg	12000	
	macrons.		5000	
	Sugar	5 kg	17600	
	Soap.	3 bars	12000	
	Tomatoes		1500	
	cassava.		4500	
	Rice	7kg	17500	
	cooking oil		4000	
	Airtime.		6000	
	Blueband.		10000	
	Drinks		5000	
24/06/14	Airtime		5000	
	cooking oil		3500	
	macrons		5000	
	Brooms.		2500	
	cabbages		3000	
	Serviets.		5000	
	Rice	6kg	15000	
	G'nut.	2kg	10000	
	Bread	2 loaves	7000	
	Sugar	5kg	17600	
	Soda		6000	

Fig. A.2. Excerpt of the left page (income) of a financial diary.



EXPENSES

Sheet No. 008

Totals Date 11/03/2014

Business Expenses = Business Value - Business Balance = _____

Household Expenses = Household Value - Household Balance = _____

Total Expenses = Business Expenses + Household Expenses = _____

		BUSINESS			HOUSEHOLD		
Date	Particulars	Qty Received	Value	Balance	Qty Received	Value	Balance
11/03/14	Went to bank		—	—		10000	—
11/03/14	Buy a hoe		—	—		10000	—
11/03/14	Ebagala enifuyi	1 lit	16000	—		—	—
11/03/14	Plant Cassava	2bags	20000	—		—	—
11/03/14	Plant Yams	2bags	40000	—		—	—
11/03/14	Seeds of Sugar cane	4bags	40000	—		—	—
11/03/14	Labour		100,000	—		—	—
12/03/14	Bought a Cow		—	—	1	800,000	—
12/03/14	" a Kettle		—	—	1	10000	—
12/03/14	" Lamp		—	—	1	10000	—
12/03/14	" Blanket		—	—	1	50000	—
12/03/14	Mattress		—	—	1	120000	—
12/03/14	Fridge	1	800000	—		—	—
12/03/14	Kalitunsi	70000	20000	—		—	—
12/3/14	Rent		90000	—		—	—
12/3/14	Games	20	25000	—			

Fig. A.3. Excerpt of the left page (expenses) of a financial diary.

Appendix B. Construction of the food consumption index, the basic numeracy and financial literacy score, the wealth index, and the numeracy score

Food consumption index

For each item, the enumerator records the number of days the food was eaten per week. For each food group, answer options are: (1) never, (2) once, (3) twice, (4) 3–5 times, (5) almost every day, (6) not last week, but during last month.

The food groups are: (1) Maize/Millet/Sorghum/Ugali/Pocho, (2) Rice, (3) Wheat flour/bread/chappati/rolex, (4) Cassava (sweet potato), (5) Irish potato, (6) Matooke, (7) Pork, (8) Chicken/duck, (9) Beef, (10) Mutton/lamb/goat, (11) Fish, (12) Dairy products, (13) Beans or peas or lentils, (14) Vegetables (any), (15) Insects (any), (16) Sugar, (17) Salt, (18) Coffee, (19) Tea, (20) Soda.

Basic numeracy and financial literacy score

Basic numeracy skills were elicited by the following four questions:

- What is 25 + 17?
- What is 49 − 23?
- What is 12*4?
- What is 56:7?

The following five questions were posed to elicit financial literacy levels, slightly adapted from Bandiera et al. (2010):

- What is 20% out of 3000 UgSh?
- If you could save UGX5,000 per month, how many months would you need to save to get UGX30,000?
- If you needed UGX180,000, how much would you need to save per month (in UgSh) to have the money within one year (12 months)?
- Assume that you saw a radio of the same model on sale in two different shops. The initial retail price was UGX 20,000. One shop offers a discount of UGX 1,500, while the other one offers a 10% discount. Which one is a better bargain?
- Suppose you have deposited UGX 100,000 in the bank for an interest of UGX 10,000 per year. If you withdraw all the money after 3 years, how much will you get?

The basic numeracy and financial literacy score was constructed as the sum of correct answers to the nine answers plus 1 if the number of commercial banks and microcredit institutions known to the respondent ('Please mention as many names of banks in Uganda as possible') was higher than the average number for all respondents, which was 6.33 financial institutions.

Wealth index

This index is constructed from:

- counts of the following assets and livestock types: chairs, tables, beds, sofas, mirrors, watches, kerosene stoves, gas stoves, televisions, radios, mobile phones, generators, solar panels, light bulbs, bicycles, motorcycles, cars, refrigerators, chicken, goats.
- housing indicator variables: roofing (= 1 if roof has tiles, = 0 otherwise), water (= 1 if water is piped, = 0 otherwise).

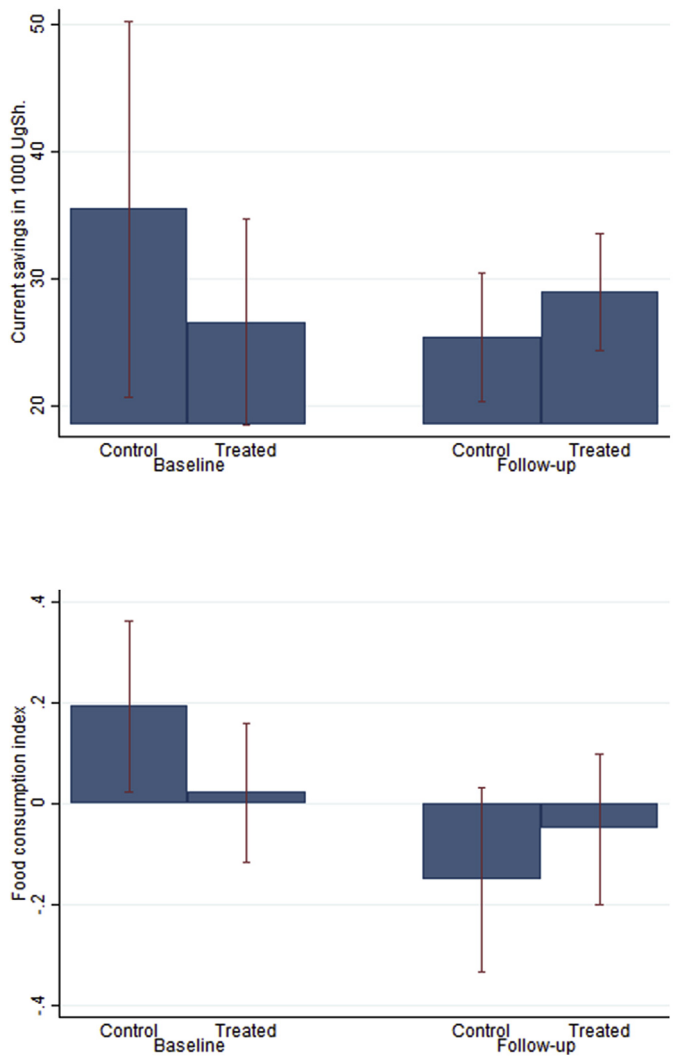
Numeracy score

You have 5 min to answer as many of the following calculation questions as possible. I will be keeping time with a stopwatch. Fill in the answer in the blank behind each question on the paper. Your answer needs to be readable. You will receive Sh. 200 for each question that you answer correctly. You are free to decide how much time you spend on each question, but you cannot change your answers after the 5 min are over. 4 min after the start of the quiz, I will remind you that 1 min is left. No electronic or other devices are allowed to be used; only the paper provided to you and the pen. You cannot be helped by anyone during the quiz. Help from other people in any form will invalidate your results, in which case we cannot pay you out. If you are running a business and a customer enters your business, we will take the paper and stop the time, so that you can help the customer and continue the quiz after the customer has left.

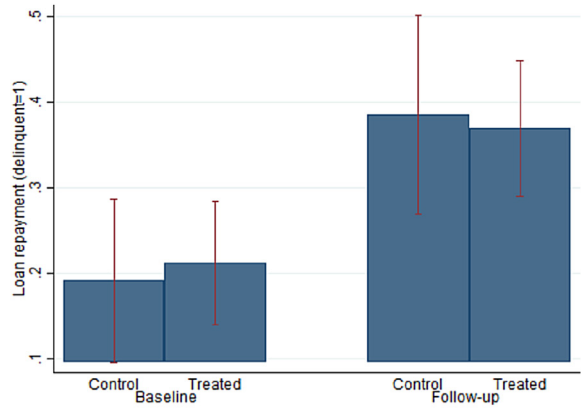
0 + 12 =
3 + 91 =
13 − 6 =
12 − 8 =
3 × 4 =
6 × 9 =
6/1 =
8/2 =
11 + 7 +2 =
14 + 25 + 3 + 15 =
72 − 23 =
89 − 14 =
13 × 7 =
5 × 12 =
56/7 =

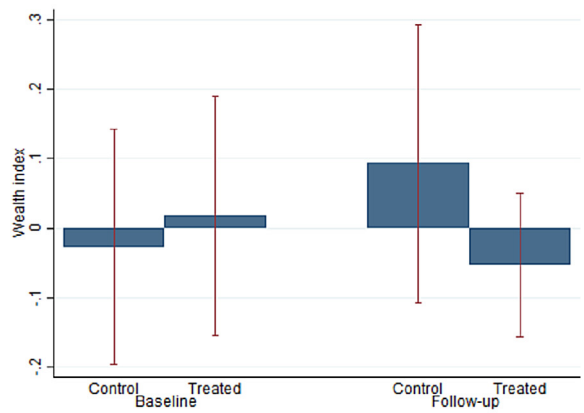
110/5 =
26 + 38 + 7 =
2 + 6 + 70 + 17 + 9 + 11 =
66 − 49 =
141 − 56 =
211 − 77 =
66 × 2 =
96 × 5 =
330/5 =
104/8 =
3 × 14.5 =
16 × 8.5 =
45/6 =

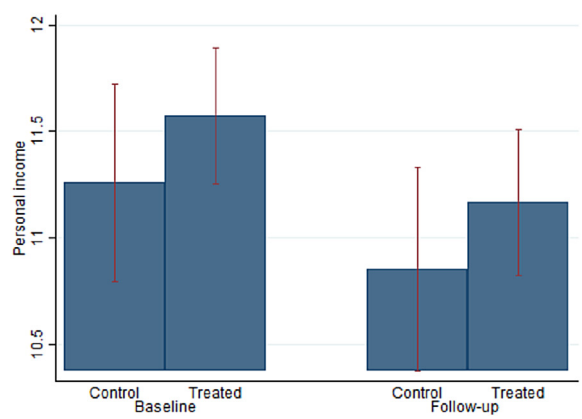
Appendix C. Difference-in-difference bar chart



Note: Means of outcome variables for treated and control groups at baseline and follow-up, with 95% confidence bars.







References

Akee, R., 2011. Errors in self-reported earnings: the role of previous earnings volatility and individual characteristics. J. Dev. Econ. 96 (2), 409–421.

Angrist, J.D., Pischke, J.S., 2008. Mostly Harmless Econometrics: an Empiricist's Companion. Princeton University Press.

Azzarri, C., Carletto, G., Covarrubias, K., De la O, A.P., Petracco, C., Kinnon, S., Zezzo, A., 2010. Measure for Measure: Systematic Patterns of Deviation between Measures of Income and Consumption in Developing Countries. Evidence from a New Dataset Technical Report.

Bandiera, O., Goldstein, M., Rasul, I., Burgess, R., Gulesci, S., Sulaiman, M., 2010. Intentions to participate in adolescent training programs: evidence from Uganda. J. Eur. Econ. Assoc. 8 (2–3), 548–560.

Beegle, K., De Weerdt, J., Friedman, J., Gibson, J., 2012. Methods of household consumption measurement through surveys: experimental results from Tanzania. J. Dev. Econ. 98 (1), 3–18.

Bruhn, M., McKenzie, D., 2009. In pursuit of balance: randomization in practice in development field experiments. Am. Econ. J. Appl. Econ. 1 (4), 200–232.

Carletto, G., 2013. Methodological Validation and Research on Root Groups. Collaboration LSMS Global Strategy: Measuring Casava Productivity in Zanzibar and Malawi Technical Report.

Collins, D., Morduch, J., Rutherford, S., Ruthven, O., 2009. Portfolios of the Poor: How the World's Poor Live on 2 USD a Day. Princeton University Press.

Crossley, T.F., Winter, J.K., 2013. Asking Households about Expenditures: what Have We Learned? Technical Report. National Bureau of Economic Research.

De Mel, S., Herath, D., McKenzie, D., Pathak, Y., 2016. Radio frequency (un) identification: results from a proof-of-concept trial of the use of RFID technology to measure microenterprise turnover in Sri Lanka. Dev. Eng. 1, 4–11.

De Mel, S., McKenzie, D.J., Woodruff, C., 2009. Measuring microenterprise profits: must we ask how the sausage is made? J. Dev. Econ. 88 (1), 19–31.
Grosh, M., Glewwe, P., 2000. Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study. World Bank, Washington, DC.
Gutknecht, D., 2011. Nonclassical Measurement Error in a Nonlinear (Duration) Model. University of Warwick, Department of Economics.
Hannagan, A., Morduch, J., 2015. Income Gains and Month-to-month Income Volatility: Household Evidence from the US Financial Diaries. NYU Wagner Research Paper No. 2659883.
Hoderlein, S., Winter, J., 2007. Recall Errors in Surveys Technical Report. Mannheim University.
Joshi, K., Hasan, R., Amoranto, G., 2011. Sinformal sector enterprises: some measurement issues. Rev. Income Wealth 57 (s1), 143–165.
Kamath, R., Ramanathan, S., 2015. Informal businesses and micro-credit–Evidence from financial diaries: a study in Ramanagaram, India. IIMB Manag. Rev. 27 (3), 149–158.
Kemsley, W., Redpath, R., Holmes, M., 1980. Family Expenditure Survey Handbook. Her Majesty's Stationary Office, London.
Kikafunda, J.K., Lukwago, F.B., 2005. Nutritional status and functional ability of the elderly aged 60 to 90 years in the Mpigi district of central Uganda. Nutrition 21 (1), 59–66.
Kolenikov, S., Angeles, G., 2009. Socioeconomic status measurement with discrete proxy variables: is principal component analysis a reliable answer? Rev. Income Wealth 55 (1), 128–165.
McKenzie, J., 1983. The accuracy of telephone call data collected by diary methods. J. Market. Res. 417–427.
Millimet, D.L., 2011. The elephant in the corner: a cautionary tale about measurement error in treatment effects models. Adv. Econom. 27 (1), 1–39.
Sankoh, A.J., Huque, M.F., Dubey, S.D., 1997. Some comments on frequently used multiple endpoint adjustment methods in clinical trials. Stat. Med. 16 (22), 2529–2542.