RUHR
ECONOMIC PAPERS

Nathan Fiala
Lise Masselus

Whom to ask? Testing Respondent Effects in Household Surveys

rwi #935

Imprint

Ruhr Economic Papers

Published by

RWI – Leibniz-Institut für Wirtschaftsforschung
Hohenzollernstr. 1-3, 45128 Essen, Germany

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Editors

Editorial Office

Ruhr Economic Papers #935

Responsible Editor: Manuel Frondel

ISSN 1864-4872 (online) – ISBN 978-3-96973-095-9

Ruhr Economic Papers #935

Nathan Fiala and Lise Masselus

Whom to ask? Testing Respondent Effects in Household Surveys

Bibliografische Informationen der Deutschen Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at **http://dnb.dnb.de**

RWI is funded by the Federal Government and the federal state of North Rhine-Westphalia.

http://dx.doi.org/10.4419/96973095
ISSN 1864-4872 (online)
ISBN 978-3-96973-095-9

Nathan Fiala and Lise Masselus[1]

Whom to ask? Testing Respondent Effects in Household Surveys

Abstract

The common practice in household questionnaires of surveying the most knowledgeable household member can lead to inaccurate data if they have limited information. Using survey experiments in Paraguay and Uganda, we investigate whether there are discrepancies in intra-household reporting on income and consumption when multiple household members are interviewed. We use data from 4,100 households where we randomly vary whether the survey is administered to one spouse only, both spouses together or both spouses separately. We do not find meaningful systematic differences in the mean or distribution of household income and consumption and conclude that the magnitude of respondent effects for these variables is unlikely to bias most empirical analyses. However, a within-household analysis reveals large, but mostly unsystematic, reporting discrepancies. Taken together, the results indicate that respondent selection may matter for obtaining accurate information for a given household, but not for aggregate analysis of households.

JEL-Code: O1, C8, J16, D13, I32

Keywords: Survey methods; respondent effects; proxy reporting; intra-household economics

January 2022

1 Nathan Fiala, University of Connecticut and RWI; Lise Masselus, RWI and University of Passau .– All correspondence to: Lise Masselus, RWI, Hohenzollernstr. 1/3, 45128 Essen, Germany, e-mail: lise.masselus@rwi-essen.de

1. Introduction

The surge in primary data collection through household surveys is accompanied by an increasing need to understand how to collect accurate data through interviews (De Weerdt et al., 2020; Dillon et al., 2020; McKenzie and Rosenzweig, 2012). Survey design decisions, such as whom to survey or what to ask, can have substantial implications, which may affect conclusions from econometric analyses and key stylized facts of development (De Weerdt et al., 2020). Our study provides experimental evidence on the effects of respondent selection as a potential source of measurement error. As outcome variables, we look at income and consumption, two key economic indicators used by researchers and policymakers.

Our goal is to document any discrepancies between three common methods of respondent selection and identify a cost-effective way to obtain high quality data. We implement a household questionnaire in a sample of 4,100 loan applicants in rural Uganda and Paraguay. We randomly vary whether the survey is administered to one spouse only, both spouses together or both spouses separately. This allows us to compare whether reported income and consumption vary between individual and household interviews.

The most common interview approach is to survey the most knowledgeable household member, often the household head, who then serves as proxy for other household members. This is done to minimize survey costs and non-response, and because it is often assumed that information flows perfectly within the household. However, surveying only one household member could potentially lead to measurement error because of intentional hiding, selective reporting or incomplete or erroneous information within the household. These concerns can be solved by surveying multiple household members who report for themselves, an approach that is generally considered as best practice (de Nicola and Giné, 2014; De Weerdt et al., 2020). Intuitively, self-reporting may result in more accurate information but comes at a logistical and monetary cost compared to reporting by proxy. It is also not clear whether several household members should be surveyed together or separately from each other. While interaction between spouses in a joint interview may be beneficial, information hiding between spouses could still result in biased responses. Instead, privately administered interviews with multiple household members may be ideal.

We find limited sensitivity of household income and consumption to respondent selection. For aggregate income, we find small effect sizes that are insignificant at conventional levels. We also do not see a difference in the distribution of aggregate income. For consumption (administered in Uganda only), we do not see differences between joint surveys and surveys with one respondent alone. However, respondents who are surveyed separately from their partner report lower consumption compared to respondents alone or couples together, though the effect size is small. Together, these findings suggest that respondent effects are unlikely to bias empirical analyses that focus on average treatment effects and so the most cost-effective survey strategy is fine.

Nevertheless, while we do not observe large systematic differences in the mean and variance of the aggregates, we do find evidence that is in line with imperfect intra-household information flow. When looking at reports from spouses in the same household, we find large discrepancies between members of the same household, though these do not seem to be systematic. Additionally, when we assess how spouses report on each other's income, we find that survey errors are moving in opposing directions and are cancelling each other out for aggregate income. Our data is consistent with each respondent underreporting their own income and overreporting their spouse's income in the joint survey. This suggests that respondent selection may matter for obtaining accurate information through a household questionnaire, depending on the objective of the data collection.

To inform survey design choices, we also document the tracking rates and survey costs of each approach. Our calculations show that, in both study settings, there is a trade-off between costs and tracking rates when surveying multiple household members. Separate surveys lead to higher tracking rates, but they also impose the highest survey costs per household, compared to joint surveys. While costs and tracking rates will differ across study settings, this analysis allows readers to assess potential trade-offs for respondent selection.

Finally, as enumerators were randomly matched with respondents, we also shed light on enumerator effects and survey design effects as potential sources of measurement error that may interact with the identified respondent effects. Similar to Di Maio and Fiala (2020), we find no evidence of enumerator effects. We also argue that survey design effects are unlikely to affect our findings. We end by elaborating on some mechanisms that may explain respondent effects.

This article adds to a burgeoning literature that explores survey design issues in the Global South [1]. Apart from respondent effects, this literature addresses questionnaire design issues, survey mode effects and enumerator effects [2]. Typically, the effects of respondent selection are evaluated by relying on comparisons within the same household. In these studies, two spouses are surveyed separately to assess if reporting on female bargaining power, income or property values differs (Ambler et al., 2021a; Chen and Collins, 2014; de Nicola and Giné, 2014; Doss et al., 2018; Fisher et al., 2010). This design is interesting as it allows researchers to estimate discrepancies and their determinants within the same household. However, the main caveat of interviewing multiple household members is that it could lead to conditioning by either the enumerator or the respondents, and hence, this approach does not reflect a business-as-usual scenario where one household member is surveyed (De Weerdt et al., 2020). Survey experiments are needed to assess the causal effect of respondent selection and to understand how discrepancies affect the averages across treatment arms.

To date, there are only few experimental studies assessing the impact of respondent selection in developing countries, and the studies that exist, draw mixed conclusions. Bardasi et al. (2011); Dillon et al. (2012) and Serneels et al. (2017) randomly vary the respondent (self or proxy) in a large-scale survey experiment in Tanzania. They find that proxy reporting has no effect on reporting of female employment rate but substantially lowers the male employment rate (Bardasi et al., 2011). In contrast, for child labor reporting and returns to education, respondent selection does not seem to play an important role (Dillon et al., 2012; Serneels et al., 2017). Kilic and Moylan (2016) test five approaches to respondent selection and conclude that the respondent type influences the individual-level measurement of ownership and control of assets. We add to these experimental studies using data from two large samples in diverse countries. Rather than comparing between proxy and self-reporting, we introduce experimental variation using three data collection methods used in practice. Finally, our data also allows us to assess within-household discrepancies and therefore also offers a useful complement to the non-experimental literature.

The structure of this article is as follows. The next section discusses the experimental design

[1] While for a long time most of the existing evidence stemmed from high-income countries (e.g. Sudman and Bradburn (1974)), this and other studies are motivated by the desire to understand the implications of the main strategies commonly used for collecting data in the Global South.

[2] De Weerdt et al. (2020) and Carletto et al. (2021) provide a comprehensive overview of this literature. For studies on survey design effects, see for example Abay et al. (2021); Ambler et al. (2021a); Beaman and Dillon (2012); Dillon et al. (2012); Beegle et al. (2012a,b); de Mel et al. (2009); de Nicola and Giné (2014); Heath et al. (2021); Laajaj and Macours (2019); Nillesen et al. (2021); Peterman et al. (2021); Reitmann et al. (2020). For survey mode effects, see Garlick et al. (2020). For enumerator effects, Di Maio and Fiala (2020).

and compliance rates. Section 3 outlines the econometric framework and results. Section 4 digs into potential confounders and mechanisms and section 5 concludes.

2. Experimental design

2.1. Study sample and experimental design

Data for this study was collected during the baseline survey for an ongoing experimental evaluation of microfinance programming in rural Paraguay and Northern Uganda. Respondents were pre-identified as farmers and entrepreneurs who expressed interest in a microloan. In Uganda, loan applicants are mostly subsistence farmers who applied for an agricultural loan. In Paraguay, loan applicants are female, own a business or intend to start one, and expressed interest in receiving a business loan. The sample for the survey experiment consists of all loan applicants who have a spouse or partner, leaving us with 2,564 loan applicants in Uganda and 1,576 in Paraguay.

We developed a survey module in which we ask the respondent(s) to report on the different sources of income for every household member during the past 4 weeks and a module on household consumption in the past 7 days. We randomly assigned a respondent to this survey module using three treatment arms. Figure 1 presents the study design. In the first treatment arm (T1), the survey is administered to the loan applicant only. In the second treatment arm (T2), the survey is conducted with the loan applicant and his/her spouse in a joint interview. In a third treatment arm (T3) both spouses are surveyed separately, which means that we collect two data points for each household. Due to budget restrictions, T3 and the consumption questions were implemented in Uganda only. Appendix I and II provide details about the survey instrument, along with a description of how field work was organized.

Our study design was pre-specified and registered under the American Economic Association RCT Registry[3].

	Paraguay	Uganda
Survey modules	*Income*	*Income* *Consumption*
T1 Loan applicant only	108 clusters n = 1063 *100% compliers*	100 clusters n = 1283 or *100% compliers*
T2 Joint interview	57 clusters n = 513 *70.8% compliers*	50 clusters n = 649 *76.9% compliers*
T3 Separate interview		50 clusters n = 632 *89.4% compliers*

Figure 1: Experimental design

[3]ID AEARCTR-0005162. The pre-analysis plan outlines the study design in Uganda and the main hypothesis for income.

2.2. Compliance rates and costs

Tracking is an important component of data quality. In data analysis, low tracking rates result in a loss of statistical power and can lead to unbalanced samples. Potentially, more accurate data through surveying multiple household members may come at the expense of internal validity due to low tracking rates. To illustrate this potential tracking-accuracy trade-off, we document the tracking rates of each method. Ultimately, the decision on who to survey should depend on these trade-offs and the intended use of the data (Dillon et al., 2020).

Figure 1 shows the compliance rates for each treatment arm. Despite substantive efforts, tracking rates are imperfect. For the joint survey (T2), we tracked a second respondent in 77% of the households in Uganda and 71% in Paraguay. Tracking respondents for the separate interviews (T3) was easier, as they did not have to be available at the same time. The compliance rate for separate interviews in Uganda was 89%. If no second respondent was available, we surveyed the loan applicant alone.

Tracking respondents has also proven to be challenging in other studies. Kilic and Moylan (2016), who have a set-up similar to ours, document difficulties to track respondents. They have compliance rates that vary between 30 to 60% per treatment arm[4].

We also estimate the cost of a survey in each treatment arm by differentiating between costs resulting from increased survey duration and costs related to tracking. The aggregate cost differences between the different treatment arms will vary depending on the local context, the organization of field work and the ease of tracking spouses. In Uganda and Paraguay respectively, joint surveys (T2) cost 13% and 14% more, compared to surveying only one respondent in T1. In Uganda, the costs for separate surveys (T3) are 33% higher, compared to T1 (see Appendix III for a detailed description). These cost calculations reveal an important trade-off between costs and tracking. While surveys in T3 are most expensive, they also resulted in higher tracking rates.

2.3. Benchmark income

One shortcoming of this and many other survey experiments is that we do not have a validation or gold standard measure of the outcome variables. Unfortunately, approximating the true value for income or consumption is nearly elusive for respondents with little to no record keeping (De Weerdt et al., 2020). In a developing country context, gold standard measures are sometimes constructed through diaries (e.g., Beegle et al. (2012b) for consumption data and de Mel et al. (2009) for profit data) or by using administrative data (e.g., de Nicola and Giné (2014) for income data). However, these approaches are extremely costly, not entirely error-free, and often suffer from high non-compliance (De Weerdt et al., 2020; Kilic and Moylan, 2016). Due to the absence of a true gold standard, we cannot speak to the nature of measurement errors.

We construct a benchmark for household income in Uganda by using two data points for the same household in T3. We add up income earned by female household members, reported by the wife and income earned by male household members, reported by the husband[5]. We assume that each respondent reports most truthfully on his or her own income, when being surveyed privately. Additionally, we assume that household members of the same gender have the most knowledge

[4]Own calculations based on Table 2 in Kilic and Moylan (2016)

[5]For consumption, we do not have data disaggregated by household member, and are not able to construct a similar benchmark.

of each other's income[6]. While our constructed benchmark for household income gives a sense of the direction of misreporting, we cannot interpret this as a gold standard measure as we cannot rule out that other reporting errors, such as recall errors, telescoping or survey fatigue, affect the reported levels.

3. Econometric framework and results

3.1. Econometric framework

To test whether spouses report different household income and consumption when they are surveyed alone, together or separately, we compare the data reported by loan applicants surveyed with their partner in T2, with the data reported by loan applicants in T1 and spouses separately in T3 (Uganda only). To accommodate for imperfect compliance, we estimate the following intention-to-treat (ITT) effect:

$$Y_{ia} = \alpha + \beta_1 T_{1ia} + \beta_3 T_{3ia} + \delta X_i + \epsilon_{ia} \tag{1}$$

where Y_{ia} represents the measure for income or consumption for household i in village a. A detailed description of our outcome variables can be found in Appendix VI. T_{1ia} and T_{3ia} are dummy variables with value one if the household is allocated to T1 or T3 (Uganda only), respectively. The control group is T2 (joint interview). Treatment is clustered at the village level for logistical and ethical considerations and randomization is stratified at the subcounty (Paraguay) or district (Uganda) level. We adjust standard errors ϵ_{ia} for clustering at the village level and add region fixed effects to account for stratification. The coefficients of interest are β_1 and β_3.

We test for balance on socio-demographic variables and, as expected, we find that the randomization resulted in a well-balanced sample (Appendix IV-V).

Ex-post power calculations for our main outcome variables reveal that have sufficient power to detect meaningful changes. Minimum Detectible Effect (MDE) sizes for aggregate income range from 22-32% of the control group mean in Uganda and 17-25% in Paraguay. For consumption, the MDE is 9-12%.

3.2. Main results on household income and consumption

We present the results for the impact of each treatment arm on household income and consumption in Table 1. For both countries, we do not find statistically significant differences in household income between the individual survey (T1) and the joint survey (T2). In Uganda, we also do not see statistically significant differences between our benchmark measure of income (T3) and T1 or T2, though the p-value for T1-T3 is 0.14, which is rather close to conventional levels of statistical significance. The estimated effect sizes are small in magnitude and insignificant at conventional confidence levels.

For household consumption, we find that loan applicants report lower household consumption in T3 compared to T2 or T1. The effect size represents a small effect that corresponds to 6.7% compared to the control group mean and has a p-value of 0.03. We do not see any differences between household consumption reported by loan applicants in T1 and household consumption reported by both spouses together in T2 (p-value = 0.65).

[6]The former assumption is based on several other studies (Bardasi et al., 2011; De Weerdt et al., 2020; Kilic and Moylan, 2016). The latter assumption is a reasonable assumption given the dynamics of gender relations in Northern Ugandan households.

These results are robust to imputation and winsorization and to using different constructs for T3 income (Table A6-A7 in Appendix VI).

Table 1: Main treatment effects for household income and consumption

	Household income		Household consumption
	(1)	(2)	(3)
	Uganda	Paraguay	Uganda
Loan applicant only (T1)	-11.029	-34.910	-1.157
	(18.87)	(54.63)	(2.53)
Spouses separately (T3)	24.579		-6.072**
	(26.14)		(2.76)
Region FE	Yes	Yes	Yes
Control (T2) Mean	275.338	1056.919	90.009
N	2564	1578	2564
R2	0.021	0.012	0.039
P-value T1 = T3	0.144		0.033
P-value T1 = T2	0.560	0.524	0.648
P-value T2 = T3	0.348		0.029

This table reports the OLS regression results for different treatments on household income and consumption. Outcome variables are converted to USD PPP and winsorized at the 99th percentile by treatment arm. SE in parenthesis and clustered at the village level. ***, ** and * indicate significance at the 1, 5 and 10 percent critical level.

Apart from differences in means across treatment arms, discrepancies could also appear in other moments of the distribution. We construct kernel density functions for household income and consumption for the different treatment arms and conduct Kolmogorov-Smirnov equality-of-distribution tests to assess whether the distributions across treatment arms are equal. The figures and test statistics (Shown in Appendix VII) indicate that the spread of household income does not vary across treatment arms. The p-values for the Kolmogorov-Smirnov test range from 0.15 to 0.98. For household consumption, we find a difference between the distribution of consumption for T3, versus the other two treatment arms (p-value 0.01-0.02). Again, there does not seem to be a systematic difference in the distribution of T1 and T2 (p-value 0.76).

Taken together, these findings suggest that household income and consumption are quite robust to respondent selection. Any differences we find are small when compared to the T2 mean and not statistically significant. For household income, we conclude that surveying only one household member would be a cost-effective approach for obtaining an estimate that does not introduce bias, compared to surveys with multiple household members. For consumption, this is less clear.

3.3. Within-household analysis

Hitherto, most studies on respondent effects relied on interviews with two spouses to document high levels of discrepancy within a household, that are often systematic when aggregating by gender (Ambler et al., 2021a; de Nicola and Giné, 2014; Doss et al., 2018; Fisher et al., 2010). Our findings complement these studies: while we do not find large differences in the average levels across the three treatment arms, non-systematic or random discrepancies within the household could still exist. This has important implications for policymakers and researchers interested in the levels of certain variables, for example when questionnaires are used for program targeting. Additionally, conducting a within-household analysis helps to gauge the validity of using means.

We use data from both spouses in 547 households in T3 to compare within-household discrepancies in reported household income and consumption. We calculate the percentage difference in household income and consumption as the difference between the value reported by the wife minus the value reported by the husband, divided by the average of the two values. Figure 2 pictures the results for income and consumption. Each bar shows the percentage of couples (y-axis) that differ a certain percentage of income/consumption (x-axis) in their reports. Values to the left of zero indicate higher male reports. Values to the right indicate higher female reports. Values close to zero imply that the husband's and wife's reporting is almost identical. The vertical lines show the mean percentage difference. We find high levels of disagreement between income reported by the wife and husband in the same household, but no systematic under or over-reporting as the mean is close to zero and the spread is more or less similar to the right and left of zero. For household consumption, the percentage difference between two spouses is smaller, and again, we do not see systematic higher reports by spouses from a particular gender.

Figure 2: Difference between husbands' and wives' reports of household income and consumption

This figure presents the values of discrepancies in reporting of household income between men and women from the same household. Differences in income and consumption are calculated as (value reported by wife − value reported by husband)/[(value reported by wife + value reported by husband)/2]. Values of 200% occur when one of the spouses reported 0 income while these other spouse did not.

4. Confounders and mechanisms

Our goal is not merely to document levels of reporting, but also to shed light onto why discrepancies may occur. We first elaborate on enumerator effects and questionnaire design effects as two potential confounders and then discuss potential mechanisms.

4.1. Confounders

Enumerator effects occur when enumerator characteristics, behavior, and skills influence survey responses. As enumerators were randomly assigned to respondent households, we can test for the presence of enumerator effects by estimating how much variation in the observed responses is explained by individual enumerators. We estimate the R^2 in an enumerator fixed effects regression and find very low values for the R^2 for our main outcome variables, ranging from 0.004 to 0.071. We conclude that enumerator effects have little to no impact on the reported outcome variables in our experiment. Following Di Maio and Fiala (2020), we also add enumerator fixed effects to our main regressions and find that our main results hold (see Appendix VIII).

Questionnaire design could also affect responses (Abay et al., 2021; Ambler et al., 2021b; Kilic and Sohnesen, 2019; Laajaj and Macours, 2019). To rule out survey effects as potential confounders, similar survey instruments were administered across treatment arms, to the extent possible. In Uganda, the position of the income section was consistent across the treatment arms to rule out survey fatigue and the position of questions as a confounder. In Paraguay, we cannot disentangle the respondent effect from the survey fatigue effect as the position of the income section varied across treatment arms.

4.2. Mechanisms

While we find similar effects across two diverse samples, we are careful in translating our findings to other settings. Understanding the cognitive activities behind the survey processes is insightful to gauge whether or not our results extrapolate to other contexts (De Weerdt et al., 2020). We discuss three mechanisms that could drive differences between treatment arms.

A first mechanism that could explain reporting discrepancies is different levels of cognitive effort across the treatment arms. It is possible that loan applicants unintentionally exert varying levels of effort when they are surveyed alone, versus with their spouse versus alone but knowing that their spouse would be surveyed after. Additionally, for variables that suffer from rounding up errors or recall bias, two memories may be better able to recall information compared to one. Finally, the main respondent may already be fatigued or primed after answering a full questionnaire, while having the spouse join in could lead to a renewed cognitive effort that may lead to more accurate responses.

A second mechanism is strategic misreporting, which can occur when respondents over- or underreport their income to increase their chances of participating in a program. Our sample may be particularly prone to strategic misreporting as all respondents are applying for a microloan. We employed several strategies to minimize this risk. Loan groups were already formed at this stage and respondents were aware that only randomization would determine if their village was to be included in the treatment group. We thus informed respondents that the survey would not influence their chances of receiving a loan, and we believe this was believed by respondents. Moreover, if strategic misreporting were present in our sample, we would expect it to affect loan applicants more than their spouses. Since there was no consultation between spouses in T2 and T3, and since we do not find large discrepancies across the treatment arms, we conclude that strategic misreporting is unlikely in our study.

A last mechanism is asymmetric information and income hiding. We test for this in three ways. First, we regress each spouse's income on treatment status[7]. Table 2, columns 1 and 2 compare

[7] In Uganda, both spouses were surveyed separately in T3. We use data from the survey with the loan applicant.

the loan applicant's income as reported by the respondent alone (T1), with the main respondent's income as reported by the spouses together (T2). We find that the respondent reports higher own income when being surveyed alone, compared to being surveyed in a joint survey. Columns 3 and 4 show the spouse's income as reported by the respondent alone, versus the spouses together (T2). We see that the spouse income is lower when reported by the main respondent, compared to the joint survey. This data is consistent with the idea that respondents underreport their own income and overreport their spouse's income in the joint survey, compared to interviews with the loan applicant only. The first effect could be driven by income hiding, whereas the second effect could be caused by information asymmetries in the household. Interestingly, in magnitude, the two errors offset each other, which produces an accurate estimate of aggregate household income.

Table 2: Treatment effects for respondent and spouse income (Uganda and Paraguay)

	Loan applicant income		Spouse income	
	(1)	(2)	(3)	(4)
	Uganda	Paraguay	Uganda	Paraguay
Loan applicant only (T1)	11.065	53.412**	-14.923*	-69.259*
	(14.82)	(26.09)	(7.99)	(35.61)
Spouses separately (T3)	1.118		-11.007	
	(17.90)		(9.34)	
Region FE	Yes	Yes	Yes	Yes
Control (T2) Mean	166.415	316.301	83.277	587.773
N	2564	1578	2564	1578
R2	0.019	0.011	0.014	0.017
P-value T1 = T3	0.518		0.641	
P-value T1 = T2	0.456	0.042	0.063	0.053
P-value T2 = T3	0.950		0.240	

This table reports the OLS regression results for different treatments on income. The dependent variable is either loan applicant income or spouse income. Outcome variables are converted to USD PPP and winsorized at the 99th percentile. SE in parenthesis and clustered at the village level. ***, ** and * indicate significance at the 1, 5 and 10 percent critical level.

In Paraguay, the main respondent is always female. The findings in Table 2 imply that the wife's income is lower in joint surveys, as opposed to interviews with the wife only. The male income, on the other hand, is higher in joint surveys, as opposed to surveys with the wife alone. In Uganda, the main respondent is either male or female. The underreporting of own income and overreporting of spouse's income in the joint survey is driven mostly by male loan applicants (Appendix IX).

Second, we asked respondents whether they hide income from their spouse or suspect their spouse to hide income from them. 7% of the respondents in Paraguay and 12% in Uganda stated that they hide income from their spouse, while 8% (Paraguay) and 16% (Uganda) suspect their spouse to hide income from them. We look at heterogeneous treatment effects with regard to income hiding (Appendix X) and find that income hiding may be part of the explanation for the different income levels reported by the respondent and spouse in Uganda, but not in Paraguay.

Third, we look at which sources of income are more prone to reporting discrepancies. In general, we expect regular and stable sources of income, such as income from wage employment, to vary less across respondents. Farm income, on the other hand, is seasonal but often earned by multiple household members, and thus, less prone to information asymmetries within the household. In contrast, irregular and private sources of income, such as income from business activities, are expected to be most prone to misreporting.

We decompose respondent and spouse income by activity to assess which activities drive the above discrepancies. We are underpowered for this analysis and are careful in drawing strong conclusions. This exploratory analysis shows that, for spouse income, higher values in joint surveys are mainly driven by business income in both countries. In Uganda, the spouses' business income is 31% higher in the joint survey, compared to spouse income reported by the loan applicant. In Paraguay, business income earned by the loan applicant's spouse is 135% higher in the joint survey, compared to the survey with the loan applicant alone. Both effects are statistically significant at the 5% level (Appendix XI).

5. Discussion and conclusion

To date, there is limited understanding of the implications of survey design choices for the measurement of key welfare indicators. Interviewing one household member as a proxy is logistically easier and less costly, but potentially more prone to measurement error. The trade-off between joint versus separate reporting is less clear: surveying spouses separately potentially leads to more truthful reporting but comes at a higher survey cost.

Using a large-scale survey experiment in Paraguay and Uganda, we conclude that the presence of spouses in household surveys, either in a joint or in a separate interview, is unlikely to affect most empirical analyses. At the same time, we do find large discrepancies between spouses from the same household. Moreover, when looking into proxy effects, we find misreporting of disaggregated income, with survey errors moving into different directions and cancelling each other out. Taken together, the results indicate that respondent selection may matter for obtaining accurate information for a given household, but not for the aggregate analysis of households.

Ultimately, respondent selection for household surveys should depend on the research objective. For researchers interested in estimating causal relationships, the question of whom to survey will depend on the research question. The differences in aggregates across the different treatment groups are unlikely to affect most types of analyses. This implies that the most cost-effective survey strategy can often be used. However, when asking about income earned by separate household members, respondent selection may matter. As a first step, researchers should ensure balance with regard to respondent selection between any comparison group in order to reduce the likelihood of bias. An imbalance could occur if the survey strategy consists of surveying multiple household members, and the availability of household members correlates with some kind of treatment status. Another scenario where an imbalance could occur is when researchers allow for respondents to consult with their spouses. Again, in cases where the availability or willingness of spouses to collaborate correlates with a treatment status, this could cause a systematic difference in reporting individual income. Simply controlling for respondent selection would not solve the problem.

For policymakers who are interested in the levels of income or consumption, our findings suggest that data needs to be interpreted cautiously as the levels of income and consumption may differ depending on whom you ask. This may be problematic when household-level data is used for program targeting or poverty analysis at the household level. At the very least, researchers should be clear about how respondents were selected, and maintain consistency where possible.

Acknowledgements

We thank Frédéric Cochinard for excellent research assistance and Innovations for Poverty Action (Paraguay) and Filder Aryemo and her team from Gaplink (Uganda) for their support and dedication to collecting high-quality data. We also thank all study participants for generously giving their time. We are grateful for the comments and feedback provided by our colleagues at RWI – Leibniz Institute for Economic Research and participants of the IPA/GPRL Methods and Measurement conference 2021. In particular, we would like to thank Gunther Bensch, Steve Glazerman, Michael Grimm, Florian Neubauer, Julian Rose, and Jörg Peters for their thorough feedback and Cecilia Michelle Argueta for proofreading the paper.

Funding

The data collection for this work was funded by a special grant from the Research Methods Initiative by Innovations for Poverty Action.

References

Abay, K. A., Berhane, G., Hoddinott, J. F. and Tafere, K. (2021). Assessing Response Fatigue in Phone Surveys Experimental Evidence on Dietary Diversity in Ethiopia.
URL: *http://hdl.handle.net/10986/35483*

Ambler, K., Doss, C., Kieran, C. and Passarelli, S. (2021a). He says, she says: Spousal disagreement in survey measures of bargaining power, *Economic Development and Cultural Change* **69**(2): 765–788.
URL: *https://doi.org/10.1086/703082*

Ambler, K., Herskowitz, S. and Maredia, M. K. (2021b). Are we done yet? Response fatigue and rural livelihoods, *Journal of Development Economics* **153**.
URL: *https://doi.org/10.1016/j.jdeveco.2021.102736*

Bardasi, E., Beegle, K., Dillon, A. and Serneels, P. (2011). Do labor statistics depend on how and to whom the questions are asked? Results from a survey experiment in Tanzania, *World Bank Economic Review* **25**(3): 418–447.
URL: *https://doi.org/10.1093/wber/lhr022*

Beaman, L. and Dillon, A. (2012). Do household definitions matter in survey design? Results from a randomized survey experiment in Mali, *Journal of Development Economics* **98**(1): 124–135.
URL: *https://doi.org/10.1016/j.jdeveco.2011.06.005*

Beegle, K., Carletto, C. and Himelein, K. (2012a). Reliability of recall in agricultural data, *Journal of Development Economics* **98**(1): 34–41.
URL: *https://doi.org/10.1016/j.jdeveco.2011.09.005*

Beegle, K., De Weerdt, J., Friedman, J. and Gibson, J. (2012b). Methods of household consumption measurement through surveys: Experimental results from Tanzania, *Journal of Development Economics* **98**(1): 3–18.
URL: *https://doi.org/10.1016/j.jdeveco.2011.11.001*

Carletto, C., Dillon, A. and Zezza, A. (2021). Agricultural data collection to minimize measurement error and maximize coverage, *Handbook of Agricultural Economics* (July).
URL: *https://doi.org/10.1016/bs.hesagr.2021.10.008*

Chen, J. J. and Collins, L. A. (2014). Let's talk about the money: Spousal communication, expenditures, and farm production, *American Journal of Agricultural Economics* **96**(5): 1272–1290.
URL: *https://doi.org/10.1093/ajae/aau059*

de Mel, S., McKenzie, D. and Woodruff, C. (2009). Measuring microenterprise profits: Must we ask how the sausage is made?, *Journal of Development Economics* **88**(1): 19–31.
URL: *https://doi.org/10.1016/j.jdeveco.2008.01.007*

de Nicola, F. and Giné, X. (2014). How Accurate Are Recall Data ? Evidence from Coastal India, *Journal of Development Economics* **106**: 52–65.
URL: *https://doi.org/10.1016/j.jdeveco.2013.08.008*

De Weerdt, J., Gibson, J. and Beegle, K. (2020). What Can We Learn from Experimenting with Survey Methods ?, *Annual Review of Resource Economics* **12**: 431–450.
URL: *https://doi.org/10.1146/annurev-resource-103019-105958*

Di Maio, M. and Fiala, N. (2020). Be Wary of Those Who Ask: A Randomized Experiment on the Size and Determinants of the Enumerator Effect, *World Bank Economic Review* **34**(3): 654–669.
URL: *https://doi.org/10.1093/wber/lhy024*

Dillon, A., Bardasi, E., Beegle, K. and Serneels, P. (2012). Explaining variation in child labor statistics, *Journal of Development Economics* **98**(1): 136–147.
URL: *https://doi.org/10.1016/j.jdeveco.2011.06.002*

Dillon, A., Karlan, D., Udry, C. and Zinman, J. (2020). Good identification, meet good data, *World Development* **127**(1).
URL: *https://doi.org/10.1016/j.worlddev.2019.104796*

Doss, C. R., Catanzarite, Z., Baah-Boateng, W., Swaminathan, H., Diana Deere, C., Boake-Yiadom, L. and Suchitra, J. Y. (2018). Do men and women estimate property values differently?, *World Development* **107**: 75–86.
URL: *https://doi.org/10.1016/j.worlddev.2018.02.012*

Fisher, M., Reimer, J. J. and Carr, E. R. (2010). Who should be interviewed in surveys of household income?, *World Development* **38**(7): 966–973.
URL: *https://doi.org/10.1016/j.worlddev.2009.11.024*

Garlick, R., Orkin, K. and Quinn, S. (2020). Call Me Maybe: Experimental Evidence on Frequency and Medium Effects in Microenterprise Surveys, *The World Bank Economic Review* **34**: 418–443.
URL: *https://doi.org/https://doi.org/10.1093/wber/lhz02*

Heath, R., Mansuri, G., Rijkers, B., Seitz, W. and Sharma, D. (2021). Measuring Employment : Experimental Evidence from Urban Ghana, *The World Bank Economic Review* **35**(3): 635–651.
URL: *https://doi.org/10.1093/wber/lhaa014*

Kilic, T. and Moylan, H. (2016). *Methodological Experiment on Measuring Asset Ownership From a Gender Perspective.*
URL: *ttps://openknowledge.worldbank.org/handle/10986/33653*

Kilic, T. and Sohnesen, T. P. (2019). Same Question But Different Answer: Experimental Evidence on Questionnaire Design's Impact on Poverty Measured by Proxies, *Review of Income and Wealth* **65**(1): 144–165.
URL: *https://doi.org/10.1111/roiw.12343*

Laajaj, R. and Macours, K. (2019). Measuring Skills in Developing Countries, *Journal of Human Resources* **56**(4): 1–46.
URL: *https://doi.org/10.3368/jhr.56.4.1018-9805R1*

McKenzie, D. and Rosenzweig, M. (2012). Symposium on measurement and survey design, *Journal of Development Economics* **98**(1): 1–2.

Nillesen, E., Grimm, M., Goedhuys, M., Reitmann, A. K. and Meysonnat, A. (2021). On the malleability of gender attitudes: Evidence from implicit and explicit measures in Tunisia, *World Development* **138**: 105263.
URL: *https://doi.org/10.1016/j.worlddev.2020.105263*

Peterman, A., Schwab, B., Roy, S., Hidrobo, M. and Gilligan, D. O. (2021). Measuring women's decisionmaking: Indicator choice and survey design experiments from cash and food transfer evaluations in Ecuador, Uganda and Yemen, *World Development* **141**: 105387.

URL: *https://doi.org/10.1016/j.worlddev.2020.105387*

Reitmann, A. K., Goedhuys, M., Grimm, M. and Nillesen, E. E. (2020). Gender attitudes in the Arab region – The role of framing and priming effects, *Journal of Economic Psychology* **80**(April): 102288.
URL: *https://doi.org/10.1016/j.joep.2020.102288*

Serneels, P., Beegle, K. and Dillon, A. (2017). Do returns to education depend on how and whom you ask?, *Economics of Education Review* **60**(July): 5–19.
URL: *https://doi.org/10.1016/j.econedurev.2017.07.010*

Sudman, S. and Bradburn, N. M. (1974). *Response Effects in Surveys: A Review and Synthesis.*

Appendix

I. Survey design

The income survey consists of questions on the income of each household member from a detailed set of activities. The sources of income comprise crop farming, livestock farming, salaried employment, occasional employment, and self-employment. Transfer payments, either from public or private sources are also included. We additionally asked about income from rent, infrequent sale of assets, and bushmeat selling in Uganda. For each household member and each source of income, we asked the following question: *"How much money did [household member x] earn as income from [source y] in the last 4 weeks? By income, we mean all revenue minus any expenses for that activity"*. The primary variable of interest is aggregate household income, which we calculate by adding up the reported income earned in the past 4 weeks from all sources of income, for all household members.

In Uganda, we also asked about household consumption in the past 7 days. To obtain an estimate of the total value of household consumption, we looped through a list of 28 food items and asked whether anyone in the household consumed this item in the past 7 days, either inside or outside the household. Then, we asked to provide the value of consumption if the food item was consumed. Household consumption is estimated by adding up the values for all food items.

II. Field protocol

The data was collected between March and August 2019 in Paraguay, and between January and March 2020 in Uganda.

This survey experiment is part of an ongoing RCT on microcredit. The total sample of the microcredit RCT consists of 3,064 male and female loan applicants in Northern Uganda and 2,035 female loan applicants in rural Paraguay. The study sample for this survey experiment was reduced to all household with a spousal pair. In polygamous households in Uganda, we focused on one principal couple. In the vast majority of cases, the main spouses were considered to be the male household head and the first wife. In less than 1% of the couples, the second or third wife were surveyed. In Paraguay, we surveyed the loan applicant alone if spouses were not available. In Uganda, protocol allowed replacement of the spouse by another knowledgeable household member[1].

To organize the field work efficiently, we designed two separate survey instruments. First, the main survey was always conducted with the loan applicant. The main survey collected data on socio-economic characteristics, and a wide range of economic and non-economic variables. After that, a separate survey on household income was administered, which was either administered immediately after the main survey, if the required respondents were available, or at a later time. In T2, the joint survey was conducted immediately after the loan applicant's interview[2]. In T3, the protocol recommended that the loan applicant should be surveyed first, and the spouse after[3][4].

In T2 and T3, respondents were informed beforehand that their partner would also be surveyed. Surveys were conducted at or close to the house of the respondent and in complete privacy. To ensure independent responses, the enumerator sought a quiet place where the respondent could be surveyed without interruption,

[1] In T2, the other household member consisted of the son (13 cases) or daughter (13 cases), mother (4 cases), sister (1 case) or grandfather (1 case). In T3, the other household member consisted of the son (9 cases), daughter (8 cases) or brother (1 case) of the respondent.

[2] Once the main survey was finished, the spouse was asked to join for the income section. This was feasible for the majority of households. The median time gap between the end of the loan applicant survey and the beginning of the spouse survey is 4 minutes. For 40 surveys (6% of all T2 surveys), the time gap was larger than 30 minutes. Only 3 joint surveys in T2 were administered on a different day than the loan applicant surveys.

[3] This was not adhered to in 11 cases (1.7% of all T3 surveys) where the spouse was only available before the loan applicant was surveyed.

[4] Spouses were mostly surveyed by the same enumerator. However, for logistical reasons, we allowed two different enumerators to survey spouses from the same household when tracking was very difficult. In these cases, another enumerator revisited the household at a scheduled time. This was the case in 14% of the T3 surveys.

and without other household members overhearing the interview. To ensure high compliance rates, the survey protocol allowed for flexibility in scheduling interviews, meaning that enumerators could return to the village to complete joint surveys or spouse surveys later.

III. Cost analysis

In joint surveys, survey duration increases slightly because the enumerator needs to ask consent to the spouse too and because the discussion between two spouses adds to the survey duration. In separate surveys, the same survey section is conducted twice. Cost calculations are shown in Table A1 below. We first calculate the average number of enumerator days per village. For enumerators who visited more than one village per day, we count each village proportionally to the number of villages visited that day, but we add 50% to account for additional transport costs. We calculate the average costs per village by adding the costs for enumerator days and supervisor days. We then divide this by the average number of respondents per village to obtain the average cost per respondent. Finally, we add tracking costs. In Paraguay, these consist of costs for rescheduling interviews for 34% of the sample in T2. In Uganda, tracking costs consist of community mobilizer days that are used for scheduling surveys with each respondent.

We used different tracking approaches in each country, tailored to the local context and available survey budget. In Uganda, we used community mobilizers to visit each village before the enumeration team and announce who would be surveyed. In Paraguay, we asked at the onset of the interview if the spouse would be available for a joint survey after. In case the spouse was absent, we made an appointment via phone and came back at a later time.

We calculate the average costs per survey by treatment arm by first calculating the average number of enumerator days and supervisor days per village. We then divide this by the average number of surveys per village and add tracking costs.

Note that these cost comparisons serve to show the cost differences relative to other treatment arms. We did not factor in costs that are constant across treatment arms, such as data quality monitoring (backchecks, auditing and high-frequency checks), costs for training or programming or overhead.

Table A1: Cost per survey by treatment arm

	Average survey duration (min)	Average number of enumerator days/village	Supervisor day per enumerator day	Average cost per village	Average number of main respondents per village	Cost per survey (enumerator + transport)	tracking costs per village	Costs for rescheduling interviews/survey	**Cost per survey (survey + tracking)**	**Cost differences (T1 = base)**
	survey and transportation costs						tracking costs			
Uganda										
T1	100	5.45	0.15	$ 280.46	12.8	$ 21.86	$ 43.00	$ 3.35	**$ 25.21**	100%
T2	107	5.55	0.15	$ 285.42	13.0	$ 21.99	$ 86.00	$ 6.63	**$ 28.61**	+13%
T3	121	6.59	0.15	$ 339.06	12.64	$ 26.82	$ 86.00	$ 6.80	**$ 33.63**	+33%
Paraguay										
T1	66	3.74	0.08	$ 207.93	9.8	$ 21.30			**$ 21.30**	100%
T2	76	3.90	0.08	$ 216.55	9.2	$ 23.55	34%	$ 0.69	**$ 24.23**	+14%

IV. Summary statistics

Column 1 and 5 of Table A2 show the most common sources of income in our sample in both countries. Multiple sources of income are possible within one household. In Uganda, where the sample consists of subsistence farmers, the survey was administered immediately after the harvest of the second agricultural season of 2019. 61% of households earned an income from crop farming in the month preceding the survey, while 53% of households earned an income from business activities in the last 4 weeks. The third most common occupation type in Uganda is occasional employment (34%). In Paraguay, the three most frequent sources of household income are occasional employment (57%), self-employment (47%) and salaried employment (44%).

Columns 2-4 and 6-8 show the total household income from each activity in the 4 weeks preceding the survey, across the different treatment arms. In Uganda, crop farming, business income and income from the infrequent sale of items contributed most to the total household income. In Paraguay, the highest amount of income comes from salaried employment, occasional employment, and self-employment.

Table A2: Summary statistics: household income (Uganda and Paraguay)

	Uganda				Paraguay			
	dummy	mean	sd	max	dummy	mean	sd	max
Crop farming	0.61	91.53	191.21	1291	0.22	61.01	194.84	1564
Livestock farming	0.12	9.73	50.54	540	0.32	80.94	205.27	1329
Salaried employment	0.07	12.58	63.05	470	0.44	354.90	558.22	2658
Occasional employment	0.34	16.73	43.21	327	0.57	264.00	402.36	2306
Self-employment	0.53	49.71	100.83	783	0.47	177.01	334.85	1955
Pension payments					0.03	14.36	99.86	860
Social assisistance	0.20	9.48	34.49	313	0.16	26.95	81.95	508
Government transfers	0.01	0.26	3.28	59	0.24	37.54	75.55	391
Non-government transfers	0.02	0.47	3.66	39				
Rents	0.11	6.20	23.99	176				
Infrequent sales	0.20	44.11	153.31	1252				
Selling wildlife	0.05	0.59	3.40	31				
Other					0.01	0.87	9.13	102
Aggregate hh income		262.74	360.77	2270		1027.61	855.96	5043
Observations	2564				1578			

Column "dummy" shows a dummy variable with value 1 if anyone in the household earned an income from this activity. The other columns show the mean, standard deviation and maximum earned by a household in a certain activity. The minimum is equal to 0 for each activity.

Table A3 below reports the average income per household member in each country, reported by the different groups of respondents. In Uganda, the loan applicant (male in 66% of the sample) contributes about 65% of the total household income. In Paraguay, the loan applicant's husband contributes a little over half of the household income on average, whereas the average loan applicant contributes 34%. In Uganda, 87% of loan applicants and 59% of spouses earned an income in the last 4 weeks across the three treatment arms. In Paraguay, this number is 80% for loan applicants and 83% for spouses. Other income earners in the household are the son of the loan applicant (8% in Uganda and 16% in Paraguay) and the daughter (4% in Uganda and 6% in Paraguay). In a small minority of households (1% in Uganda and 7% in Paraguay), other household members contributed to the household income.

Table A3: Income split up per household member, across the three treatment arms

	Uganda				Paraguay			
	dummy	mean	sd	max	dummy	mean	sd	max
Loan Applicant	0.87	170.23	289.30	1800	0.80	347.57	440.41	2150
Spouse	0.59	72.94	158.52	1017	0.83	539.54	562.24	3245
Son	0.08	5.73	29.06	273	0.16	79.25	245.95	1439
Daughter	0.04	1.00	6.22	63	0.06	18.84	86.57	704
Other male	0.01	0.13	1.42	16	0.04	12.68	74.28	715
Other female	0.00	0.00	0.05	.94	0.03	4.72	29.75	215
Aggregate hh income		262.74	360.77	2270		1027.61	855.96	5043
Observations	2564				1578			

Column "dummy" shows a dummy variable with value 1 if anyone in the household earned an income from this activity. The other columns show the mean, standard deviation and maximum earned by a household in a certain activity. The minimum is equal to 0 for each activity.

V. Balance tests

We test for balance of socio-demographic variables for Uganda and Paraguay. The results are reported in Tables A4 and A5. From Table A4, we see that respondents in Uganda were 41 years old on average, mostly male (69%), mostly married and self-identified as the household head. The average Ugandan household in our sample consists of 3 adults and 4 children. In Uganda, the majority of the sample relies on subsistence farming as their main occupation.

From Table A5, we see that respondents in Paraguay were 40 years old on average, all female (in line with the study's eligibility criteria), and most self-identified as the household head. About half of the sample is married while the other half lives with a partner. The average Paraguayan household in our sample consists of 5 household members. The sample consists of women who either have a business or intend to start one. 31% of the respondents report being self-employed as their main occupation.

The final column(s) in Table A4 and Table A5 report the p-value(s) for a test of balance between the different treatment arms. In all, we find most variables to be well balanced, as we would expect from randomization. The last two rows show the F-test for joint orthogonality. We conclude that the randomization resulted in comparable groups across the different treatment arms.

Table A4: Balance on socio-demographic variables (Uganda)

		Loan applicant only (T1)			Spouses joint (T2)			Spouses separately (T3)			difference	
	n	mean	sd	n	mean	sd	n	mean	sd	(1)-(2)	(1)-(3)	(2)-(3)
Age	1283	40.79	12.10	649.00	40	12.09	632.00	40.55	12	0.825*	0.082	-0.578
										(0.497)	(0.509)	(0.609)
Female	1283	0.35	0.48	649.00	.31	0.46	632.00	0.33	.47	0.032	0.015	-0.023
										(0.026)	(0.026)	(0.030)
Loan applicant is household head	1283	0.71	0.45	649.00	.74	0.44	632.00	0.72	.45	-0.029	-0.008	0.025
										(0.023)	(0.022)	(0.026)
Married	1283	0.86	0.35	649.00	.87	0.33	632.00	0.87	.34	-0.009	-0.001	0.008
										(0.015)	(0.015)	(0.015)
Post-primary education	1283	0.29	0.45	649.00	.3	0.46	632.00	0.30	.46	-0.013	-0.011	0.004
										(0.026)	(0.021)	(0.026)
Number of household members	1283	6.68	2.58	649.00	6.6	2.51	632.00	6.67	2.4	0.061	0.001	-0.052
										(0.117)	(0.112)	(0.107)
Number of adults (>18)	1283	2.78	1.24	649.00	2.7	1.18	632.00	2.68	1.1	0.117**	0.100*	-0.013
										(0.057)	(0.057)	(0.062)
Subsistence farming as main occupation	1283	0.85	0.36	649.00	.83	0.38	632.00	0.88	.33	0.021	-0.018	-0.050*
										(0.024)	(0.021)	(0.026)
Poverty probability	1283	45.78	24.44	649.00	43	23.34	632.00	45.45	24	1.794	0.518	-1.273
										(1.299)	(1.256)	(1.391)
Number of loans	1279	2.84	2.63	646.00	3	2.76	631.00	2.71	2.7	-0.242	0.132	0.312
										(0.165)	(0.148)	(0.196)
Total household savings, USD ppp	1128	2.70	4.36	564.00	3.1	4.18	547.00	2.91	5.7	-0.419*	-0.253	0.095
										(0.227)	(0.288)	(0.287)
Satisfaction with life (0-10)	1281	4.53	2.49	649.00	4.7	2.48	632.00	4.61	2.4	-0.104	-0.074	0.053
										(0.109)	(0.128)	(0.130)
F-Test										1.47	1.48	1.44
P-value F-test										0.15	0.14	0.17

The values displayed for t-tests is the difference in the means across the groups. Standard errors are clustered at the village level. Branch fixed effects are included in all estimation regressions. Missing values are imputed at the treatment group mean. ***, ** and * indicate significance at the 1, 5 and 10 percent critical level.

Table A5: Balance on socio-demographic variables (Paraguay)

	Loan applicant only (T1)			Spouses joint (T2)			
	n	mean	sd	n	mean	sd	difference
Age	1063	39.82	12.31	515	39.02	11.61	-0.907
							(0.673)
Female	1063	1.00	0.00	515	1.00	0.00	0.000
							(0.000)
Loan applicant is household head	1057	0.92	0.26	512	0.92	0.27	-0.005
							(0.017)
Married	1063	0.52	0.50	515	0.52	0.50	0.006
							(0.036)
Post-primary education	1063	0.19	0.39	515	0.24	0.43	0.049
							(0.031)
Number of household members	1063	4.65	1.85	515	4.66	1.69	0.030
							(0.101)
Number of adults (>18)	1062	2.67	1.15	515	2.68	1.16	0.015
							(0.067)
Self-employment as main occupation	1063	0.31	0.46	515	0.30	0.46	-0.017
							(0.034)
Poverty probability	1063	4.71	7.41	515	3.89	6.79	-0.835
							(0.535)
Number of loans	1061	0.79	1.01	514	0.81	0.96	0.030
							(0.069)
Total household savings, USD ppp	1025	0.98	7.26	492	0.96	6.05	-0.033
							(0.370)
Satisfaction with life (0-10)	1042	6.81	2.40	503	7.02	2.44	0.195
							(0.142)
F-test							1.47
P-value F-test							0.61

The values displayed for t-tests is the difference in the means across the groups. Standard errors are clustered at the village level. Branch fixed effects are included in all estimation regressions. Missing values are imputed at the treatment group mean. ***, ** and * indicate significance at the 1, 5 and 10 percent critical level.

VI. Outcome variables

Aggregate household income is constructed as the sum of income for each activity and each household member, as reported by the loan applicant (T1) or the couple together (T2). T3 income is constructed as the sum of income earned by female household members, reported by the female respondent and income earned by male household members, reported by the male respondent. Aggregate household income was set to zero if the household did not earn an income. Missing values are imputed at the treatment group mean.

Aggregate household consumption is constructed as the sum of the value of all food items consumed inside and outside the household by all household members in the past 7 days, as reported by the loan applicant (T1) or the couple together (T2), or the loan applicant, who is surveyed separately before their spouse is surveyed (T3). Missing values are imputed at the treatment group mean. Outcome variables are converted to USD PPP and winsorized at the 99th percentile by treatment arm.

Following our pre-analysis plan, we trim outcome variables at the 99th percentile by treatment group in order to prevent outliers from determining the results. We also impute missing values by treatment group mean. Our main results are robust to trimming and winsorizing (Table A6).

Table A6: Robustness to imputation and winsorization

	Winsorized and imputed	Not winsorized, not imputed	Only winsorized
	(1)	(2)	(3)
Panel A: Household income (Uganda)			
Loan applicant only (T1)	-11.029	-11.722	-10.851
	(18.87)	(22.11)	(19.21)
Spouses separately (T3)	24.579	16.473	24.466
	(26.14)	(28.20)	(26.33)
Region FE	Yes	Yes	Yes
Control (T2) Mean	275.338	283.678	
N	2564	2531	2531
R2	0.021	0.017	0.021
P-value T1 = T3	0.144	0.264	0.149
P-value T1 = T2	0.560	0.597	0.573
P-value T2 = T3	0.348	0.560	0.354
Panel B: Household income (Paraguay)			
Loan applicant only (T1)	-34.910	-17.107	-34.485
	(54.63)	(60.97)	(54.92)
Region FE	Yes	Yes	Yes
Control (T2) Mean	1056.919	1066.093	1056.919
N	1578	1578	1574
R2	0.012	0.011	0.012
P-value T1 = T3			
P-value T1 = T2	0.524	0.779	0.531
P-value T2 = T3			
Panel C: Household (Uganda)			
Loan applicant only (T1)	-1.157	-1.176	-1.157
	(2.53)	(2.62)	(2.53)
Spouses separately (T3)	-6.072**	-6.198**	-6.072**
	(2.76)	(2.84)	(2.76)
Region FE	Yes	Yes	Yes
Control (T2) Mean	90.009	90.531	
N	2564	2564	2564
R2	0.039	0.039	0.039
P-value T1 = T3	0.033	0.037	0.033
P-value T1 = T2	0.648	0.654	0.648
P-value T2 = T3	0.029	0.030	0.029

This table reports the OLS regression results for different treatments on household income. The dependent variable is aggregate household income, constructed with and without imputation and winsorization at the 99th percentile by treatment arm. Outcome variables are converted to USD PPP. SE in parenthesis and clustered at the village level. ***, ** and * indicate significance at the 1, 5 and 10 percent critical level.

Table A7 exploits the different data points for T3 in Uganda. Column 1 shows our preferred (benchmark) construct of income, where we use data from the interviews with both spouses to construct household income. Column 2 uses the data from the loan applicant, who reports on the income earned by all household members. Column 3 uses the data from the spouse survey, reporting on the income earned by all household members.

Table A7: Main treatment effect for aggregate household income for with T3 constructed from different data sources (Uganda)

	Aggregate income		
	(1)	(2)	(3)
	T3 = Benchmark	T3= Loan applicant	T3 = Spouse
Loan applicant only (T1)	-11.029	-10.440	-10.678
	(18.87)	(18.58)	(19.00)
Spouses separately (T3)	24.579	-17.953	-3.807
	(26.14)	(22.99)	(26.58)
Region FE	Yes	Yes	Yes
Control (T2) Mean	275.338	275.338	275.338
N	2564	2564	2497
R2	0.021	0.020	0.023
P-value T1 = T3	0.144	0.720	0.780
P-value T1 = T2	0.560	0.575	0.575
P-value T2 = T3	0.348	0.436	0.886

This table reports the OLS regression results for different treatments on household income. The dependent variable is aggregate household income, either constructed as the gold standard measure of income (see Table 1), data from the interview with the loan applicant and data from the interview with the spouse. Outcome variables are converted to USD PPP and winsorized at the 99th percentile by treatment arm. SE in parenthesis and clustered at the village level. ***, ** and * indicate significance at the 1, 5 and 10 percent critical level.

VII. Distribution analysis

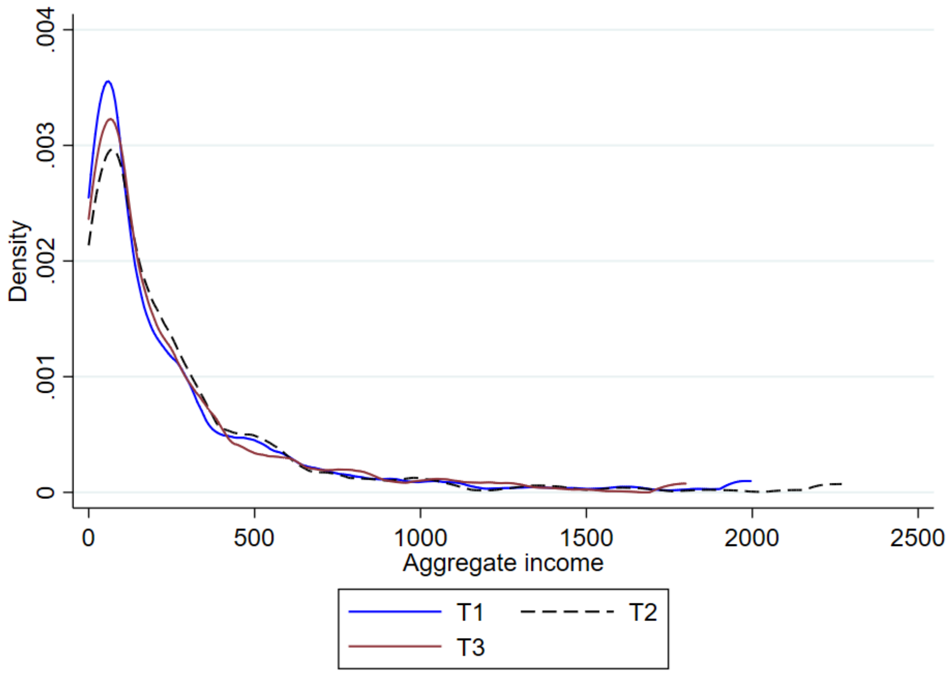

Figure A1: Kernel density of aggregate income by treatment (Uganda)

Two-sample Kolmogorov-Smirnov test for equality of distribution functions for aggregate income in Uganda:

- T1 vs T2: D = 0.0549, p-value = 0.148
- T3 vs T1: D = 0.0226, p-value = 0.982

- T3 vs T2: D = 0.0511, p-value = 0.374

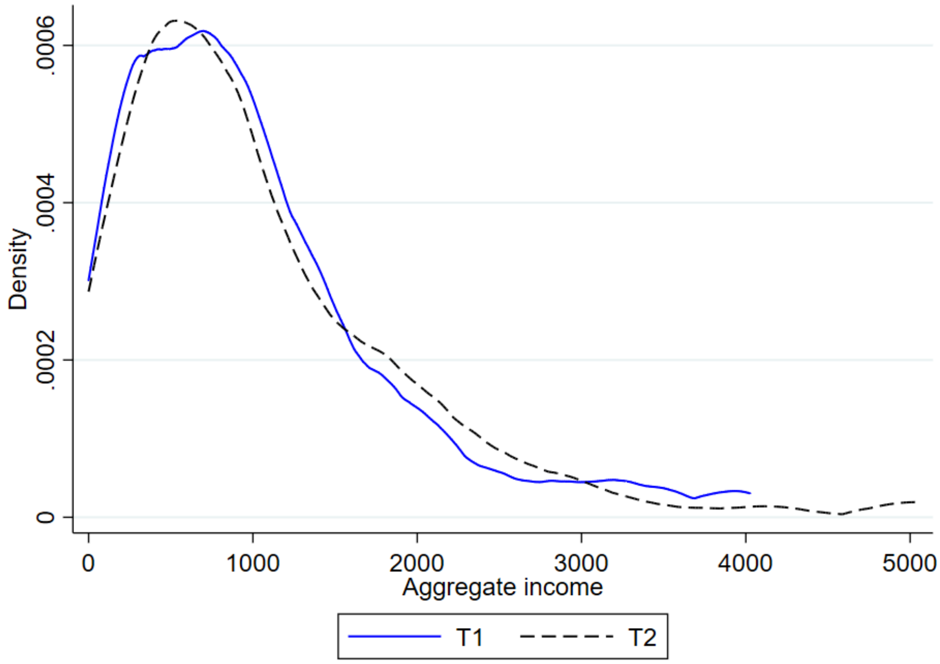

Figure A2: Kernel density of aggregate income by treatment (Paraguay)

Two-sample Kolmogorov-Smirnov test for equality of distribution functions for aggregate income in Paraguay:

- T1 vs T2: D = 0.0441, p-value = 0.511

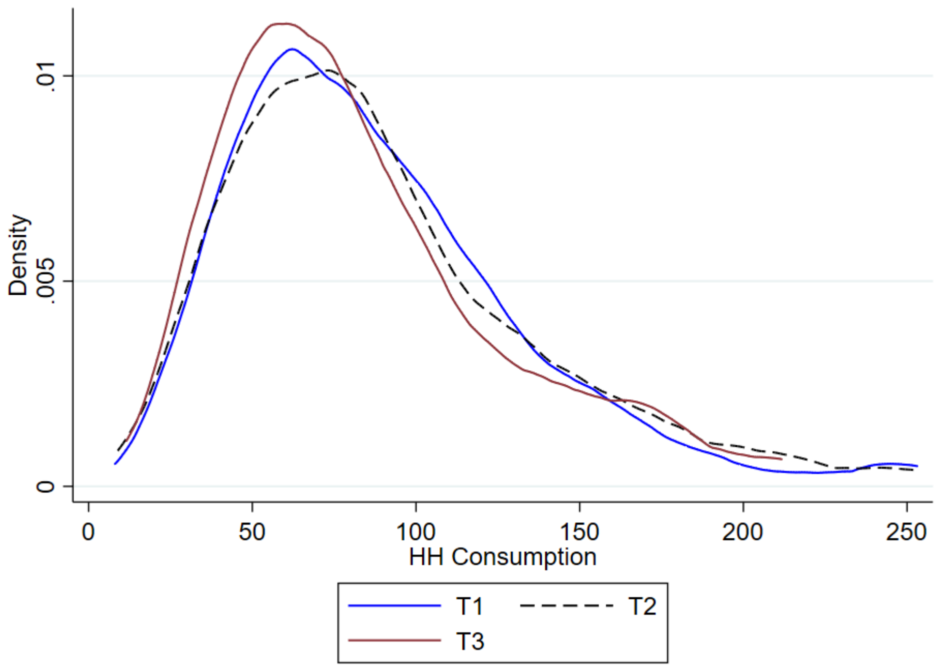

Figure A3: Kernel density of household consumption by treatment (Uganda)

Two-sample Kolmogorov-Smirnov test for equality of distribution functions for consumption in Uganda:

- T1 vs T2: D = 0.0324, p-value = 0.757

- T3 vs T1: D = 0.0745, p-value = 0.018
- T3 vs T2: D = 0.0893, p-value = 0.012

VIII. Enumerator effects

Table A8: Enumerator effects for main outcome variables.

	Adj. R^2	
	Uganda	Paraguay
Aggregate income	0.024	0.004
Household consumption	0.071	

This table reports the adjusted R^2 for each of the main outcome variables. The R^2 is obtained from regressing the outcome of interest on enumerator dummies and a constant.

Table A9: Treatment effects for aggregate income, with enumerator fixed effects

	Household income		Household consumption
	(1)	(2)	(3)
	Uganda	Paraguay	Uganda
Loan applicant only (T1)	-12.515	-47.278	0.532
	(19.25)	(53.17)	(2.29)
Spouses separately (T3)	20.414		-5.327**
	(26.65)		(2.50)
Region FE	Yes	Yes	Yes
Enum FE	Yes	Yes	Yes
Control (T2) Mean	275.338	1056.919	90.009
N	2564	1578	2564
R2	0.052	0.023	0.119
P-value T1 = T3	0.171		0.007
P-value T1 = T2	0.516	0.375	0.817
P-value T2 = T3	0.445		0.034

This table reports the OLS regression results for different treatments on household income and consumption. Outcome variables are converted to USD PPP and winsorized at the 99th percentile by treatment arm. All regressions include region and enumerator fixed effects. SE in parenthesis and clustered at the village level. ***, ** and * indicate significance at the 1, 5 and 10 percent critical level.

IX. Gender

In Paraguay, the respondents are all female, so we can interpret Table 2 directly through a gender lens. For consumption, a breakdown by gender is not possible since we only collected data on household income. Table A10 shows a breakdown by the loan applicants' gender for income effects in Uganda. Columns 1 and 2 show the loan applicant's income and spouse's income, reported by the female loan applicant. Columns 3 and 4 show the own income and spouse's income if the loan applicant is a man.

Table A10: Treatment effects by gender (Uganda)

	Female loan applicant		Male loan applicant	
	(1)	(2)	(3)	(4)
	Own income	Husband income	Own income	Wife income
Loan applicant only (T1)	-18.617	-25.656	25.330	-15.099**
	(19.78)	(17.52)	(18.16)	(7.60)
Spouses separately (T3)	-35.614*	-3.927	21.680	-16.598**
	(19.01)	(22.18)	(21.87)	(7.78)
Region FE	Yes	Yes	Yes	Yes
Control (T2) Mean	126.326	138.307	184.793	58.050
N	859	859	1705	1705
R2	0.031	0.031	0.025	0.019
P-value T1 = T3	0.214	0.253	0.859	0.804
P-value T1 = T2	0.348	0.145	0.165	0.048
P-value T2 = T3	0.063	0.860	0.323	0.034

This table reports the OLS regression results for different treatments on income for the loan applicant and spouse in Uganda, separated by gender of the loan applicant. Outcome variables are converted to USD PPP and winsorized at the 99th percentile by treatment arm. SE in parenthesis and clustered at the village level. ***, ** and * indicate significance at the 1, 5 and 10 percent critical level.

X. Heterogeneous treatment effects

Table A11: Heterogeneous treatment effects - Uganda

	Aggregate	Respondent	Spouse	Other members
Loan applicant only (T1 and T3)	-12.842	3.277	-10.241	-0.714
	(18.63)	(15.05)	(7.42)	(1.79)
Income hiding (applicant)	19.834	9.571	-2.180	-1.957
	(64.82)	(28.80)	(31.21)	(4.79)
Loan applicant only (T1 and T3)×Income hiding (applicant)	3.968	34.327	-17.917	1.067
	(69.58)	(35.57)	(32.32)	(5.31)
Region FE	Yes	Yes	Yes	Yes
T1/3 + Income hiding	-8.87	37.60	-28.16	0.35
p-value	0.89	0.27	0.37	0.94
N	2560	2560	2560	2560
R2	0.020	0.021	0.015	0.006
Loan applicant only (T1 and T3)	-8.233	1.603	-6.847	-0.049
	(19.09)	(15.46)	(7.84)	(1.82)
Income hiding (spouse)	-5.054	-52.510***	20.161	4.107
	(38.73)	(17.05)	(20.23)	(5.32)
Loan applicant only (T1 and T3)×Income hiding (spouse)	-9.776	57.274**	-39.977*	-4.478
	(43.18)	(25.75)	(21.55)	(5.69)
Region FE	Yes	Yes	Yes	Yes
T1/3 + Income hiding suspect	-18.01	58.88	-46.82	-4.53
p-value	0.65	0.02	0.02	0.38
N	2541	2541	2541	2541
R2	0.020	0.021	0.016	0.007
Loan applicant only (T1 and T3)	-17.484	-6.251	-8.165	0.655
	(19.96)	(16.03)	(8.31)	(1.87)
Any income hiding in HH	-11.161	-39.830**	10.309	4.191
	(38.49)	(18.75)	(20.55)	(4.81)
Loan applicant only (T1 and T3)×Any income hiding in HH	21.658	66.597***	-25.756	-5.925
	(43.57)	(25.53)	(21.67)	(5.08)
Region FE	Yes	Yes	Yes	Yes
T1/3 + Income hiding in HH	4.17	60.35	-33.92	-5.27
p-value	0.91	0.01	0.08	0.24
N	2564	2564	2564	2564
R2	0.020	0.021	0.015	0.007

This table reports the OLS regression results for different treatments on household income. The loan applicant interviews from T1 and T3 are pooled. T2 serves as the control group. The treatment dummy is interacted with a categorical variable indicating whether or not there is income hiding in the household.

Table A12: Heterogeneous treatment effects - Paraguay

	Aggregate	Respondent	Spouse	Other members
Loan applicant only (T1)	-28.477	65.597**	-76.983**	-6.464
	(56.38)	(26.47)	(35.55)	(22.51)
Income hiding (applicant)	247.160	237.776***	-30.260	66.423
	(169.65)	(89.52)	(96.02)	(73.41)
Loan applicant only (T1)×Income hiding (applicant)	-105.491	-166.579	92.867	-60.056
	(201.81)	(106.37)	(123.15)	(82.71)
Region FE	Yes	Yes	Yes	Yes
T1/3 + Income hiding	-133.97	-100.98	15.88	-66.52
p-value	0.49	0.33	0.90	0.41
N	1572	1572	1572	1572
R2	0.015	0.018	0.018	0.009
Loan applicant only (T1)	-37.652	63.107**	-85.331**	-9.541
	(57.84)	(27.21)	(35.94)	(23.00)
Income hiding (spouse)	-1.486	117.065	-134.007	15.232
	(171.00)	(94.08)	(94.49)	(61.95)
Loan applicant only (T1)×Income hiding (spouse)	-36.300	-121.425	125.742	-31.504
	(189.00)	(105.08)	(104.32)	(68.68)
Region FE	Yes	Yes	Yes	Yes
T1/3 + Income hiding suspect	-73.95	-58.32	40.41	-41.05
p-value	0.68	0.56	0.70	0.53
N	1550	1550	1550	1550
R2	0.013	0.012	0.020	0.008
Loan applicant only (T1)	-37.376	65.557**	-84.872**	-9.330
	(58.50)	(27.67)	(35.43)	(23.73)
Any income hiding in HH	18.854	118.444	-93.436	0.066
	(136.65)	(71.64)	(81.06)	(51.59)
Loan applicant only (T1)×Any income hiding in HH	20.294	-100.819	129.352	-10.028
	(156.97)	(82.25)	(94.49)	(59.84)
Region FE	Yes	Yes	Yes	Yes
T1/3 + Income hiding in HH	-17.08	-35.26	44.48	-19.36
p-value	0.91	0.64	0.65	0.72
N	1578	1578	1578	1578
R2	0.012	0.013	0.019	0.008

This table reports the OLS regression results for different treatments on household income. The treatment dummy is interacted with a categorical variable indicating whether or not there is income hiding in the household.

XI. Decomposing aggregates

Figure A4 and A5 show a breakdown of loan applicant's income per source of income in Uganda and Paraguay, respectively. Figure A6 and A7 show spouse's income.

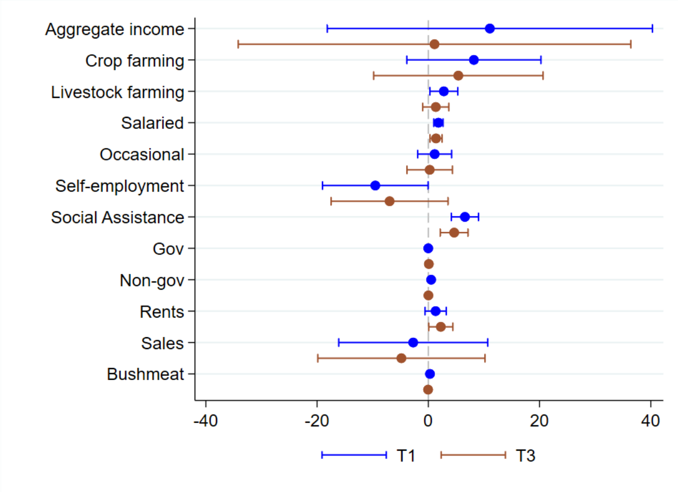

Figure A4: Loan applicant income per source of income (Uganda)

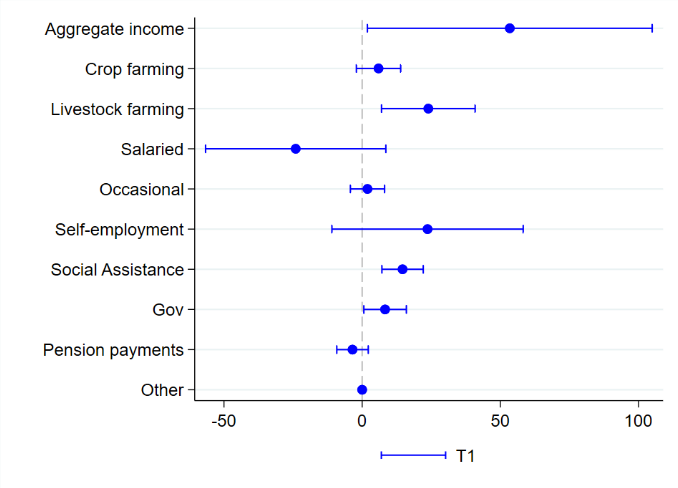

Figure A5: Loan applicant income per source of income (Paraguay)

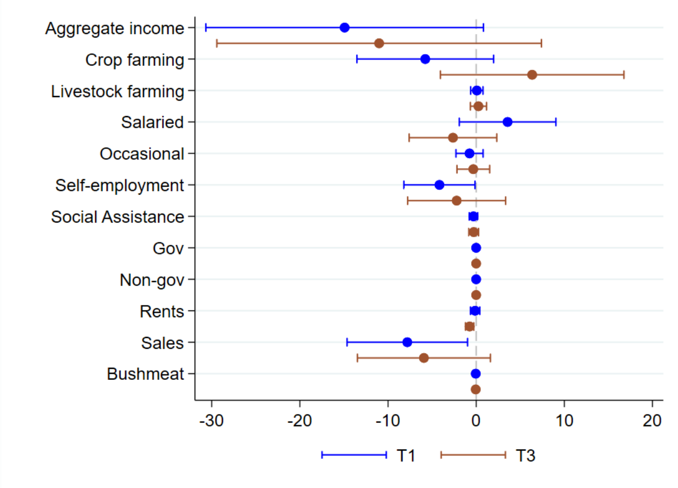

Figure A6: Spouse income per source of income (Uganda)

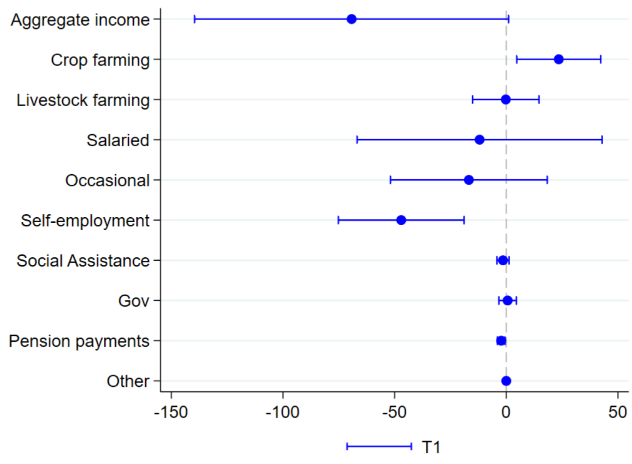

Figure A7: Spouse income per source of income (Paraguay)