Journal of African Economies, 2022, Vol. 00, number 00, 1–31
doi: 10.1093/jae/ejac025
Advance Access Publication Date:
Article

Are You Being Asked? Impacts of Respondent Selection on Measuring Employment in Malawi

Talip Kilic[a], Goedele Van den Broeck[b], Gayatri Koolwal[a,*] and Heather Moylan[a]

[a]Development Data Group, World Bank, 1818 H Street, NW Washington, DC 20433, USA, [b]Earth and Life Institute, University of Louvain, Croix du Sud 2, Louvain-la-Neuve, Belgium

*Corresponding author: Gayatri Koolwal. E-mail: koolwalg@devscience.org

Abstract

Accurate estimates of men's and women's employment are necessary for understanding sources of productivity and growth and designing well-targeted, gender-sensitive labour policies. This paper aims to address a key question—how respondent selection in household and labour force surveys affects these estimates—by leveraging two concurrent national surveys in Malawi that relied on the same questionnaire and field teams but differed in their approach to respondent selection. As compared with direct and private interviews with respondents, the 'business-as-usual' approach that allows for proxy reporting when targeted respondents are not available, as well as a mix of other standard survey approaches often used under time and resource constraints, is associated with significantly lower reporting of employment across a range of wage and self-employment activities. Although the effects are seemingly limited in absolute terms, they are quite large in relative terms, vis-à-vis the average participation rates and they tend to be more pronounced for women respondents and concerning questions with longer/12-month recall periods. The analysis also examines how household wealth, proxy reporting and difficulties associated with interpreting questions may be linked to lower reporting in the business-as-usual approach, and which can be examined in future methodological experimentation.

Keywords: Sub-Saharan Africa, Malawi, household surveys, respondent selection, labour, gender

JEL classification: C83, J21, J16

1. Introduction

Accurate estimates of men's and women's employment are at the heart of understanding sources of productivity and economic growth and designing well-targeted, gender-sensitive labour policies. National surveys that provide the required data, including labour force surveys and multi-topic household surveys with labour modules, are usually structured to

ask working-age individuals about their participation and time allocation to a range of employment-related activities over the last week or the last 12 months, depending on the question. However, while aimed towards working-age individuals themselves, the common approach in these surveys can often involve a combination of proxy respondents reporting on behalf of individuals who are not available for interviews, respondents potentially having other family members or individuals present during the interview and other issues tied to survey time constraints, including lack of follow-up questions linked to difficult-to-measure or seasonal activities and on intra-household roles in enterprises. Much of this is due to resource and time constraints faced in survey implementation. All of these variations, however, can potentially lead to errors in reporting employment—particularly when questions involve longer recall periods, or for types of employment that are more difficult to measure, including seasonal/informal work.

A key question, as a result, is whether survey approaches that place a greater emphasis on eliciting self-reported survey data in private interviews with respondents, as well as other efforts to ensure that respondents have the time and ability to report their outcomes carefully, have a significant effect on reporting of employment over the 'business-as-usual' practice that follows standard survey approaches discussed above. However, across low-income contexts, methodological research on this topic using nationally representative data has been more limited.

This study contributes to this discussion by leveraging a unique opportunity to compare how survey design affects labour reporting in Malawi. We examine two national surveys that were implemented concurrently by the Malawi National Statistical Office in 2016–17, which had the same questionnaire design and field teams but used different approaches to survey design and respondent selection in collecting individual-disaggregated survey data on economic outcomes. Specifically, the Integrated Household Panel Survey (IHPS) 2016, sought direct, intra-household, private interviews with adult household members on individual asset ownership and rights (Kilic *et al.*, 2021), as well as on employment, including gender-matching of enumerators and respondents. Separately, the Fourth Integrated Household Survey (IHS4) 2016–17, covered the same topics (with the same field teams and questionnaires) but employed a set of business-as-usual approaches including allowing for proxy reporting. Our study examines how reporting of work and employment varies across surveys and provides a descriptive analysis that presents hypotheses regarding drivers behind these differences. Our prior is that the IHPS approach to selecting and interviewing respondents provides more accurate responses on labour, underscored by recent international guidance around collecting self-reported data on individual economic outcomes. However, in the absence of a true survey experiment, our conclusions are focused on comparing relative differences across the surveys (as a package of design differences, but within which different approaches to respondent selection is a key element) as opposed to making conclusions about true estimates.

As compared with individual interviews conducted in the IHPS, our findings show that the business-as-usual approach, which comprises a set of common survey approaches to survey design and respondent selection, is associated with significantly lower reporting of employment in livestock-related activities and household non-farm enterprises (NFEs; in management and worker capacity), with stronger effects for women's participation and under a recall period of 12 months, as opposed to 7 days. Although the effects are seemingly limited in absolute terms, they are quite large in relative terms, vis-à-vis the average participation

rates. Among women, the business-as-usual approach in the IHS4 leads to a 2.6 percentage point decrease in livestock participation (which, given mean participation for women in the IHS4, is a relative decrease of 15%) and 3.6 percentage points in running/managing NFEs (a relative decrease of about 45%). For men, relative decreases in reporting over the last week are about 20% for NFE management. We also find that weekly hours spent across different sectors fall in the IHS4. The analysis also controls for individual interview date and enumerator fixed effects to help control for the dimensions of IHPS design around gender-matching of enumerators and respondents, as well as other variations related to interview duration and timing. The analysis also reveals that the business-as-usual approach is associated with lower reporting of wage employment among men. Looking at how the choice of survey approach affects time allocation to employment, we find that most differences arise within agriculture, with the business-as-usual approach associated with lower reported weekly hours by men and women in livestock activities as well as weekly hours devoted by men to crop agriculture.

Although the comparison of the two survey approaches is not based on an experimental design to be able to fully isolate the effects of specific design differences, we are able to descriptively examine potential channels that may be linked with our findings. When compared against the IHPS, for example, the IHS4 proxy sample contributes more to lower reporting than the own-reporting/non-proxy sample, particularly in agricultural and NFE activities. There is also some evidence that respondents in the IHS4 may have faced greater difficulty in interpreting and reporting on concepts related to businesses or enterprises (as evidenced by greater discrepancies in reporting in the IHS4 between the survey modules on labour and NFEs). We also discuss how household wealth is significantly associated with the extent of reporting differences across the individual-interview and business-as-usual approaches. Further exploration of these potential channels in experimental studies, as well as additional country contexts, can shed greater light on future guidance around respondent selection and survey design.

In general, a better understanding of issues around respondent selection and survey measurement, particularly using nationally representative studies on labour survey methods, is essential for developing more accurate statistics on men's and women's employment and improving the design of policies to bridge gender gaps in economic opportunities. This paper aims to contribute to the growing literature in this area across low- and middle-income contexts, and also support momentum that has increased on this front within the international statistical community as well. The UN Statistical Division (UNSD, 2019), for example, as well as the World Bank (Durazo *et al*., 2021), have published guidance for labour and socioeconomic surveys on the importance of self-reporting for individual-level data collection on outcomes across wealth and labour, at the same time recognising practical constraints around resources and implementation that surveys often have to manage.

Recommendations by the International Conference of Labour Statisticians (ICLS), which is hosted by the International Labour Organization (ILO) every 5 years, also provides guidance on the design and implementation of country labour force surveys, including respondent selection. As part of an effort to better characterise and delineate different areas of paid and unpaid work, for example, the 19th ICLS (ILO, 2013) not only narrowed the definition of employment strictly to work for pay or profit, but also recommended the collection of additional data on individuals' unpaid activities, including subsistence agricultural production (which would have been included under employment in the pre-

19th ICLS framework) and other unpaid work within and outside the household. Without collecting additional data on these unpaid activities, including the careful collection of data on work that is not paid regularly, important dimensions of employment could be missed through standard survey approaches.[1] Improving the accuracy of employment statistics can also aid in monitoring of Sustainable Development Goals targets related to small-scale agricultural and informal employment.[2] Overall, nationally representative studies on survey respondent selection have important implications for gender-sensitive policy design and an accurate understanding of how men and women in the economy respond to macroeconomic fluctuations (as captured through development research outputs and official statistics).

2. Literature review

There is a longstanding literature on respondent selection and labour reporting in nationally representative surveys, but mainly for high-income countries. Key questions centre around the role of proxy response, which can reduce survey costs and potentially provide less biassed information on sensitive topics,[3] but also carries the risk of inaccurate reporting due to informational gaps or differences in preferences and motivation between the proxy and intended respondent (see Bound *et al.*, 2001 and Cobb, 2018). Cognitive issues related to respondents' interpretation of questions, and ability to recall past outcomes accurately, additionally need to be considered (see Bickart *et al.*, 1990 and Moore *et al.*, 2000).

Several survey experiments in high-income settings, for example, have aimed to better understand sources of survey measurement error in earnings by matching individual survey data with a concurrent, 'validation' data set. In a study from Sweden, for example, Koijen *et al.* (2015) find that household budget surveys yield lower consumption estimates compared with tax records for higher-income and higher-wealth households. In comparing the U.S. Census Bureau's Survey of Income and Program Participation (SIPP) with administrative tax records from the U.S. Internal Revenue Service (IRS), Tamborini and Kim (2013) find that proxy reporting in the SIPP leads to significant underreporting of earnings for single female workers (also see Hill, 1987). Cristia and Schwabish (2009) demonstrate that other covariates (in particular, earning higher incomes) lead to underreporting in the SIPP and that

1 The Women's Work and Employment Partnership (WWEP), a collaborative initiative of the World Bank, the International Labour Organization (ILO), the Food and Agriculture Organization of the United Nations, was also launched in 2014 to promote interagency collaboration to operationalise the 19th ICLS definitions of work and employment.

2 Relevant SDG targets include, for example, the following: Target 2.3 to double the agricultural productivity and incomes of small-scale food producers by 2030, through secure and equal access to land, other productive resources and inputs, knowledge, financial services, markets and opportunities for value addition and non-farm employment; as well as Target 8.3 to promote development-oriented policies that support productive activities, decent job creation, entrepreneurship, creativity and innovation, and to encourage formalisation and growth of micro-, small- and medium-sized enterprises including through access to financial services.

3 Some studies on health outcomes in high-income countries, for example, have shown that proxy reporting can provide more accurate and detailed information on other household members' health, including fewer issues with social desirability bias (see Mathiowetz and Groves, 1985, for a discussion using data from the U.S. National Health Interview Survey).

individual characteristics that are associated with higher earnings (including age and marital status) are systematically associated with underreporting, leading to bias in estimates of these commonly used predictors of earnings. Quality issues have also been raised with earnings data across survey as well as administrative sources, however (see Abowd and Stinson, 2013, in a discussion of the SIPP and IRS records; and Britton *et al.*, 2019, in comparing administrative tax records with student loan data in the U.K.).

Outside of high-income settings, methodological survey research on measuring labour outcomes has been limited. The quality and regularity of administrative data is one reason, also due in large part to the prevalence of informal and irregularly paid work in these countries. Recent validation studies have involved smaller, more focused experiments where concurrent household surveys contrast alternative methodological approaches, or modules testing for different approaches are randomised within the same survey. Bardasi *et al.* (2011), for example, randomised the implementation of a 'short' and 'detailed' module in the Survey of Household Welfare and Labour in Tanzania, a survey of 1,344 households across seven districts in the country, to examine how employment statistics are affected by (a) the level of detail of questions and (b) self-reporting as opposed to using proxies. Their findings have strong gender implications; proxy response significantly lowers male employment in agriculture but has no effect for women (the proxy effect is mitigated when the spouse is selected as the proxy and when the proxy has some education). A few additional screener questions to identify work mitigates over-reporting of employment among women engaged completely in domestic work, as well as higher average weekly hours conditional on working, especially for men. Dillon *et al.* (2012) use the same data and experiment to study the effects of proxy reporting and question detail on child labour, and in Serneels *et al.* (2017) to examine the returns to education. In both of these studies, proxy reporting did not have an effect, but screening questions (i.e., a detailed versus short module) did matter.

The findings around respondent selection can vary by context and experimental design. In a separate experiment from Ethiopia of 1,200 Fairtrade coffee households (Galdo *et al.*, 2019), survey respondents were randomly selected to test, across three different seasons, the effects of proxy reporting on reporting of child labour (children themselves versus the household head or spouse). Proxy reporting was found to significantly underreport work of girls in agricultural settings, relative to children's own reports. The importance of self-reporting may also be more pronounced for labour outcomes such as hours or earnings, the specifics of which may be more challenging for respondents to recall for others (Posel and Casale, 2005). Measurement error stemming from lack of information about others' activities might be mitigated among more educated respondents, as well as those that participate jointly in economic activities and decision-making (Cobb, 2018). Recent research has also discussed, where resources are available, how self-reporting can also potentially assist in reducing respondent fatigue (in situations where one respondent is asked to report for others' outcomes) and related measurement concerns (Ambler *et al.*, 2021). These channels need to be tested further, however.

Recent international guidance on survey design and implementation has also emphasised the importance of self-reporting across individual economic outcomes, when possible. This includes guidelines by the UNSD (UNSD, 2019), based on several country pilots supported by the Evidence and Data for Gender Equality (EDGE) initiative, which recommends self-reported data on individuals' asset ownership and rights. Country pilots conducted by the ILO between 2015 and 2016 also aimed to inform guidance on implementing the 19th ICLS

recommendations on work and employment; as compared with direct respondents, the pilots found that proxy reporting led to a lower number of activities reported across countries and lower reporting of desire and need to work (see Benes and Walsh, 2018a, 2018b). Barring a sizable sample and true experimental design across different country contexts, however, it is difficult to make definitive conclusions about the true effects of proxy reporting. At a minimum, there is recognition that surveys collect information on whether the respondent self-reported, or whom reported for them, given the potential of measurable differences in how direct versus proxy respondents might report specific activities. Recent World Bank guidance on the measurement of employment and own-use production in household surveys (Durazo *et al.*, 2021) has also underscored the importance of self-reporting on labour, as well as collecting metadata on who reported for the respondent, if direct interviews were not possible.[4]

Our paper builds on the emerging literature focused on low-income settings, with a particular emphasis on evidence using nationally representative surveys, and examines whether, and how, implementing individual interviews makes a significant difference in labour reporting. We also investigate the heterogeneity of the impact of individual interviews in accordance with different individual and household characteristics.

3. Survey design and sampling across the Malawi IHS4 and IHPS

Malawi is a predominantly rural country situated in southeast Africa, with a little over half of its households below the national poverty line and 20% living in extreme poverty (World Bank, 2019). Agriculture makes up 26% of the GDP and is a source of employment for 83% of Malawian households (Davis *et al.*, 2017). A high share of households, particularly in rural areas, are engaged in off-farm[5] activities as well, such as wage employment (which can be seasonal or full-year round), managing or working in an NFE and *ganyu*—the local term for casual, part-time labour (Beegle *et al.*, 2017). Based on our fieldwork experience, *ganyu* labour takes place mostly in agriculture for both men and women in rural areas, in industry for urban men and in services for urban women.[6]

3.1 Survey design and questionnaires

Our analysis is informed by two surveys that were implemented concurrently by the Malawi National Statistical Office: (1) the Fourth Integrated Household Survey (IHS4), a national cross-sectional survey conducted between April 2016–April 2017, and (2) the Integrated

4 See Durazo *et al.*, 2021, p. 32: 'Information in the labour module should always be collected at the individual level, with questions administered directly to each household member. For persons who cannot respond on their own behalf, another household member may respond for them (a proxy respondent). However, proxy-reported data may not be as reliable as self-reported data. Every effort should therefore be made to ensure each person responds for himself or herself. Nonetheless, information reported by a proxy respondent is greatly preferred to missing data. In such cases, metadata should be collected indicating if a proxy respondent was used and which household member responded on behalf of the individual whose data were being collected.'

5 Off-farm refers to the location of the work taking place, i.e., outside the own farm household.

6 Van den Broeck and Kilic (2019) use the same insight regarding *ganyu* while determining its farm versus off-farm nature in rural versus urban areas.

Household Panel Survey (IHPS), the third wave of a national longitudinal household survey that was implemented from April 2016 to January 2017.

Like other national multi-topic surveys under the Living Standards Measurement Study (LSMS) program, the IHS4 and IHPS collected detailed data on a range of individual and household outcomes critical for understanding poverty and welfare. In particular, both surveys had the same questionnaire structure, flow of modules and question wording,[7] and each covered the same range of topics across household members' age, relationship to the household head, marital status, education, labour and health; members' ownership of assets; household access to infrastructure and services; household consumption expenditure; NFE activity; credit; subjective well-being; and shocks faced by the household. Both surveys also had the same enumerator teams, which visited both the IHPS and IHS4 enumeration areas (EAs), and followed the same protocol for administering interview questions across the two surveys. Appendix Figure A1 shows that the two national survey samples were similarly distributed across interview months and survey strata.[8]

By design, the two surveys differed in their approach to selecting respondents. The IHS4 allowed for greater reporting by proxy when respondents were not available (40% of respondents in the IHS4 labour module did not report for themselves). In comparison, the IHPS was supported by the World Bank LSMS Plus (LSMS+) initiative that followed the 2019 UNSD-EDGE guidance on self-reported, individual-level data collection on asset ownership and rights, as well as guidance supported by the WWEP. The IHPS attempted to interview up to four adult members per household directly,[9] and in private for specific modules on labour, health, education and asset ownership.[10] Within the LSMS+ supported modules, the IHPS enumerators also aimed to conduct within-household interviews simultaneously so that respondents could be interviewed privately, along with emphasising a gender match-up

7 The data, questionnaires and basic information document for the IHS4 2016–17 can be accessed here: https://microdata.worldbank.org/index.php/catalog/2936. The data, questionnaires and basic information document for the IHPS 2016 can be accessed here: https://microdata.worldbank.org/index.php/catalog/2939. Both the IHS4 2016–17 and the IHPS 2016 were implemented with technical and financial assistance from the World Bank LSMS-ISA, using the Surveys Solutions Computer-Assisted Personal Interviewing (CAPI) platform. The implementation of the individual interviews as part of the IHPS 2016 was made possible by technical and financial assistance from the World Bank LSMS+ Program.

8 There were six survey strata, by region (North/Central/South) and within each region, by urban/rural.

9 This was an upper limit that only applied to 1% of the sampled household population that had more than four adults. If a sampled household had more than four adult household members, following the preference given to the head of the household, and his/her spouse if applicable, the remaining interview targets (two or three depending on the presence of a spouse) were selected at random from the remaining pool of adult household members.

10 The survey data production and methodological research activities of the World Bank LSMS+ program aims to enhance the availability and quality of individual-disaggregated survey data on key dimensions of economic opportunity. Over the period of 2016–21, the thematic focus of the program is on asset ownership, work and employment and entrepreneurship. The national household surveys supported by the World Bank LSMS+ initiative include the Cambodia LSMS+ Survey 2019–20, Ethiopia Socioeconomic Survey 2018–19, Malawi IHPS 2016, Nepal LSMS+ Phone Survey 2022, Sudan Labor Market Panel Survey 2022 and Tanzania National Panel Survey 2019–20. For more information, please visit: www.worldbank.org/lsmsplus.

between enumerators and respondents to be able to collect sensitive information about assets and economic opportunities without cultural factors potentially biasing responses. Appendix I includes the protocol for administering the IHPS individual questionnaire.[11,12]

3.2 Comparing the relative merits of each survey approach

A comparison of the IHPS with the IHS4 sheds light on how responses vary by survey approach across similar subgroups, although arriving at the 'true' estimates of men's and women's labour participation can be elusive in either survey approach. In keeping with the literature on measurement issues discussed earlier (Bickart *et al.*, 1990; Moore *et al.*, 2000; Bound *et al.*, 2001; Cobb, 2018) and recent survey guidance discussed in Section 2, our prior is that the approach used by the IHPS in interviewing respondents directly and individually would help reduce measurement error in labour reporting. Experimental studies have shown this is more likely, for example, in lower-income contexts where seasonality, prevalence of secondary occupations and informality of work (such as in agriculture) can make recall more challenging for proxy respondents (Bardasi *et al.*, 2011).

One complexity that arises in the context of operationalising private interviews with adult household members is non-response, which is 16.8% in the case of IHPS 2016. Our analysis employs a regression-based correction of non-response, which is discussed in greater detail in the next section but is common in economics/statistics literature, including for correction of attrition in the use of longitudinal household survey data.

Apart from measurement issues, an added concern associated with the IHPS approach is the higher cost vis-à-vis the business as usual. The survey metadata on interview dates allow us to get a better understanding of the added burden of the IHPS approach. On average, the field teams spent approximately 3.4 days in the IHS4 EAs for the business-as-usual interviews, while the same field teams spent an average of 4.5 days in the IHPS EAs. The additional

11 For more information on the organisation and implementation of the individual-disaggregated data collection as part of the IHPS, please consult the survey's basic information document, which can be accessed here: https://microdata.worldbank.org/index.php/catalog/2939/download/47216.

12 An additional difference between the IHS4 2016–17 and the IHPS 2016 is that there are individuals that were interviewed by the IHPS 2016 and that had previously been interviewed in 2013 and/or 2010. Recall that the IHPS 2016 is a longitudinal survey, primarily of individuals. Starting with a baseline household survey sample in 2010, the IHPS had been fielded again in 2013 to track all individuals that were at least 12 years of age and that had been interviewed in 2010, provided that they were known to be living in Malawi. In 2016, the same tracking rules were applied, including attempts to interview individuals that had been interviewed in 2010 but not in 2013. In both 2013 and 2016, the sample grew by bringing in additional households that were established or joined by individual movers that moved away from their interview locations in the prior survey round. As such, by 2016, there were new individuals that may have been interviewed for the first time, as well as individuals that may have been interviewed once or twice since 2010. Concerning the individuals that were interviewed in 2016 and that may have had received the same questions in the past, we assume any potential Hawthorne effects to be minimal given the passage of time between the interviews (3–6 years, depending on the individual).

1.1 days per IHPS EA (approximately 32% of the IHS4 average) were used for organising call backs to the sampled households and interviewing individual respondents.[13]

On the whole, in the absence of a true survey experiment, our analysis in this paper is limited to discussing the relative differences in labour statistics between the two surveys, without making absolute conclusions on the true estimates; although, we contend that the IHPS is closer to the truth. As discussed below, we do find a strong association of the individual-interview approach in the IHPS on higher reporting of labour outcomes relative to the IHS4, particularly among women and for longer recall periods. In the analysis, we also descriptively examine whether specific channels (such as lower incidence of proxy reporting and greater time spent with respondents in the IHPS) might be more likely to be associated with the overall differences we observe—understanding that isolating these effects is difficult without an experimental design. The findings point to important channels in respondent selection, which can be useful for further experimentation in national surveys.

3.3 Survey samples and correcting for non-response

Our analysis focuses on individuals aged 18–64 years, for whom the information was elicited during the overlapping months of IHS4–IHPS fieldwork.[14] In the IHPS sample, there were 4,377 self-reporting individuals aged 18–64 years, who were interviewed in private. The IHS4 sample had 13,718 individuals in the same age range (a combination of individuals that self-reported and individuals for whom the information was reported by a proxy).

To compare the IHS4 with the IHPS, we pool the data on 13,718 IHS4 respondents together with 4,377 IHPS respondents who self-reported and were interviewed in private. The combined sample includes 18,095 observations. Within the IHS4, relative to respondents reporting by proxy, we find that 60% of self-reporters tend to be significantly older and less educated, more likely to be female, to be the household head or spouse, and to live in smaller, poorer households with a higher dependency ratio than respondents who did not

13 We have also analysed the more detailed metadata and paradata collected as part of the Survey Solutions CAPI application used for the LSMS+-supported data collection as part of the Tanzania National Panel Survey (NPS) 2019–20. which operationalised the same approach to conducting private interviews with adult household members as in Malawi IHPS. The CAPI paradata in particular allow us to compute precise interview duration estimates for household versus individual interviews and specific questionnaire modules and questions within. In Tanzania, at the national-level, on average, enumerators spent a total of 270 minutes at a household, with 79 of those minutes devoted to the individual-level modules administered to each adult in the household—approximately 30% of their time. Considering only the individual-level modules, at the national-level, an interview with an adult household member took about 30 minutes, on average, with no significant differences between men and women. For the labour module, specifically, this took approximately 6.7 minutes, on average. Dividing the total survey budget for the Tanzania NPS 2019–20 by the total number of minutes spent across all interviews—both household and individual level, we found that a minute of survey data collection costs USD 1.75. Based on this estimate, we computed that, on average, it costs USD 334 to administer the household questionnaire in Tanzania, and an average of an additional USD 52 to interview an adult household member.

14 In view of seasonality of work and employment activities in Malawi, the sample is limited to the calendar months that were covered by both surveys (April 2016 through January 2017) to ensure cross-survey temporal comparability of the individual-level data.

report for themselves.[15] These differences are also in line with other studies that compare self-respondents and respondents-by-proxy within surveys (Bardasi *et al.*, 2011; Serneels *et al.*, 2017). We do not discard any IHS4 observations on the basis of self- versus proxy reporting since the official statistics would be based on the totality of these survey data, irrespective of the deviations from the international best practices.

Conversely, there are 969 IHPS respondents who were eligible but were unavailable for the individual interview, constituting a non-response rate of 16.8% (of a total of 5,776). To use the IHPS data set in a way that can gauge the accuracy of the IHS4 data, we focus only on the IHPS individuals who participated in individual interviews and correct for non-response. To do the latter, we calculate weights by first running a logistic regression of individual response status among the 5,776 adults eligible for individual interviews. The results from the logistic regression are presented in Appendix Table A1 and show that the correlates of respondents who participated in individual interviews were also quite similar to those self-reporting in the IHS4—more likely to be women, household head or spouse, living in smaller households with a lower wealth index, and with a higher dependency ratio.[16] Subsequently, we (1) take the inverse of the predicted response probability to construct the response weight variable for each IHPS adult household member who was interviewed in private; (2) winsorise the response weights at the top 3% to account for potential outliers; and (3) set it equal to 1 for all adults in the IHS4 sample. In the paper, all statistics are weighted using the response weight.

A general caveat to our regression-based non-response correction is that it is conditional on observables and that it can indirectly account for unobservable drivers of non-response depending on the extent of correlation between observable and unobservable attributes. Since the latter is not an empirically testable hypothesis, we concede that there may remain unobservable determinants of non-response that are not catered to.

3.4 Respondent and household characteristics across surveys

Table 1 presents descriptive statistics on selected respondent and household attributes, by the two surveys as well as response type. The overall IHS4 and IHPS samples are statistically

15 These results are available upon request.

16 Van den Broeck and Kilic (2019) use a similar approach to correct for attrition bias in panel data samples, in a study of labour market dynamics in Sub-Saharan Africa. Our right-hand-side predictors of 'response' include the following: (i) fixed effects for districts, interview months and enumerators; (ii) individual covariates including age, a dichotomous variable identifying females, a series of dichotomous variables on educational attainment, dichotomous variables identifying whether the individual is currently married and, separately, whether he/she is head/spouse of head and individual's number of months living away from the household over the past year; and (iii) household covariates, including household size, dependency ratio, and wealth index. The latter is a factor analysis-based index that is composed of (i) a series of dichotomous variables that capture the ownership of mortar, bed, table, chair, fan, air conditioner, radio, radio with flash drive/micro CD, TV, VCR, sewing machine, kerosene/paraffin stove, electric/gas stove, refrigerator, washing machine, bicycle, motorcycle, car, minibus, lorry, beer-brewing drum, sofa, coffee table, cupboard/drawers, lantern, desk, clock, iron, computer, satellite dish, solar panel and generator; and (ii) a series of dwelling covariates, including number of dwelling rooms per capita and categorical variables that identify construction material (permanent, semi-permanent, traditional), roof type (grass, iron sheets, clay tiles/concrete/plastic sheeting/other), floor type (sand, smoothed mud, smooth cement/wood/tile/other), water source (piped/well, borehole, other) and toilet facility (flush/VIP toilet, traditional latrine, other/none).

similar; individuals in the IHS4 are somewhat more likely to never have attended school, although the other educational (primary/secondary schooling) outcomes are not statistically different. IHS4 households also tend to be smaller and with a slightly higher dependency ratio, although the magnitude of the difference is small.[17] Appendix Figure A2 also presents the age distributions of both samples; there is no statistical difference for women, as well as across most ages for men, except for a slightly higher share of men aged 20–25 years in the IHPS. We do see significant differences between the IHPS sample and the respondents in the IHS4 that self-reported versus having a proxy report for them. Interestingly, the differences between the IHPS and each of these sub-samples tend to move in opposite directions. For example, the IHS4 self-reporting sample is significantly more likely than the IHPS sample to be female, but the IHS4 proxy reporting sample is substantially less so. The IHS4 self-reporting sample is also more likely to be married, the household head or spouse and have a smaller household size and lower wealth compared with the IHS4 proxy sample—with corresponding (and divergent) differences from the IHPS. There are also some differences in average age and education, although the magnitude of these differences is not large. In general, these comparisons also underscore factors associated with self-reporting versus proxy reporting status in the IHS4 and which we control for in the regression analysis along with geographic and interview fixed effects.

3.5 Labour market outcomes across surveys and gender

Table 2 presents the descriptive statistics for labour market outcomes of interest, by gender. As discussed earlier, the questionnaires for the IHS4 and IHPS were the same, and followed the same question wording and implementation by the same teams of enumerators, coverage of the two samples by survey strata and interview month were also similar. The estimates are coupled with the results from the tests of mean differences across the two surveys. First, we analyse a range of dichotomous variables regarding participation in specific activities during the past 7 days and during the past 12 months—constructed from the answers given to the screening questions asked at the beginning of the labour module to account for participation in the following activities:

(1) **crop production** (i.e., household farming activities whether for sale or for household food);

(2) **raising livestock** (i.e., household livestock activities whether for sale or for household food);

(3) **managing an NFE** (i.e., run or do any kind of non-agricultural or non-fishing household business, big or small, for yourself);

(4) **working in an NFE** (i.e., help in any of the household's non-agricultural or non-fishing household businesses);

(5) **engaging in off-farm wage employment** (i.e., any work for a wage, salary, commission or any payment in kind, excluding *ganyu*); and

(6) **ganyu employment** (i.e., engage in casual, part-time or *ganyu* labour).

17 These differences are due to the survey setting, as the IHS4 aims to represent the national population of Malawi in 2016, while the IHPS is based on the national population in 2010.

Table 1: Descriptive Statistics by Survey Approach and Response Type

	IHPS	IHS4					
	IHPS: total	IHS4: total	*T*-test (1)–(2)	IHS4: self-report	*T*-test (1)–(3)	IHS4: proxy report	*T*-test (1)–(4)
	(1)	(2)		(3)		(4)	
Age (years)	33.0	33.3		35.3	***	30.4	***
Female	0.52	0.53	*	0.62	***	0.39	***
Never attended school	0.09	0.12	***	0.12	***	0.12	***
No primary degree	0.52	0.51		0.55		0.45	***
Primary degree	0.12	0.12		0.11		0.13	
Lower secondary degree	0.12	0.12		0.10	*	0.14	**
Higher secondary degree	0.15	0.12		0.10	**	0.15	
Married	0.68	0.66		0.73	***	0.55	***
Household head or spouse	0.73	0.75		0.89	***	0.55	***
Log household size	1.61	1.53	***	1.45	***	1.64	*
Dependency ratio	0.89	0.94	*	0.99	***	0.84	
Wealth index	0.12	0.19		0.08		0.35	**
Observations	4,377	13,718		8,219		5,499	

Notes: IHPS, individual interviews; IHS4, business-as-usual survey approach. Estimates are weighted with the response weight.
*$p < 0.1$.
**$p < 0.05$.
***$p < 0.01$.

For crop production, our analysis also distinguishes between crop production intended mainly for sale versus household consumption, to better understand how the survey approach affects the measurement of crop employment under the revised 19th ICLS definitions of work and employment. This has been a particularly challenging area for economies with a large share of smallholder farmers, for whom paid and subsistence work often vary seasonally throughout the year and where women often have economically productive but hidden roles (Benes and Walsh, 2018c; Koolwal, 2021). Over the past few years, country labour force surveys and the surveys supported under the World Bank LSMS-ISA initiative have been introducing individual-level questions on the main intended destination of agricultural output to better elicit employment in agriculture under the new definitions.

Subsequently, we examine hours allocated to each of these activities during the past 7 days, as well as annual hours allocated to and earnings from wage and, separately, *ganyu* employment.[18] For now, we do not distinguish within the IHS4 between self-reporting

18 Additional results including zeroes for the non-employed (i.e., unconditional sample) are available upon request. There were more significant differences in the unconditional sample owing to differences in participation across the two surveys. For weekly hours, we assume that no one can work more than 12 hours per day during the past 7 days such that weekly hours are capped at 84 hours. We replace missing values with the median value of weekly hours within the EA and take the logarithm conditional on participation. Annual hours are calculated based on the average number of hours per week spent on wage employment multiplied with the average number of weeks per month and the

Table 2: Comparing Individual-Level Labour Outcomes by Survey Approach and Gender

	Men			Women			Gender differences	
	IHPS (1)	IHS4 (2)	*T*-test (1)–(2)	IHPS (3)	IHS4 (4)	*T*-test (3)–(4)	IHPS (1)–(3)	IHS4 (2)–(4)
Labour activities (Y = 1, N = 0)								
Any crop work (7 d)	0.26	0.30	*	0.29	0.33	*	−0.03	−0.03***
At least 1 crop mainly for sale (7 d)[†]	0.42	0.39		0.33	0.36		0.09***	0.03***
All for household own use (7 d)[†]	0.44	0.52	**	0.59	0.56		−0.15***	−0.04***
Any crop work (12 m)	0.69	0.72		0.74	0.75		−0.05***	−0.03***
Livestock (7 d)	0.19	0.18		0.20	0.17	*	−0.01	0.01*
Livestock (12 m)	0.33	0.31		0.31	0.30		0.02	0.01
Management of NFE (7 d)	0.14	0.12	***	0.13	0.08	***	0.01	0.04***
Management of NFE (12 m)	0.20	0.16	***	0.20	0.14	***	0.005	0.02***
Working in NFE (7 d)	0.04	0.03	**	0.04	0.03	*	−0.001	−0.001
Working in NFE (12 m)	0.09	0.05	***	0.08	0.06	*	0.01*	−0.01***
Wage employment (7 d)[‡]	0.18	0.15	**	0.06	0.06		0.12***	0.09***
Wage employment (12 m)[‡]	0.22	0.18	**	0.08	0.07		0.14***	0.11***
Ganyu employment (7 d)	0.22	0.24		0.17	0.16		0.05***	0.08***
Ganyu employment (12 m)	0.47	0.45		0.36	0.36		0.11***	0.09***
Weekly hours, among employed								
Any crop work	15.4	14.8		12.9	14.0		2.5***	0.8***
Livestock	7.4	6.5		5.1	5.4		2.3***	1.1***
Management of NFE	30.1	27.7		22.2	21.3		7.9***	6.4***
Working in NFE	19.5	19.3		19.0	16.0		0.5	3.3**
Wage employment	39.7	36.4	***	32.1	35.0		7.6***	1.4***
Ganyu employment	18.6	17.4		10.3	11.8	**	8.3***	5.6***

(Continued)

Table 2: Continued

	Men			Women			Gender differences	
	IHPS (1)	IHS4 (2)	*T*-test (1)–(2)	IHPS (3)	IHS4 (4)	*T*-test (3)–(4)	IHPS (1)–(3)	IHS4 (2)–(4)
Annual hours, among employed								
Wage employment	1,401.2	1,358.6		1,107.8	1,247.0		293.4***	111.6**
Ganyu employment	438.1	474.9	*	329.6	328.0		108.5***	146.9***
Annual earnings, among employed (1000 s Kwacha)								
Wage employment	851.3	1,897.6		1,038.8	846.6		−187.5	1,051
Ganyu employment	151.2	131.8		63.5	58.7		87.7***	73.1***
Observations	1,966	6,447		2,411	7,271			

Notes: IHPS, individual interviews; IHS4, business-as-usual survey approach. Estimates are weighted with the response weight.

*$p < 0.1$.

**$p < 0.05$.

***$p < 0.01$.

†Indicates that means are calculated among those involved in crop agriculture (1,215 individuals in the IHPS and 5,724 in the IHS4).

‡Wage employment excludes *ganyu*.

respondents and respondents-by-proxy, although we do discuss potential issues around proxy reporting in the IHS4 further in the paper.

Table 2 shows that in the IHS4 nearly all activities (with the exception of men's own-use work in agriculture) have the same or lower participation rates as compared with the IHPS. The difference in reporting of NFE employment, particularly for men, also tends to widen when the recall period is longer. Although participation rates might differ, the ranking of importance of activities is similar in both scenarios, with cropping as the most common activity, then *ganyu* and livestock, management of an NFE, followed by off-farm wage employment and supporting work in an NFE. Further, Table 2 shows that hours worked and earnings among those employed are not statistically different (or the differences are not very large) across survey approaches for most activities, with some variations by gender (wage employment for men and *ganyu* for women). The choice of survey approach may also matter more with longer recall periods as well as seasonality of different activities among this population.

Looking at gender differences, the last two columns of Table 2 show that women are significantly more likely than men under both survey approaches to work in crop agriculture over the last 12 months—particularly in own-use production as opposed to market activity—and less likely to work in wage employment and *ganyu*. While men are significantly more likely to report management of an NFE than women under the business-as-usual approach, this difference disappears in the IHPS. Generally, Table 2 also shows that men's participation and weekly hours across wage and *ganyu* work, as well as hours in agriculture and running an NFE rises in the IHPS relative to the IHS4 sample by a greater factor as compared with women, resulting in wider gender gaps for the outcomes in the IHPS sample. Annual hours in wage employment are also higher for men in the IHPS sample (about 290 hours more on average, or roughly 35 eight-hour working days). Significant gender differences also emerge for annual *ganyu* earnings in the IHPS and IHS4—again with greater differences in the IHPS.

Figure 1 looks at gender differences across samples by age and demonstrates that under the IHPS, the levels of reported wage employment and NFE management are higher for nearly all men across the age distribution, but for women the largest increases in these activities are focused on younger age groups (around 30 years of age; for NFE management there is an additional increase among older women (55+) in the IHPS as well). For women, lower time constraints outside their child-rearing years may be one reason for these trends.

Figure 2 depicts the kernel density distribution of conditional annual hours (top panel) and earnings (lower panel) across wage and *ganyu* employment, for both men and women. Annual hours in wage employment for men and women tend to follow a bimodal distribution, indicating that people are either part-time or full-time wage employed, whereas hours in *ganyu* employment tends to be unimodal and focused around the lower concentration of wage employed (possibly reflecting, to some degree, that part-time wage employed may also

total number of months during the last 12 months. For *ganyu* employment, we do not have data on the number of hours per week, so we assume an average of 6 hours per day multiplied with the number of days per week spent on *ganyu*. We replace unrealistically high values with a maximum of 4,368 hours as we assume that no one can work more than 12 hours per day all year round. Annual earnings represent net income and are derived from all cash and in-kind remuneration that a worker received during the past 12 months. We winsorise earnings at top 1%, replace missing values with the median value within the EA and take the logarithm conditional on employment.

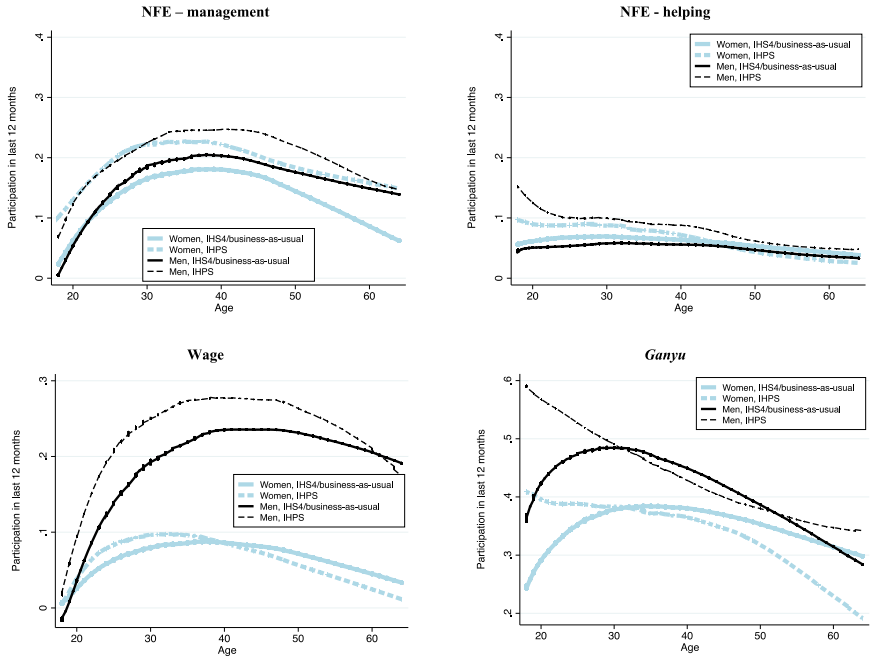

Figure 1: Locally Weighted Regressions: Participation (Last 12 Months) in Off-Farm Activities, By Survey Approach and Gender

be supplementing their income with *ganyu*). For women, relative to the IHPS, there is a higher concentration of individuals employed full-time and a lower share among those working part-time/fewer hours under the business-as-usual scenario. The distributions for men, on the other hand, indicate a higher share of both part-time and full-time wage employed (albeit a smaller increase among part-time workers/those with fewer hours). And regarding hours in *ganyu*, the business-as-usual approach leads to a higher concentration of men and women around the middle of their respective distributions. As indicated in Figure 2, the cross-survey differences are statistically significant in the comparisons of the distributions of (i) *ganyu* hours and earnings for women and men and (ii) wage hours for women. The IHPS distributions in these cases take on lower values vis-à-vis their IHS4 counterparts.

4. Empirical analysis

To understand the links between choice of survey approach on labour reporting, we estimate the following equation separately for the samples of men and women aged 18–64 years:

$$y_{ih} = \propto + \beta_1 T_{1ih} + \gamma D + \varepsilon_{ih}, \quad (1)$$

where i and h represent individual and household, respectively; y represents individuals' labour outcomes (as presented in Table 2); and α and ε represent constant and error terms, respectively. T_1 is a binary variable taking the value 1 for the IHS4 sample, and 0 for individ-

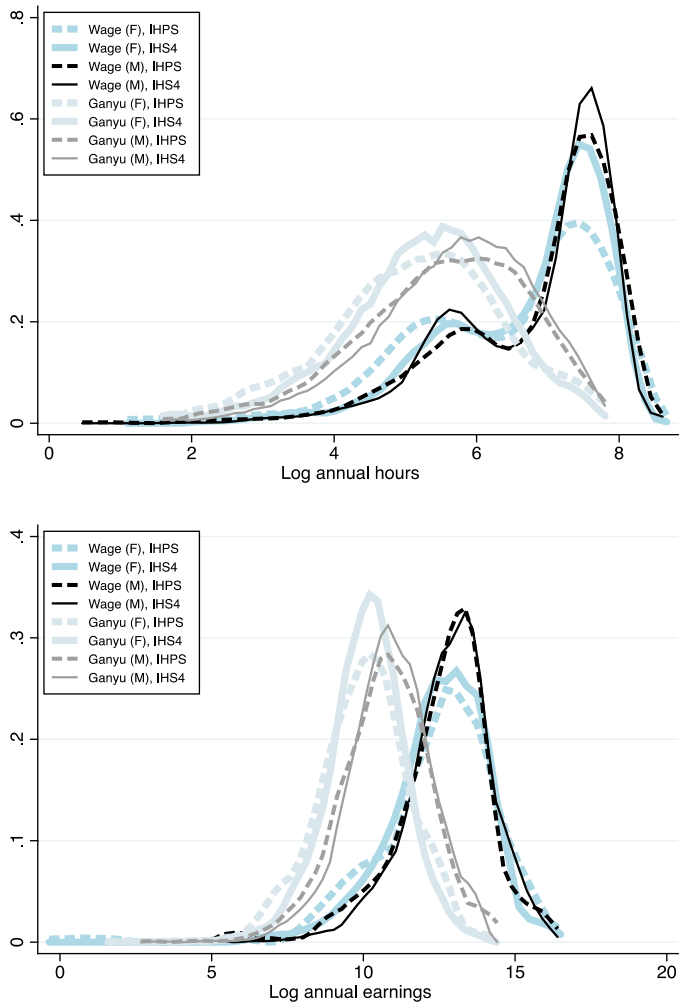

Figure 2: Kernel Density Distribution of Log Annual Hours in and Log Earnings From Wage and *Ganyu* Employment, By Survey Approach and Gender

uals in the IHPS; and D is a vector of individual and household attributes, as well as fixed effects for districts, interview month and enumerators, to capture any remaining unobserved heterogeneity that may also jointly determine both the dependent variable and the likelihood of being selected for one survey versus the other.[19] As noted above, besides the emphasis

19 Individual covariates in D include age, whether the respondent is female, a series of binary variables on educational attainment (never attended school/attended but did not complete primary/completed primary/completed secondary), whether the respondent is married, and whether he/she is head/spouse of head. Household covariates in D include whether the household is in a city, household size, dependency ratio and wealth index, the construction of which is discussed in Section 3.2.

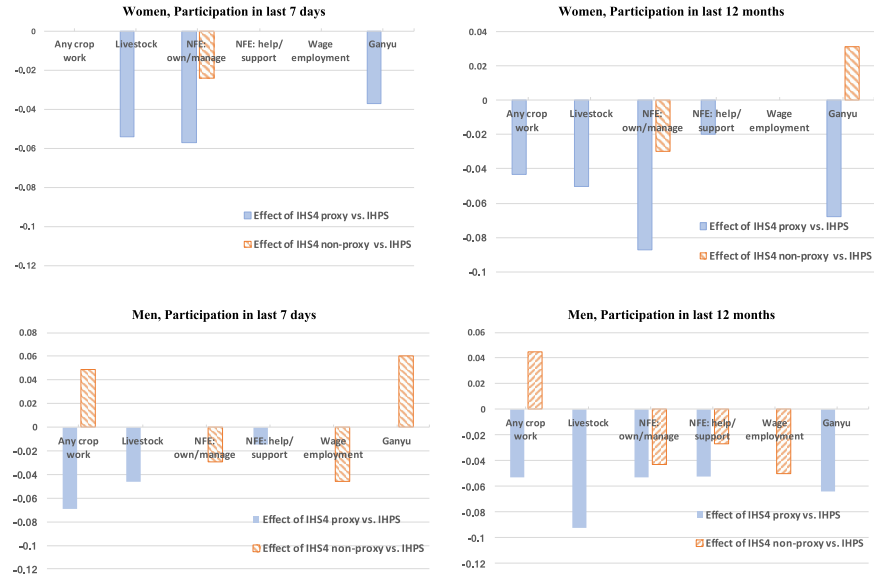

Figure 3: Significant Effects ($p \leq 0.05$) of the Business-as-Usual Survey Approach, by Whether Individual Self-Reported

on conducting private interviews with adult household members, the IHPS also attempted, albeit imperfectly, to (i) achieve a gender match between the respondents and the enumerators and (ii) conduct multiple intra-household interviews simultaneously. These were not among the design features of the IHS4. In the absence of a true experimental design that allows us to gauge the relative impacts of (i) self-reporting in private interviews, (ii) respondent-enumerator gender match and (iii) simultaneous interviews (in applicable households with multiple interview targets), the analysis aims to gauge whether the measurement of labour outcomes is impacted by the business as usual survey setting vis-à-vis the individual interview approach that should be understood *as a package of design choices*. The enumerator and interview date fixed effects do assist to some extent in partialing out the potential role of gender-matching of respondents and enumerators in the IHPS, as well as other variations related to interview duration and timing—to better link, as best as possible, the comparisons of IHPS and IHS4 to respondent selection. Nevertheless, the coefficient β_1 in Equation (1) still reflects to a large degree this set of design choices within the business-as-usual approach as opposed to one design element.

For the binary outcome variables, Equation (1) is estimated as a linear probability model with probability weights adjusting for non-response.[20] For weekly hours across activities, and annual hours and earnings in wage and *ganyu*, OLS regressions are run on the conditional sample of respondents (those who reported working in that activity over the relevant

20 The reported regression coefficients are qualitatively similar to the marginal effects derived from logit and probit regressions, which are available upon request.

reference period), and logging the outcome variables.[21] Standard errors are clustered at the EA level, which is the community unit on which survey sampling was based.

Following the estimations of Equation (1), Tables 3 and 4 summarise the association of the business-as-usual survey setting with labour reporting (for the full sample, and for men and women separately). After controlling for demographic and survey characteristics, we find that relative to the IHPS, the estimates of participation in several different areas of employment, for men and women, are lower under the business-as-usual scenario. This includes participation in livestock activity (for women, across both the 7-day and 12-month recall periods; for men, with the 12-month recall). In addition, for NFE-related activities, the business-as-usual scenario is associated with lower reporting of men's and women's work in (i) managing an NFE, across both recall periods, and (ii) supporting work in an NFE in the last 12 months. Relative to the IHPS, the estimates of men's participation in wage employment are lower under the business-as-usual scenario. The magnitude of lower reporting in the aforementioned activities under the business-as-usual approach also tends to increase with the longer (i.e., 12 months) recall period.

Although the effects are seemingly limited in absolute terms,[22] they are in fact quite large in relative terms, vis-à-vis the average participation rates in Table 2, and the effects tend to be stronger for women over the 7-day recall period. Among women, for example, Table 4 shows lower reporting for the last 7 days in the business-as-usual approach ranges from 2.6 percentage points for livestock (which, given mean participation for women in the IHS4, is a relative decrease of 15%) and 3.6 percentage points for NFE management (a relative decrease of about 45%); for men, relative decreases in reporting over the last week is 21% for NFE management. We do observe higher men's participation in *ganyu* over the last week under the business-as-usual approach, although the effect disappears for the 12-month recall period (and there is no significant effect for the overall sample).

Looking at how the choice of survey approach affects time in employment, we find that under the business-as-usual approach, reported weekly hours are lower in (i) livestock activities for women (coefficient of −0.11, which vis-à-vis average hours worked in Table 2, amounts to about 35 minutes less per week on the conditional sample of employed) and (ii) across a broader range of activities for men, including crop agriculture, livestock, NFE management and wage employment (effects when evaluated at average weekly hours for the conditional sample, range from about 1.1 and 1.8 hours less per week in livestock and crop agriculture, respectively, to 3 and 3.5 hours less per week in NFE and wage work, respectively). We find that relative to the IHPS, annual hours in *ganyu* employment for both men and women and reported earnings in *ganyu* for men are higher under the business-as-usual scenario.

The identified differences across the surveys entail several implications. The business-as-usual approach in the IHS4 tends to underestimate employment rates and employment duration relative to the IHPS, especially for livestock activities and activities taking

21 Since the survey experiment/setting is exogenous, and we are focused on the conditional sample, we do not use a selection model in the analysis.

22 Other studies examining recall-related bias do find larger effects (Bardasi *et al.*, 2011; Arthi *et al.*, 2018); however, it remains difficult to compare our effects directly with these studies, since we do not have a single 'treatment' of proxy reporting.

Table 3: Estimates of the Impact of Business-as-Usual Survey Approach on Labour Outcomes of All Adults Aged 18–64 Years

	Any crop work (all HH)	Within crop agriculture:		Livestock	NFE:		Wage	*Ganyu*
		At least one crop mainly for sale	All crops for HH own use		Ownership/ management	Helping/ supporting		
	(1)	(2a)	(2b)	(3)	(4a)	(4b)	(5)	(6)
Total sample								
Participation (7 d)	−0.004	0.029	−0.008	−0.021*	−0.030***	−0.007*	−0.007	0.01
	[−0.22]	[1.26]	[−0.33]	[−1.90]	[−3.91]	[−1.66]	[−1.06]	[1.15]
R-squared	0.26	0.14	0.17	0.10	0.06	0.03	0.22	0.11
Participation (12 m)	−0.011	-	-	−0.038***	−0.049***	−0.028***	−0.014*	−0.01
	[−0.93]			[−2.84]	[−5.88]	[−4.35]	[−1.76]	[−0.88]
R-squared	0.38	-	-	0.16	0.08	0.05	0.24	0.22
Observations	*18,095*	*5,507*	*5,507*	*18,095*	*18,095*	*18,095*	*18,095*	*18,095*
Weekly hours (cond.)	−0.027	-	-	−0.131***	−0.039	−0.270**	−0.038	0.031
	[−0.66]			[−2.74]	[−0.91]	[−2.55]	[−0.91]	[0.75]
R-squared	0.25	-	-	0.25	0.30	0.35	0.29	0.21
Observations	*5,565*	-	-	*3,273*	*1,969*	*662*	*1,943*	*3,557*

(Continued.)

Table 3: Continued

	Any crop work (all HH)	Within crop agriculture:		Livestock	NFE:		Wage	*Ganyu*
		At least one crop mainly for sale	All crops for HH own use		Ownership/ management	Helping/ supporting		
	(1)	(2a)	(2b)	(3)	(4a)	(4b)	(5)	(6)
Annual hours (cond.)	-	-	-	-	-	-	0.027	0.125***
							[0.46]	[3.21]
R-squared	-	-	-	-	-	-	0.26	0.17
Annual earnings (cond.)	-	-	-	-	-	-	0.024	0.083*
							[0.29]	[1.88]
R-squared	-	-	-	-	-	-	0.48	0.21
Observations	-	-	-	-	-	-	*2,274*	*7,292*

Notes: The results are from linear probability models for binary outcomes and from linear regressions for hours and earnings. Coefficients represent the impact of the business-as-usual survey approach, originating from regressions that also include the control variables in Table 1, and district, month and enumerator fixed effects. The regressions are weighted by the response weight, and *t*-statistics accounting for clustering at the enumeration-area level are presented in brackets.

*$p < 0.1$).

**$p < 0.05$.

***$p < 0.01$.

Table 4: Estimates of the Impact of Business-as-Usual Survey Approach on Labour Outcomes of Men and Women Aged 18–64 Years

		Within crop agriculture:		Livestock	NFE:		Wage	*Ganyu*
	Any crop work (all HH)	At least one crop mainly for sale	All crops for HH own use		Ownership/ management	Helping/ supporting		
	(1)	(2a)	(2b)	(3)	(4a)	(4b)	(5)	(6)
Women								
Participation (7 d)	0.006	0.033	−0.043	−0.026**	−0.036***	−0.005	−0.0002	−0.005
	[0.29]	[1.37]	[−1.61]	[−2.13]	[−3.75]	[−1.19]	[−0.02]	[−0.48]
R-squared	0.29	0.16	0.17	0.11	0.07	0.04	0.18	0.12
Participation (12 m)	−0.018	-	-	−0.030**	−0.050***	−0.016**	−0.001	0.003
	[−1.37]			[−2.11]	[−4.89]	[−2.28]	[−0.10]	[0.21]
R-squared	0.39	-	-	0.16	0.09	0.05	0.21	0.23
Observations	*9,682*	*3,072*	*3,072*	*9,682*	*9,682*	*9,682*	*9,682*	*9,682*
Weekly hours (cond.)	0.055	-	-	−0.112**	0.029	−0.147	0.042	0.158***
	[1.09]			[−2.08]	[0.51]	[−1.05]	[0.49]	[2.75]
R-squared	0.28	-	-	0.28	0.33	0.45	0.48	0.22
Observations	*3,107*	-	-	*1,726*	*929*	*359*	*586*	*1,601*
Annual hours (cond.)	-	-	-	-	-	-	0.233*	0.103**
							[1.85]	[1.99]
R-squared	-	-	-	-	-	-	0.36	0.20
Annual earnings (cond.)	-	-	-	-	-	-	0.101	0.046
							[0.60]	[0.74]
R-squared	-	-	-	-	-	-	0.55	0.18
Observations	-	-	-	-	-	-	*688*	*3,481*

(Continued)

Table 4: Continued

	Any crop work (all HH)	Within crop agriculture: At least one crop mainly for sale	All crops for HH own use	Livestock	NFE: Ownership/ management	Helping/ supporting	Wage	*Ganyu*
	(1)	(2a)	(2b)	(3)	(4a)	(4b)	(5)	(6)
Men								
Participation (7 d)	−0.01	0.012	0.047	−0.015	−0.025**	−0.009	−0.019*	0.029**
	[−0.57]	[0.36]	[1.58]	[−1.15]	[−2.51]	[−1.41]	[−1.66]	[2.27]
R-squared	0.24	0.17	0.20	0.11	0.08	0.04	0.23	0.12
Participation (12 m)	−0.003	-	-	−0.047***	−0.049***	−0.042***	−0.031***	−0.023
	[−0.18]			[−2.94]	[−4.05]	[−4.53]	[−2.62]	[−1.40]
R-squared	0.39	-	-	0.17	0.09	0.06	0.25	0.23
Observations	*8,413*	*2,435*	*2,435*	*8,413*	*8,413*	*8,413*	*8,413*	*8,413*
Weekly hours (cond.)	−0.123**	-	-	−0.171**	−0.111**	−0.135	−0.100**	−0.081
	[−2.33]			[−2.47]	[−2.01]	[−0.77]	[−2.46]	[−1.46]
R-squared	0.26	-	-	0.28	0.33	0.54	0.32	0.22
Observations	*2,458*	-	-	*1,547*	*1,040*	*303*	*1,357*	*1,956*
Annual hours (cond.)	-	-	-	-	-	-	−0.044	0.154***
							[−0.74]	[3.07]
R-squared	-	-	-	-	-	-	0.28	0.15
Annual earnings (cond.)	-	-	-	-	-	-	0.026	0.120**
							[0.34]	[2.12]
R-squared	-	-	-	-	-	-	0.50	0.19
Observations	-	-	-	-	-	-	*1,586*	*3,811*

Notes: The results are from linear probability models for binary outcomes and from linear regressions for hours and earnings. Coefficients represent the impact of the business-as-usual survey approach, originating from regressions that also include the control variables (apart from the female dummy) in Table 1, and district, month and enumerator fixed effects. The regressions are weighted by the response weight, and *t*-statistics accounting for clustering at the enumeration-area level are presented in brackets.

*$p < 0.1$.

**$p < 0.05$.

***$p < 0.01$.

place outside the farm-household. This implies that participation in the (non-farm) labour market might be higher in reality, which crop production might comprise a lower share in overall economic activities, and that labour productivity might be lower in those activities that are covered by surveys. This all bears consequences for measuring sources of income more accurately and, in turn, poverty and tracking the process of structural transformation. Moreover, the gender differences imply that entrenched gender employment gaps might be even higher for some types of activities, such as working in an NFE and wage employment.

5. Mechanisms behind labour reporting differences across surveys

What are underlying factors that may be associated with these trends? In this section, we conduct a descriptive analysis of potential channels, by looking at key differences across the IHS4 and IHPS survey approaches. Figure 3, for example, presents estimates from the same regressions in Table 4, but instead comparing the effects of (a) the IHS4 proxy sample and (b) the IHS4 non-proxy sample, with the IHPS. Full results are presented in Appendix Tables A2 and A3. The results show that one possible hypothesis for the differences in labour reporting across the surveys might be associated with proxy reporting in the IHS4, although again isolating this effect from other design differences across the surveys (including time spent with respondents) would require additional methodological experimentation. When examining the IHS4 individuals by whether they reported for themselves and estimating the same regressions as in Tables 3 and 4, we find that the IHS4 proxy sample contributes more to lower reporting than the own-reporting/non-proxy sample, particularly in agricultural and NFE activities.

Another potential hypothesis for lower reporting under the business-as-usual approach may be due to how respondents interpret questions when they have less dedicated time with the enumerator. This mechanism helps to explain the results in Figure 3 showing that the IHS4 non-proxy sample reports lower NFE and wage employment. Recent cognitive studies conducted by the ILO across countries, for example, have found that many respondents in highly informal and agricultural contexts can have difficulty—without detailed contextual and follow-up questions—on identifying their work in an enterprise or casual wage work (Benes and Walsh, 2018c).

Table 5 aims to better understand whether this mechanism is at work for NFE activity, where lower reporting under the business-as-usual approach is more systematic across the wealth distribution and for both men and women. Specifically, Table 5 compares discrepancies in reporting of NFE outcomes in the labour module with reporting of the same outcomes from the enterprise module of each survey. Specifically, the enterprise module in each survey creates a roster of all enterprises that household members are involved in, and for each enterprise, one person (almost always the owner/manager, discussed below) reported on roles across household members on owners, managers, as well as those in supporting roles. In the IHPS, self-reporting of outcomes in the enterprise module—i.e., selecting the owner/manager to report, as opposed to another household member—was emphasised. There were cases of proxy respondents in the enterprise module (those that neither owned/managed nor worked in the enterprise), although for the sample of individuals used in the analysis, the share of

proxy response was lower in the IHPS (6.7%, or 64 individuals) compared with the IHS4 (10.6%).[23]

Table 5 examines whether there are discrepancies across the labour and NFE modules in the shares of men and women reporting they run (own/manage) the NFE. We compare these differences across men and women in the IHPS and IHS4, for all individuals and among those living in households with an enterprise. We find that the survey approach in the IHPS tends to result in lower significant differences between the labour and NFE modules for the overall sample, compared with the IHS4. For example, among men and women in the IHPS, about 20% reported owning/managing an NFE in the labour module, compared with 18% in the NFE module, and these differences were not statistically significant. In the IHS4, on the other hand, there are statistically significant differences in reporting across the labour and NFE modules. These gaps also widen significantly among individuals in households with an enterprise (an increase from 37% to 51% in women's ownership/management moving from the labour module to the NFE module, or a 14-percentage point increase) as compared with men (a 12-percentage point increase).

The last two columns of Table 5 also show that proxy response in the NFE module leads to a wider variation in statistics on ownership/management of NFEs, particularly for men. There was, otherwise, no significant difference in the share of NFE owners across the self-reporting and overall samples. Along with issues around interpreting one's role in a business, therefore, proxy reporting might be another issue to consider, although larger differences in reporting seem to emerge across the labour and NFE modules than by proxy response (as discussed as part of the related regressions below). Again, however, while it was an important difference from the IHPS, the effect of proxy reporting within the IHS4 is not completely isolated in our analysis. The comparisons with the proxy and non-proxy samples are provided to pose some important questions for survey design which could be confirmed in future experimentation.

Figure 4 further looks at specific cases where individuals were identified as owners/managers of a business in one module, but not the other. Again, the NFE module captures a greater share of individuals as owners/managers than the labour module; although the number of cases is not large, we do see much wider gaps across modules in the IHS4 compared with the IHPS, and the labour module in the IHPS also tends to capture a greater share of individuals owning/managing an enterprise. Table A4 in the Appendix also runs regressions of individuals' work in enterprises using data strictly from the enterprise module and finds very few differences by survey approach. Where lower reporting is significant under the business-as-usual approach, relative to the IHPS sample, the magnitudes of the coefficients are smaller compared with Table 4. Overall, therefore, lower reporting of enterprise/business participation under the business-as-usual approach is linked significantly to the level of detail of questions in the labour module, which the approach in the IHPS helps address.

Finally, to further understand whether reporting patterns in the business-as-usual scenario are stronger for certain subpopulations, Appendix Tables A5 and A6 summarise

23 The enterprise module provides a closer look at the types of enterprises that had a proxy respondent—nearly all (97%, or 120 out of 124) of these enterprises had only one owner/manager, with most of these (78%, or 94 out of 120) having no other supporting workers. These enterprises also tended to be owned by the husband, with the spouse reporting. Furthermore, among those enterprises with the owner/manager not reporting and which had workers, nearly all (20 out of 26) of IHPS respondents were working in that enterprise, compared with 69% in the IHS4 (69 out of 100 enterprises).

Table 5: Share of Individuals Running and Providing Supporting Work in an NFE, by Survey Approach: Comparisons Across Labour and NFE Modules

	Overall sample of men and women						NFE module: comparing overall sample with just self-reporting sample			
	Women			Men			Women		Men	
	Labour module	NFE module	*T*-test (1)–(2)	Labour module	NFE module	*T*-test (3)–(4)	NFE module: only self-reports	*T*-test (2)–(5)	NFE module: only self-reports	*T*-test (4)–(6)
	(1)	(2)		(3)	(4)		(5)		(6)	
(A) Share of individuals running NFE, in last year†										
IHPS/individual interviews										
Overall sample	0.20	0.18		0.20	0.18		0.17		0.17	
Observations	*2,411*	*2,411*		*1,966*	*1,966*		*2,395*		*1,919*	
Among HHs with enterprise	0.41	0.47	***	0.41	0.45	*	0.46		0.42	
Observations	*899*	*899*		*778*	*778*		*883*		*731*	
IHS4/business-as-usual										
Overall sample	0.14	0.17	***	0.16	0.18	***	0.16		0.16	***
Observations	*7,271*	*7,271*		*6,447*	*6,447*		*7,185*		*6,243*	
Among HHs with enterprise	0.37	0.51	***	0.41	0.53	***	0.50		0.49	**
Observations	*2,455*	*2,455*		*2,228*	*2,228*		*2,369*		*2,024*	

Notes: Estimates are weighted with the response weight.

*$p < 0.1$.

**$p < 0.05$.

***$p < 0.01$.

†Indicates that in the labour reporting module, the question asked of respondents was 'in the last 12 months, did you run a non-farm business of any size for yourself or the household, even if only for one hour?' In the NFE module (covering activity over the last 12 months), one respondent for each enterprise was asked to identify up to two household members for each of the following: (a) 'who manages this enterprise or is most familiar with it,' and (b) 'who owns this enterprise?'

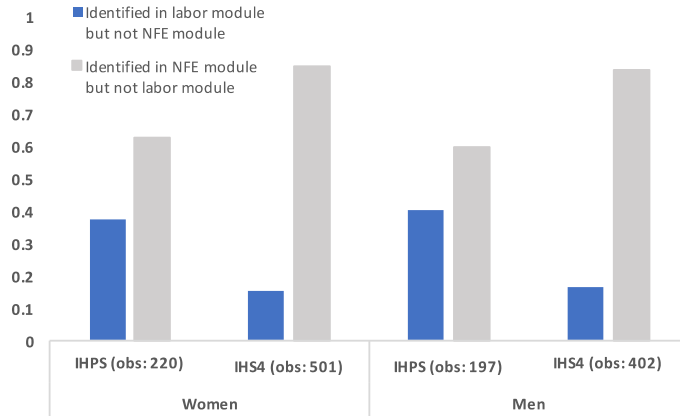

Figure 4: Discrepancies Across Labour and NFE Modules in Identifying Individuals Who Own/Run an NFE for IHPS Versus IHS4

the interaction effects (on participation in the last 7 days; results were very similar for participation in the last 12 months) of the business-as-usual survey approach with education, relation to household head, age, household wealth and location. The results are presented separately for the samples of men and women and show that along with age (being in the younger 15–24 age group), interaction effects with household wealth are associated with a greater range of labour participation outcomes for men and women.[24] For households with greater wealth under the business-as-usual approach, reporting of employment in market work in agriculture, as well as livestock, is significantly higher, but lower for NFE work (ownership/management). On the other hand, the interactions of the business-as-usual survey approach with relation to the household head, marital status and dependency ratio have relatively fewer significant effects across different labour outcomes.

Figure 5 examines how the business-as-usual approach affects reporting on participation in specific activities during the last 7 days *within* each wealth quintile (i.e., estimating equation ((1) separately for each wealth quintile). The trends in the 12-month participation are similar (though with larger effects) and are presented in Appendix Figure A3.[25] As seen in Tables 3 and 4, Figure 5 shows that the business-as-usual approach is associated with lower reporting of labour participation across most activities, relative to individual interviews, with the exception of men's employment in *ganyu* and with ambiguous effects on wage employment. Figure 5 shows that whether lower reporting is more prevalent among higher or lower wealth quintiles depends on sector of activity. Similar to the results in Appendix Tables A5 and A6, the business-as-usual approach is associated with much lower reporting of NFE work within higher wealth quintiles, which tend to be focused in these activities—with larger effects for ownership/management as opposed to supporting work. Lower reporting of work

24 We also found that wealth interactions with the business-as-usual approach also tend to have more widespread effects on participation (in the last 7 days or 12 months) as opposed to weekly/annual hours and annual earnings (results available upon request).

25 The full estimates are available upon request.

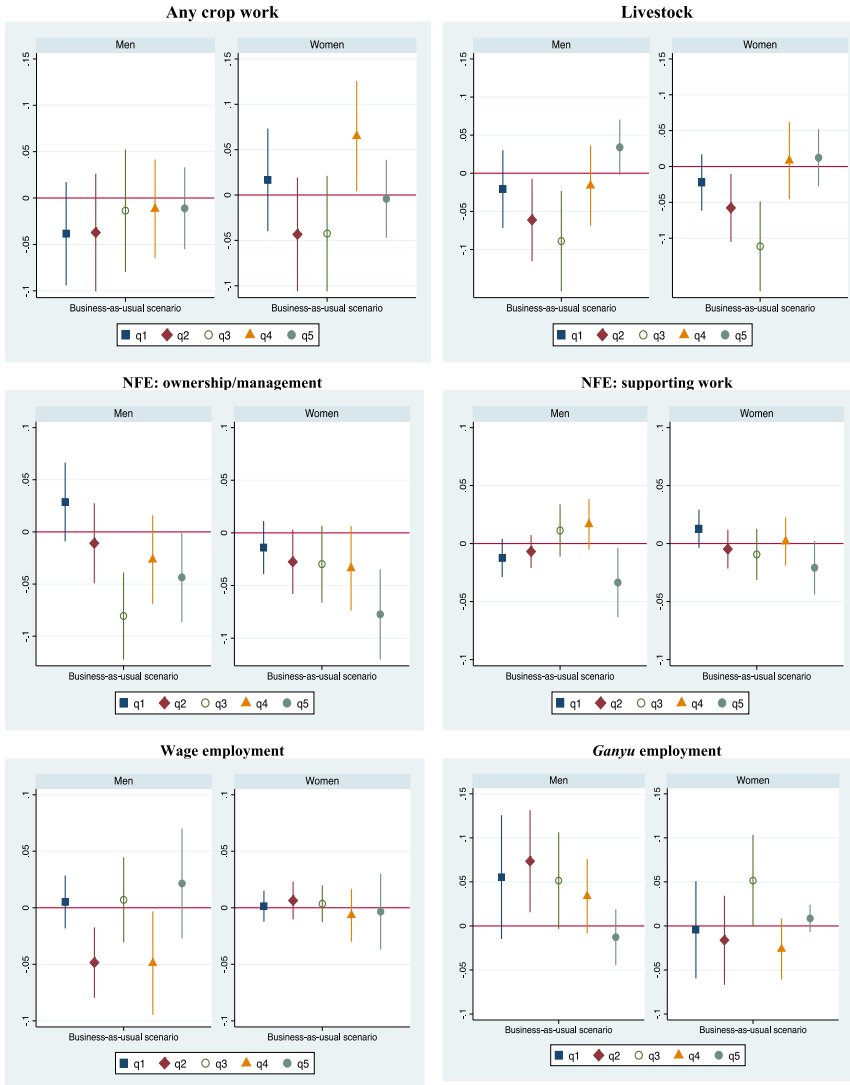

Figure 5: Effects of Business-as-Usual Approach on Labour Outcomes in the Last Seven Days by Wealth Quintile

in agriculture, on the other hand, is more likely among men and women in lower wealth quintiles, who are also more likely to be mainly concentrated in agriculture.

Generally, trends in reporting tend to be similar for women and men for self-employed activities. We do see some gender differences by quintile for wage and *ganyu* employment. On wage employment, for example, we find significantly higher reporting of men's employment associated with the business-as-usual approach in the highest quintile, but for women, relative to the individual interview approach, the business-as-usual scenario is associated with a close-to-zero effect, even though the average effects from the regressions in Tables 3 and

4 (available upon request) show that wage employment is positively associated with higher wealth quintiles for both men and women. Similarly, for *ganyu*, there is significantly greater reporting of men's employment among the lowest wealth quintiles, but again a close-to-zero effect for women for this quintile under the business-as-usual approach. Some of the gender differences may be due to norms around how women's paid work outside the household is viewed, particularly when women themselves are not reporting. Women are often engaged in several productive activities within and outside the home, for example, but their main roles may continue to be viewed as domestic instead of in paid labour (see Fox and Pimhidzai, 2013 and Comblon and Robilliard, 2017).

6. Conclusions

Accurate measurement of employment is at the heart of designing effective policies that improve economic opportunities for the poor and alleviate gender inequities in access to and returns from productive activities. Our paper leverages two national surveys that were implemented concurrently in Malawi with the same questionnaire and survey field teams and with similar distribution of survey samples across survey strata and interview months, but that relied on different survey approaches to collecting individual-disaggregated survey data. We find that, compared with conducting private individual interviews with adult household members to elicit self-reported information on labour outcomes, the business-as-usual approach (which encompasses a package of survey design differences, in addition to a greater reliance on proxy reporting) is associated with significantly lower reporting of employment in livestock-related activities and household NFEs (in management and worker capacities). These differences are more pronounced for women, and under a recall period of 12 months, as opposed to 7 days. The business-as-usual approach is also associated with lower reporting in the incidence of wage employment among men.

We also conduct a descriptive analysis to examine hypotheses on how proxy reporting in affecting these differences, as well as potential difficulties (given less time spent with enumerators) that respondents may face in interpreting and reporting on work that may be seasonal, such as casual wage work, as well as work in enterprises (as evidenced by greater discrepancies in reporting across the labour and NFE modules under the business-as-usual approach, with more pronounced differences for women). These hypotheses can be examined further in future randomised survey experiments. Finally, the impacts of the business-as-usual scenario are systematically associated with household wealth (significantly lower reporting of NFE employment in the highest wealth quintile, for example), highlighting the importance of respondent selection on specific socioeconomic subgroups as well as by gender.

Acknowledgements

The research was funded by the World Bank LSMS+ and LSMS-ISA programs. The authors would like to thank the Editor and an anonymous reviewer for their thoughtful and detailed feedback on improving the paper; and Kathleen Beegle and World Bank seminar participants for their feedback on the paper. The authors are also grateful for the exemplary hard work and exceptional professionalism on the part of the Malawi National Statistical Office management and survey teams in successful implementation of the Fourth Integrated Household Survey (IHS4) 2016–17 and the Integrated Household Panel Survey (IHPS) 2016.

Supplementary material

Supplementary material is available at *Journal of African Economies* online.

References

Abowd J. M., Stinson M. H. (2013) 'Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative Data', *The Review of Economics and Statistics*, 95(5): 1451–67.

Ambler K., Herskowitz S., Maredia M. K. (2021) 'Are We Done Yet? Response Fatigue and Rural Livelihoods', *Journal of Development Economics*, 153: 102736.

Arthi V., Beegle K., De Weerdt J., Palacios-López A. (2018) 'Not Your Average Job: Measuring Farm Labor in Tanzania', *Journal of Development Economics*, 130: 160–72.

Bardasi E., Beegle K., Dillon A., 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): 418–47.

Beegle K., Galasso E., Goldberg J. (2017) 'Direct and Indirect Effects of Malawi's Public Works Program on Food Security', *Journal of Development Economics*, 128: 1–23.

Benes E., Walsh K. (2018a) Measuring Main Activity in Labour Force Surveys: Main Findings from the ILO LFS Pilot Studies. International Labour Organization.

Benes E., Walsh K. (2018b) Measuring Unemployment and the Potential Labour Force in Labour Force Surveys: Main Findings From the ILO LFS Pilot Studies. International Labour Organization.

Benes E., Walsh K. (2018c) ILO LFS Pilot Studies Cognitive Interviewing Tests: Methodology, Process and Outcomes. International Labour Organization.

Bickart B. A., Blair J., Menon G., Sudman S. (1990) 'Cognitive Aspects of Proxy Reporting of Behavior', *Advances in Consumer Research*, 17: 189–206.

Bound J., Brown C., Mathiowetz N. A. (2001) 'Measurement error in survey data', in J.J. Heckman and E. Leamer (eds), *Handbook of Econometrics*, vol. 5. New York: Elsevier, pp. 3705–843.

Britton J., Shephard N., Vignoles A. (2019) 'A Comparison of Sample Survey Measures of Earnings of English Graduates With Administrative Data', *Journal of the Royal Statistical Society. Series A, Statistics in Society*, 182: 719–54.

Cobb C. (2018) 'Answering for someone else: proxy reports in survey research', in D.L. Vannette and J.A. Krosnick (eds), *The Palgrave Handbook of Survey Research*. Cham: Springer International Publishing, pp. 87–93.

Comblon V., Robilliard A.-S. (2017) 'Are Female Employment Statistics more Sensitive Than Male Ones to Survey Design?' *Evidence From Three West African Countries*, 87–93 DIAL Working Paper, DT/2015-22.

Cristia J., Schwabish J. A. (2009) 'Measurement Error in the SIPP: Evidence From Administrative Matched Records', *Journal of Economic and Social Measurement*, 34: 1–17.

Davis B., Di Giuseppe S., Zezza A. (2017) 'Are African Households (Not) Leaving Agriculture? Patterns of Households' Income Sources in Rural Sub-Saharan Africa', *Food Policy*, 67: 153–74.

Dillon A., Bardasi E., Beegle K., Serneels P. (2012) 'Explaining Variation in Child Labor Statistics', *Journal of Development Economics*, 98: 136–47.

Durazo J., Costa V., Palacios-Lopez A., Gaddis I. (2021) 'Employment and own-use production in household surveys: a practical guide for measuring labor', in *World Bank Living Standards and Measurement Study Program*.

Fox L., Pimhidzai O. (2013) 'Different Dreams, Same Bed: Collecting, Using, and Interpreting Employment Statistics in Sub-Saharan Africa: The Case of Uganda', World Bank Policy Research, Working Paper 6436.

Galdo J., Dammert A. C., Abebaw D. (2019) 'He Said, She Said: Child Labor Measurement in Ethiopia', Working Paper No. 46, Growth and Labour Markets in Low Income Countries Programme, IZA Institute of Labor Economics.

Hill D. H. (1987) 'Response errors in labor surveys: comparisons of self and proxy reports in the survey of income and program participation', in *Third Annual Research Conference*. University of Michigan: Washington D.C., 1987.

ILO (2013) 'Resolution I concerning statistics of work, employment and labour underutilization', in *19th International Conference of Labour Statisticians*. Geneva.

Kilic T., Moylan H., Koolwal G. (2021) 'Getting the (Gender-Disaggregated) Lay of the Land: Impact of survey respondent selection on measuring land ownership and rights', *World Development*, 146: 105545.

Koijen R., Van Nieuwerburgh S., Vestman R. (2015) 'Judging the quality of survey data by comparison with 'truth' as measured by administrative records: evidence from Sweden', in C. Carroll, T. Crossley and J. Sabelhaus (eds), *Improving the Measurement of Consumer Expenditures*. Cambridge: National Bureau of Economic Research.

Koolwal G. (2021) 'Improving the Measurement of Rural Women's Employment: Global Momentum and Survey Research Priorities', *World Development*, 147: 105627.

Mathiowetz N. A., Groves R. M. (1985) 'The Effect of Respondent Rules on National Health Survey Reports', *American Journal of Public Health*, 75(6): 639–44.

Moore J. C., Stinson L. L., Welniak E. J., Jr. (2000) 'Income Measurement Error in Surveys: A Review', *Journal of Official Statistics*, 16(4): 331–61.

Posel D., Casale D. (2005) 'Who replies in brackets and what are the implications for earnings estimates?' in *An Analysis of Earnings Data From South Africa*, Working Papers 07. Economic Research Southern: Africa.

Serneels P., Beegle K., Dillon A. (2017) 'Do Returns to Education Depend on How and Whom You Ask?' *Economics of Education Review*, 60: 5–19.

Tamborini C. R., Kim C. H. (2013) 'Are Proxy Interviews Associated With Biased Earnings Reports? Marital Status and Gender Effects of Proxy', *Social Science Research*, 42: 499–512.

United Nations Statistics Division (UNSD) (2019) Guidelines for producing statistics on asset ownership from a gender perspective.

Van den Broeck G., Kilic T. (2019) 'Dynamics of Off-Farm Employment in Sub-Saharan Africa: A Gender Perspective', *World Development*, 119: 81–99.

World Bank (2019) World Bank Overview: Malawi.