Different Dreams, Same Bed: Collecting, Using, and Interpreting Employment Statistics in Sub-Saharan Africa - The Case of Uganda

Louise Fox

Obert Pimhidzai

Authors can be contacted at lfox@worldbank.org and opimhidzai@worldbank.org. Useful comments and suggestions were received from Kathleen Beegle.

Abstract

Employment and earnings statistics are among the most important variables used to identify the causes of poverty as well as the pathways out - for households and for an economy. They are the key link between the size and structure of economic growth and the welfare of households which is the ultimate goal of development policy so it is important to monitor employment outcomes consistently. A cursory review of employment data for low income Sub-Saharan African countries shows both large gaps and improbable variation within countries over time and among countries, suggesting that low quality data is routinely reported by national statistics offices. Unfortunately, policies are formed and projects developed and implemented on the basis of these statistics. Therefore, errors of measurement could be having profound implications on the strategic priorities and policies of a country.

This paper explains the improbable results observed by using data from Uganda, where the labor module contains variation both within and across surveys, to show the sensitivity of employment outcomes to survey methodology. By exploiting this variation, we show that estimates of employment outcomes are unreliable if the questionnaire did not use screening questions, as labor force participation will be underestimated. Likewise, surveys which use a seven day recall period also underestimate or potentially misrepresent employment outcomes, owing to seasonality and multiple jobs. Common multivariate analysis applied on household survey data will be affected, as the errors in measurement in the dependent and independent variables will be correlated. Corrections to reduce measurement bias in existing data are tested with the survey data; none are found to be completely satisfactory. The paper concludes that we may actually know very little about employment outcomes in Sub-Saharan Africa, and that this knowledge gap will continue unless collection techniques improve.

The husband and wife share the same bed, but they have different dreams. (Chinese proverb)

1. Introduction

Employment and earnings statistics are among the most important variables used to identify the causes of poverty as well as the pathways out - for households and for an economy. Therefore, employment outcomes are especially important to monitor consistently because they are the key link between the size and structure of economic growth and the welfare of households - the ultimate goal of development policy. The quality of employment outcomes has become a key issue for governments around the world. Gallup World Poll recently reported that jobs and economic security are the number one issue for people all over the world, replacing other concerns such as war, crime, and security (Clifton, 2012).

Policy makers in Sub- Saharan African countries, facing a large youth bulge, are increasingly focusing their national development strategies on creating more productive employment. Yet data on the nature and intensity of economic activities undertaken by the labor force in low income Sub-Saharan-African countries, and the returns to these activities are hard to find. For example, the 2011 edition of the World Bank's *African Development Indicators* (ADI) reports data on the structure of employment by sector - one of the most basic statistics - for only 14 out of the 48 countries included. In the same volume, data on labor force participation rate (LFPR) is available for all 48 countries, but the values vary improbably between countries. For example, Sudan, a very poor country where most adults cannot afford to be out of the labor force, is reported to have a labor force participation rate of 52 percent in 2009, while a rate of 85 percent for Ethiopia, Sudan's neighbor and a rate of 81 percent is reported for Angola, a relatively richer country.[1]

Even for countries where micro data from household surveys are widely available, the methods of questionnaire design and implementation generate indicators which fail to capture what the survey intended to measure (non-sampling errors), and are not comparable over time. Guarcello et al (2010) found this in the case of child labor statistics in low and middle income countries. They noted substantial differences between indicators of child LFPR across similar countries, and even within the same country over a short period of time. They contrasted the variance in the measurement of child LFPR with measures of school attendance, and found that the latter did not show such high variance depending on type of survey. They attributed the unusual variance in point estimates of child LFPR to differences in the framing of questions and the methods by which the employment data are collected. These differences also plague measures of adult employment outcomes. As a result, even trends within country are hard to identify, much less trends among countries.

One reason for the poor quality of data is that until recently, employment outcomes were not a primary focus of development policy, and therefore of the household surveys. The indicators most widely tracked at the household level were social services access and outcomes, and household assets, consumption and poverty. The former were important as public social service delivery was where the largest share of donor

[1] Many of the LFPR numbers commonly reported are not data points collected recently, but projections based on old surveys. Although they are available for most countries for most years, they may not be based on recent data.

funding went in the post-HIPC period. Human welfare outcomes such as school enrollment and infant mortality were regularly scrutinized as part of MDG monitoring, while employment outcomes were not. Indicators of poverty reduction (as measured by household consumption) were given attention as well, and still remain the ultimate goal of both SSA country development strategies and donor finance (see a recent story on IDA results, 2000-2010 for an example.[2])

A second reason for the poor quality of data is that it is not easy to collect data on employment and earnings outcomes in low income Sub-Saharan African countries (hereafter, SSA). Survey research on the measurement of employment outcomes tends to take place in OECD countries, where most of the experts on the measurement of employment and earnings work. But the employment pattern in SSA is not at all similar to that found in OECD countries. In OECD countries, almost all employment takes the form of wage and salary employment, where employed labor force participants work for an unrelated individual according to an agreement covering the terms of the task and the remuneration. Job search involves entering the labor market and trying to find someone to hire you on acceptable terms. In SSA, employment takes the form of self and/or household employment, where a task is performed for family profit or gain (including agriculture for home food consumption). Most labor force participants never even enter the labor market. Yet SSA statistics offices routinely adopt surveys instruments or modules for collecting data on employment supplied by statistical agencies in OECD countries. They may modify the surveys to fit local context, but without the benefit of local survey research or knowledge on the best techniques to use when measuring household or self employment. This results in employment and earnings data that lack relevance to the country setting, and mismeasure the key features of employment (non-sampling error). Not surprisingly, in their detailed analysis of surveys from 23 countries, Guarcello et al (2010) found that the most consistently measured employment variable was wage employment.

The purpose of this paper is to *highlight the sensitivity of employment outcomes to questionnaire methodology and show the analytical difficulties in trying to use survey data from SSA countries (where questionnaire methodologies are inconsistently applied and non-sampling error appears to be large) to understand how employment and earnings patterns have changed over time, let alone forming of policies based on such analysis*. While it uses Ugandan data as an example[3], the objective of the paper is not analyzing recent trends in employment outcomes in Uganda or any other country. Rather, this paper is about why employment data from SSA countries are unreliable and what issues analysts face in using these data to try to identify employment outcomes and relate them to government policy or household behavior, for the purposes of diagnosing past problems or establishing a baseline to track progress.

The paper mainly relies on Uganda data simply because the labor module in its surveys have significantly evolved over time and contains variations both within and across surveys that can be exploited for the purpose of our analysis. By exploiting these variations, the paper shows how difficult it can be to

[2] http://www.worldbank.org/ida/results.html, accessed on August 15, 2012. Shining a spotlight on an indicator maybe necessary but clearly not sufficient to improve quality. Despite the attention paid to poverty reduction in the past twenty years, it could be argued that the collection of the consumption data for needed for calculating absolute poverty is plagued by similarly high non-sampling errors. See for example, Jerven (2012).
[3] This paper is not on how to interpret Uganda's employment trends over time; interested readers on this topic may wish to consult Fox and Pimhidzai, (2011).

obtain key labor market indicators and identify their trend over time using household survey data in SSA. We show that it is possible that different users of the same data set may pick up questions from different parts to construct their key variables, possibly coming up with different results from the same data on both the basic employment outcomes even if they are only doing simple tabulations. By implication, these results also apply to country comparisons when different countries use different instruments. The analysis suggests that we may actually know very little about employment outcomes in SSA, and how their relationship to economic policy or household welfare.

We start the paper by reviewing the definitions of the main employment outcomes, using the framework established by the UN SNA for defining employment and the conventions established by labor statisticians through the International Labor Organization (the ILO) for the measurement of these concepts. We discuss the issues to be faced when applying these definitions in a low income setting, and review the survey research that is available to guide the measurement of these concepts. The section summarizes recommendations from past literature on questions and techniques most appropriate for measuring employment outcomes in a low income setting. In section 3 we present the questionnaires for the Uganda National Household Survey (UNHS) used by UBOS from 1992-2010, which contain both good and not-so-good practice techniques, and estimate the effect that different questionnaire designs have had on non-sampling error in the measurement of employment outcomes. We conclude the paper with suggestions for SSA statistical offices on collecting employment data, for researchers on possible approaches to adjust for known measurement biases in data collected in other SSA countries, and for future survey research on measuring employment outcomes in low income countries.

2. Some basics on collecting data on employment and labor force participation[4]

The concepts used by social scientists in the analysis of labor force status and employment outcomes were initially defined in microeconomic and specifically labor economic theory and practice. The main variables commonly tracked are:

- labor force participation rate (LFPR, which is the economically active population[5] divided by the working age population);
- inactivity (and reasons for lack of participation in the case of working age people);
- primary and secondary employment characteristics (e.g. job type – that is wage or non-wage, sector, occupation, etc.);
- unemployment and underemployment;[6] and

[4] This section draws from Beegle, Bardasi, Dillon and Serneels, (2010), and Schaffner, (2000).
[5] The economically active population is thefraction of the population which is either employed or seeking employment. See http://www.businessdictionary.com/definition/economically-active-population.html.
[6] Unemployment is recognized to be of limited use as an outcome variable in low income SSA, but it is still commonly tracked. See Fox and Gaal, (2008), for a discussion. The concept of underemployment - defined as unmet demand for remunerative work – has been growing in popularity as an outcome variable. It can be measured by low hours worked and job search. See ILO, (1982) and subsequent resolutions on this topic. As ILO (2012) notes, problems remain with the international comparability of statistics on underemployment.

- earnings from employment (standardized by a unit of time, such as hourly or monthly earnings).

While the conceptual meaning of these variables is pretty simple, development of standard definitions and measurement techniques has been the subject of substantial discussion and survey research in the post World War II period.). Because statistics are a system, the definitions of employment outcomes necessary connect with other parts of the national social and economic data systems, including the System of National Accounts (SNA). *Therefore, the SNA provides the conceptual framework for the measurement of time spent in employment (economic activity)*. The objective of collecting data on employment should include measurement all labor effort spent on the production measured in the SNA. With this objective in mind, discussions, mostly led by the ILO in their annual meetings of statisticians, have established specific, precise, and lengthy technical specifications for these terms (and associated modifiers) which are accepted by almost all national statistics offices.[7] Some of these definitions are even included in legal documents (e.g. the definition of self employment vs. employment by an employer for the purposes of tax policy or the eligibility for social security benefits).

Translating these widely accepted specifications into survey questions which are both easily administered, and understood by respondents in different cultural, social, and economic settings, is not easy. Commonly used words are rarely adequate. For example, what is "work" or a "job"? For a wage earner in an OECD country - someone who leaves home 5 days a week and goes to a factory or office for a usual number of hours each day or week and gets a regular income - the answer may be clear. A simple question in a survey such as "what is your main job?" will normally provoke from the respondent an answer regarding the *economic* activity where she has spent the most time in the last week, month or year *and* where the most labor income has been earned. The vast majority of the labor force in industrial countries does not change economic activities with the season, nor do they work in several sectors or take on multiple occupations during the year.

But the situation is much more complicated in low income countries, where the most activities take place around or near the home, within a household economic system. Depending on the country, 30-50 percent of the economically active population reports undertaking two or more substantially different economic activities during the year (for example, agriculture during harvest time and running a kiosk the rest of the time), as well as doing household chores. The definition of employment is doing any *economic activity* even for 1 hour. In order to answer the question "what is your main job?" respondents would have to know which of their daily, weekly, or monthly activities are classified as economic ones (and therefore could be called "work" or a "job"). To give some examples:

- Is running errands in an unpaid apprenticeship "work"?
- Is tending livestock or small farm animals '"work"
- Is watering vegetables in a family garden "work"?

[7] See ILO (1993) "Resolution Concerning the International Classification of Status in Employment (ICSE), Adopted by the Fifteenth International Conference of Labour Statisticians".

- Is collecting (harvesting) vegetables from the garden for dinner "work"? What about washing them before cooking them for the family meal?
- Is fetching water from the communal water source for cooking and drinking "work"?
- Is helping out after school in the kiosk run by a parent "work"?
- Is making home brew beer and selling it at the market "work"?

According to the definitions of economic activities in the SNA, all the above activities are employment except for cleaning the dirt off the vegetables before serving them to the family (but if the cleaning took place outside the household, by a vendor, for example, prior to sale to the household it would be work). This is because all of the above activities *except* for preparing the vegetables for consumption are included as economic activities in the SNA. The SNA conventions exclude certain household activities (such as preparing food for the meal, cleaning the house, or caring for a household member) so these are also not counted as employment. [8]

It is obvious that most survey respondents or even survey enumerators could not possibly be aware of these subtle details. Therefore, the simple questions "do you work?" or "what is your main job?" are not likely to illicit responses which would be either consistent across respondents. Nor would the expected responses be precise enough for analyzing and comparing outcomes within the whole national statistical system. Indeed, data collected using such vague questions would be open to a variety of interpretations. Each respondent could be expected to have a "different dream" about their economic life, unknown to the interviewer and the researcher.

Survey research has been conducted to develop the best techniques and questions to use for non-wage activity in order to obtain consistent answers which would measure these economic concepts as closely as possible to the technical specifications based on the SNA and the analyst's intent. Particular attention has been paid over the last 40 years to differences in perception and behavior across genders, age and ethnic groups, etc. which would generate inconsistent responses to a question. The difference in perceptions about the economic content of home-based work done by either gender is a clear example of where different and changing perceptions can influence responses.

Recent survey research in Tanzania, conducted by Beegle, Bardasi, Dillion and Serneels, (2010), confirms that the popular understanding of the word "work" in a low income setting is quite different from the technical specifications of labor statisticians. In the Tanzania study, the authors found that simply asking "did you do any work in the last 7 days, even for one hour?" on average resulted in a 10% decline in self-reported female LFPR, compared with a detailed module which provided an extensive list of activities which are considered work. When husbands were asked the simple question regarding their wives, an even greater undercounting occurred and vice versa for wives about their husbands (see Table 4 in Beegle et al, 2010). In the absence of detailed questions, some respondents of both genders were likely to misclassify their spouses

[8] The question of which household activities are "economic" when done by a member of the household has been debated for some time. The current definition excludes activities which are primarily caring for other members of the household. See United Nations, (2009).

sector of primary activity as well. Langsten and Salem, (2008) found similar results for Egypt. This quote from a participant in a focus group interviewed for a World Bank study illustrates the difference between what a labor statistician means by "work" or "job" and what the man in the village considers "work". When asked if they would like their lives to be like their parents, a young male, the son of a farmer, responded "Not me, my father *has never worked*. He has been here on this farm, which he inherited, his whole life". [9] Most likely this father's long "non-working" life included a lot of hard, physical, economic activity.

Another common problem in low income settings is the mixing of inactivity and activity in the same question, as mutually exclusive alternatives, as well as the use of the ambiguous term "domestic duties" or "homemaker" as a reason for inactivity. In OECD countries, most students do not work, so these two activities are mutually exclusive. This may not be the case in a low income setting however. Likewise, in OECD countries "domestic duties" involving mostly those household chores such as caring for others that are not classified as economic activities if done by a member of the household. But in a low income setting, many activities performed around the house are classified as economic because: (ii) the household is a production unit as well as a consumption unit, so household members in the same day or week often undertake economic activities as contributing family workers as well as household chores not classified as economic activities, and (ii) a large share of household chores in low income countries are classified as work because they are economic services such as fetching water and wood because these are not accessible within the household (e.g. the household has no access to piped water).

Research shows that the large female presence among economically active contributing family workers may lead to an undercounting of female LFPR if the questions are not worded carefully (Dixon-Mueller and Anker, 1988). This is particularly a big problem for women and youth, as their tasks are more likely to be ignored by a respondent yet they are classified as employment according to the ILO-SNA definitions. Beegle, et. al. (2010) found that if economic activity and inactivity were included in the same question, as mutually exclusive alternatives, women are more likely to report not being economically active because they are engaged mainly in domestic activities. Yet after screening questions, many of these same women did report being engaged in economic activities. In the simple formulation, asking them if their main activity was work or domestic duties, they chose the vague phrase "domestic duties" because that described their life to them, even it did not describe the types of activities the survey was meant to measure.

Fox and Gaal, (2008) used Tanzania Household Budget Survey data on economic activities of female heads of households to show that (i) when screening questions were not used and (ii) "homemaker" and other inactivity variables were offered as responses *together* with activity variables in one activity question (to save space), the share of household heads in the working age population reporting that they were out of the labor force doubled. In urban areas, it nearly tripled as women heads of households chose homemaker as a status in much larger numbers. With increased incomes in urban areas, it is possible that some female heads of households withdrew from the labor force, but not this magnitude, given that they have to support children.

[9] From the dataset collected for World Bank, (2011); emphasis added.

The data from the Tanzanian Labor Force survey, taken in the same year, used screening questions and confirmed that the actual LFPR of female heads was indeed much higher than was reported in the HBS.

Screening questions for economic activity and unemployment are therefore important to specify clearly to the respondent and the enumerator what constitutes an economic activity or inactivity, and to ensure that all economic activities undertaken by a respondent are captured properly in the survey. Indeed, it is quite surprising that questionnaires are still developed and fielded in low income countries which do not use this approach. For the analyst, the use of data collected with poorly worded questionnaires will produce errors and biases in the results estimated - both in the dependent and independent variables, with errors likely to be correlated. Equally important, if the perception of the meaning of the word "work" changes over time, for example, as tasks performed by women are increasingly recognized as having an economic value, then the errors and biases would not be stable over time, and trends would be misidentified. The same problem can occur when the wording of an employment module changes over time, possibly reflecting changing perceptions.

An additional challenge in low income countries is measuring non-wage earnings. In this area, there is less completed survey research. One key recommendation that does emerge from experience is that this type of income needs its own module (Glewwe and Grosh, 2000). This implies separate farm income and non-farm enterprise income modules to be completed at the household level. The measurement of farm income is a particular challenge (Carletto et al, 2011). Research is ongoing under the LSMS-ISA project to test alternative approaches and offer recommendations for survey design and implementation[10]

In sum, although the task of getting consistent measurement of the SNA concept of employment in low income countries is not easy, techniques exist. Table 1 summarizes the established good practice recommendations from literature covering key issues to be considered in designing a questionnaire to measure employment outcomes in a low income setting. However, each country context is unique. Therefore a universal recommendation is that questions should be carefully tested in the variety of contexts within the country (Grosh and Glewwe, 2000). This means not just a "field test" of a questionnaire just before fielding a full survey, but actual survey research, testing alternatives with a rigorous methodology.

Table 1: What should be in labor force modules used in low income countries?

Issue	Recommendation
(1) What is work and who is in the labor force? Questionnaires need clear definitions of work (economic activity) that can be understood by all survey participants and users to get consistent and complete measures of LFP and number of economic activities.	Use screening questions at the start of a questionnaire or module. Include separate questions on non-wage work which may take place at home or away from home. Only after a rigorous screening for potential economic activities should inactivity be established – see row 5 below (Beegle et al, 2010). Screening questions should establish whether the individual undertook multiple economic activities.

[10] see: http://web.worldbank.org/ wbsite/external/extdec/extresearch/extlsms

Issue	Recommendation
	All activities identified in the screening questions should be followed up subsequent questions/modules to ensure that primary and secondary activities are captured.
(2) How long should the recall period be? The OECD and ILO standard 7 day recall will miss economic activities when seasonality is important.	Use 12 month recall period in screening questions and in the more detailed parts of the questionnaire to capture seasonality in both wage and non-wage economic activity, especially for agricultural activities (Schaffner, 2000). When a 12 month recall period is used for activity screening, a complete picture of total economic activity in the household emerges.
(3) When should a 7-day recall period be used?	A 7-day recall will not report accurately the number of labor force participants active in a particular activity such as seasonal agricultural wage work or part time nonfarm enterprise activity over the last year. However, a 7-day recall period is used in ILO international statistics for measuring wage employment, and for the calculation of unemployment (ILO, 1993). Thus surveys should also ask which of the 12 month activities were done within the last 7-days to collect data comparable internationally.
(4) Who is unemployed?	For those who did not do any of their 12 month activities within the last week, a series of screening questions are needed to determine if someone is actually unemployed. Screening according to the standard ILO definition (did not work even one hour during the last 7 days; available for work; and actively sought work through specific steps) is recommended even though it generally produces a low unemployment rate in low income countries. *Self-definitions are not* advisable, as the word "unemployment" is usually not consistently understood by the respondent and responses are therefore idiosyncratic and inconsistent.
(5) Who is inactive/ out of the labor force and why? When women do all their economic tasks at home, respondents might think they are inactive and miss their labor force participation.	Introducing types of inactivity (such as household chores) into the interview at the same time as types of activity can cause confusion for the respondent, especially if only one choice is permitted. Establish inactivity and reasons for inactivity *only after a rigorous screening for potential economic activities* to avoid missing part-time or non-monetary activities. (Beegle et al, 2010)
(6) When should questions about economic activity asked at the household level and when at the individual level? Measuring labor input and earnings in different types of household economic activities requires different methods. This makes measuring earnings and income more complicated in low income countries.	The type of module depends on whether the activity is remunerated according to *individual performance* or the performance of *the household*. Questions about hours worked in activities can be asked of individuals in the employment module. If an activity is only part time during the year, the recall period for hours worked needs to recognize this – a 7 day recall on hours would not be enough. Earnings, benefits in kind, etc. on activities where *remuneration for labor expended is given to an individual* (e.g. wage work) can be included in the individual labor module (after the screening questions). For activities which are either (i) entrepreneurial and/or (ii) communal to the household, (e.g. farming), separate modules are preferable. Usually this means including: (i) a specifically designed module to collect data on household agricultural activities. Data would be collected on land and other inputs, including household labor inputs, and on the value of production sold or retained at the household; and (ii) a specific non-farm household enterprise model. Both modules could be used to measure the inputs (hours) of all household

Issue	Recommendation
	members rather than ask these questions individually, but asking both individual questions on hours and enterprise/plot level questionsof labor inputs is preferable to make sure that all activities get covered for all individuals. See Grosh and Glewwe (2000) for examples of each approach.
Source: Authors	

These techniques are often not used. One excuse sometimes offered for not using these good practice techniques is that they make the questionnaire "too long". It is certainly true that trying to collect too much data within a multi-purpose household survey can result in a long and tedious questionnaire, and a very expensive survey – which then will become imprecise exactly because of interviewer and respondent fatigue (Grosh and Glewwe, 2000). Certainly in the case of a census, for example, to keep costs down a short questionnaire is necessary for the majority of the respondents. Some compromises are inevitably made. The problem is that the tradeoffs being made with respect to the quality and consistency of labor force and employment data may not be considered at the time because the good practice approaches are often not even tested against the "shortcut". Only after the survey is complete does it become clear that the shortcut actually cost more because the data have too much non-sampling error to be reliable, or even usable. The result of shortcuts can be poor quality data on employment outcomes, leading to ill-informed policy choices.[11] In the remainder of the paper, we demonstrate how serious these non-sampling error biases can be, using the case of Uganda.

3 Trying to track employment outcomes in SSA: lessons from Uganda

3.1 Evolution of the UNHS survey and the labor force/employment module

Uganda's National Household Surveys, (UNHS), (formerly the Integrated Household Survey, or IHS) has been conducted by the Uganda Bureau of Statistics (UBOS) at regular intervals since 1988. The UNHS are multipurpose household surveys, originally developed as a tool for poverty monitoring and analysis. They have been conducted on a random sample of households over a 12 month period since 1992. Core modules include a household demographic module, health and education modules, questions measuring household living conditions assets and access to infrastructure, and a standard LSMS-type (recall) consumption module.[12] The analysis in this paper focuses on the 1992/93 surveys onwards. This series of household surveys is noteworthy for the consistency of the overall design and quality of execution in the field. UBOS

[11] Although failure to properly screen and measure economic activity is most common in censuses, it is also found in SSA in CWIQ and DHS surveys, and in some cases in multipurpose integrated household surveys. Both the CWIQ and the DHS surveys were not designed to measure individual or household economic activity, but in both surveys employment status is often used by analysts as an the key independent variables to explain the demographic, health, and service delivery outcomes for which the surveys were designed.

[12] The guidelines for LSMS surveys can be found in Grosh and Glewwe, and online at www.web.worldbank.org (go to LSMS website).

was one of the first offices in SSA to regularly make clean, anonymized unit record data available to researchers[13], despite numerous obstacles such as funding gaps and the regional conflict in the early 1990s.

The consistency of the overall design of the household surveys did not extend to the labor force and employment modules. These modules have been altered in each survey, displaying a range of good and not so good practices. Some changes in the employment module were made to shorten the questionnaire, making room for a topic given a higher priority that year, while more recent changes were made to improve the measurement of labor force participation and employment outcomes (e.g. more extensive screening questions) as more attention was focused on better capturing labor outcome indicators in the country. Consequently the information content differs across surveys in terms of scope and content. This has meant that the methodology used to compute variables on the measurement of labor market outcomes - levels and trends –changed over time as well, posing challenges for analysis.

Table 2 below summarizes how the five Uganda UNHS surveys since 1992 have measured (i) labor force participation rate, (ii) unemployment, (iii) characteristics of *all* economic activities undertaken, including earnings. It shows where key measurement recommendations for these indicators in low income countries discussed in the previous section - such as screening questions and long recall periods to ensure that all economic activities are captured, careful measurement of unemployment, and use of both individual and household level modules to collect complete information on hours worked and income of the venture (farm or non-farm) - were and were not used.

Table 2: Evolution of the labor force module in household surveys, 1992/93 - 2009/10

Question type	SURVEY									
	IHS 1992/93		UNHS 1999/00[14]		UNHS 2002/03		UNHS 2005/06		ÙNHS 2009/10	
	Yes /no	Type	Yes/ no	Type	Yes/ no	Type	Yes/ no	Type	Yes/ no	Type
Employment screening questions? Inactivity measured separately?	No	*Activity status responses also include inactivity variables*	No	Labor module is restricted to 3 questions on activity status *Activity status responses also include inactivity variables*	Partial	"During the last 7 days did [NAME] engage in a family enterprise or in any kind of work for pay or for profit?]	Yes – 7 day recall	Detailed – separate questions for wage, family farming, HE *12 month recall has no separate screening; activity status responses also include inactivity variables*	Yes – 7 day, then 12 month recall	Detailed – separate questions for wage, family farming, HE (including commercial agric)
Measurement of multiple employment	Yes – opti	Main, Secondary Third	Yes	Usual Main, Secondary, Current Main	Yes	Last 7 days and last 12	Yes-	Main & secondary last 7 days, 12	Yes	Main – last 7 days and last 12 months;

[13] UNHS questionnaires, documentation, and manuals are available at www.ubos.org.
[14] Because of the civil conflict, the 1999/2000 household survey did not cover six districts in the North of the country, and therefore is not nationally representative.

activities?	ons	(last 12 months, 30 days & 7 days)				months		months (no sector info), also includes listing of all wage jobs past 12 months		Secondary – last 7days
Hours worked?	Yes	For main, 2nd & 3rd activity- Months in past 12 months, days in past 30 days & hours in past 7 days	Yes	-Included in specialized enterprise and agricultural modules	Yes	*Wage employment only*- Months worked past 12 months, Hours per day -	Yes	**Wage employment only** – months in past 12 months, usual days per month, hours/day. Hours worked in agriculture captured in separate agriculture modules Other employment -7 day hours	Yes	Detailed All jobs worked in last 7 days– usual hrs each week, hrs last week; Main job in last 12 months – months, usual weeks/month, usual hrs/week
Income from economic activities (earnings)?	Yes	Earnings from "gainful" activities only – past 12 months	Yes	Detailed agricultural module Brief HE module One question on income in kind/cash from each type of income (farm, HE, wage)	Yes	*Wage employment only* Some data collected from rural households on HE income	Yes	Last payment – cash and in kind value *wage employment only*. In HE module earnings last 30 days – but not hours	Yes	*Wage employment only* – main and secondary in last 7 days – earnings and in kind payments last months Selected information from rural households only on HE income
Domestic work hours?	Yes	Time use module	Not included	-	Not Included	-	Yes	Detailed – last 7 days	Yes	Detailed – last 7 days

Moving from left to right, Table 2 makes clear how the approach to measuring *labor force participation* changed over time. Detailed employment screening was completely absent in 1992/93. Only the simple main and secondary activities questions were included. Screening was partially done in 2002/03 with the use of one short screening question in the 7 day recall, and imprecise questions about economic participation in the 12 month recall. A full set of screening questions to establish LFPR was only introduced in 2005/06. Although all versions contained questions about main and secondary employment, the recall periods and optional responses provided have varied. For example, while 7 day recall period primary and secondary employment status questions were asked across all surveys, the 12 month recall version of the primary employment status question was different in 2002/03, while the 12 month recall secondary employment status question was absent in 2002/03 and 2009/10. Some surveys (e.g. the 12 month primary and secondary employment 12 months recall questions in UNHS 2005 and all employment status questions in

1992/92) included inactivity responses (such as unemployed, domestic duties, student, to young/old) together with activity responses. Only in the 2009/10 survey were both the 7 day and 12 month recall questions used, but this time only in the screening questions, not in all the detailed follow-ups. As a result, employment status has not been consistently measured across surveys.

The evolution of the questionnaire also shows that a different level of attention has been paid to measuring outcomes from different types of economic activities over the years. Until recently, the only employment outcome measured and tracked consistently was wage employment. Earnings and hours worked were generally missing for non-wage employment in the earlier surveys. Some surveys measured only non-wage agricultural earnings but ignored non-farm activities, while others measured household enterprise earnings but skipped farm earnings. As a result, it is difficult to find one survey where all types of incomes are consistently measured over the same recall period for the household. In addition, there are inconsistencies in the measurement of hours worked by individuals, a variable which is also needed to create a consistent individual earnings variable.

What is the effect of choosing different approaches (e.g. inclusion or exclusion of certain questions like screening, use of different recall periods and inclusion of inactivity responses) on measurement of labor outcomes? Below, we show that the effect can be substantial. The analysis exploits the variations documented above across surveys (and sometimes within surveys) to estimate the direction and size of the non-sampling error. We show that the changes in framing of questions and in the recall periods can produce different results illustrating how sensitive employment outcome variables are to the design and implementation of the questionnaires. We show that it is possible that different users of the data may pick up questions from different parts to construct the key variables, possibly coming up with different results from the same data on both the basic employment outcomes and the determinants of household welfare using very simple techniques. We discuss what these problems mean for economic analysis relating these employment outcomes to other key variables at the household, sectoral or macro level in order to analyze the performance of economic development policy and processes.

The outcomes focused on here are the measurement of i) LFPR, ii) primary employment status and ii) multiple economic activities and the related issues of underemployment and household livelihood portfolios. Although relevant for analysis, we do not discuss the issue of measuring individual and household earnings as we had trouble finding a baseline to work from.

3.2 The measurement of labor force participation

Until the 2009/2010 UNHS survey (except in the 2002/3 survey), in at least one employment status question, respondents were simply asked "what is your primary activity?" without any preparation to clarify the meaning of "activity" – that is work. (Table 2).The problem was compounded by the inclusion of *inactivity* response options alongside the activity response question – which often happens with failure to screen. We can't isolate the bias associated with the inactivity response from the bias associated with lack of screening, so we consider them together, first with the earliest survey and then with the 2005/6 survey.

In the IHS 1992/93 survey, no screening was used and all employment status questions (primary and secondary) included inactivity responses such as "domestic duties", "too young/old to work" and "student". The inclusion of inactivity response options in employment status questions encouraged some respondents to provide these non-economic activity responses as their *primary* employment status, and only list their economic activity as a secondary activity.[15] Table 3 shows the magnitude of the bias induced by this methodology, and it is large. One indication of the size of the bias is found by comparing primary and secondary activity. Many of those who reported a "non-economic" activity as their primary employment status also reported an economic activity as their secondary employment status. These respondents should be counted as active.

Table 3: Effect of including inactivity response questions in primary employment status question on measurement of economic participation

	1992/93 (all questions)			**2005/06** (12-month recall)		
	Unadjusted (from primary activity only)	Including secondary activity	**Difference**	Unadjusted (from primary activity only)	Including activity	**Difference**
LFPR	78.3	86.6	**8.3**	78.6	89.4	**10.8**
Labor Force	6,470,667	7,172,816	**702,149**	9,935,893	11,304,948	**1,369,055**

Source: Author calculation based on the IHS 1992/93, UNHS 2005/06 Note: 12 month recall questions used.

Further analysis shows that this failure to properly frame the questions about work and inactivity primarily affected women in 1992/93. Nine percent of females in a household reporting any agricultural activity at all provided the inactivity response "domestic duties", but only 0.3 percent of men provided this response (Table 4). 72 percent of females of working age in these households who provided this response (and hence would be coded as inactive), reported a secondary economic activity. Nearly half of the males reporting domestic duties did the same thing. Unfortunately, most analysts (statistical offices and researchers) tend to only look at the response for the "primary" or "main" economic activity when creating aggregate employment indicators such as LFPR or distribution of employment by sector. Even for unit record data analysis, many analysts do not look at the secondary activity at all. This means that all the respondents who were actually economically active, but considered such economic activities as secondary to other non-economic activities they perform (e.g. going to school or caring for members of the household), would be misclassified as inactive and out of the labor force. If they are not counted, the LFPR would be understated by at least 8 percentage points in 1992/93.

Table 4: Reports of non-domestic secondary activities among individuals whose main activities are domestic activity, 1992/93

Description	**Agriculture household**		
	Male	**Female**	**Total**
Inactive - Primarily activity is domestic work	0.7	9.2	5.1
Share of above who report non domestic secondary activity			

[15] To the extent that this happened, it also indicates that even the interviewers were not aware of how to measure economic activity either. If the literate, educated interviewers were confused, it is not surprising to find serious non-sampling error.

No	51.7	27.5	28.9
Yes	*48.3*	*72.5*	*71.1*
Total	100	100	100

Source: Author calculation based on the IHS 1992/93

Even if screening is done well, the inclusion of inactivity options in employment status is problematic. This is evident in the 2005/6 data, where screening was included in the 7-day economic activity question but not 12 month employment one. The 12 month question was similar to those used in 1992/3, where only employment status was asked, with the inclusion of inactivity responses. As seen in Table 3, once again a sizable share of those reporting an inactivity response in the primary employment status had an economic activity as their secondary response. Not shown in the tables is that of this 10 percent of respondents found to be economically active by using secondary activity, *80 percent actually reported a primary economic activity in 7 day question, where inactivity responses were not included in the activity question*. Again, those mismeasured mainly reported their activity type as "contributing family workers" in the 7 day recall. In addition to females working at home, the economic activity of students also tends to be undermeasured. Of the 20% of the respondents aged 14 -65 who reported being inactive full time students as their primary status in the 12 months recall period, 65% of these reported a undertaking secondary *economic* activity in the same recall period (see Table 5).

Table 5: Distribution of secondary activities among full time students, 2005/06

Secondary Activity	All full time students	Working age students[a]
Own account worker	0.5	1.5
Contributing family worker	54.2	62.0
Wage employment	0.5	1.6
Full time student	1.0	1.2
Children not at school	0.3	0.07
Other	1.3	1.3
None	42.2	32.4

Source: Author calculation based on the UNHS 2005/06; Note: respondents aged between 14 and 65 years, 12-month recall

In sum, when detailed screening questions are used, broader economic participation is effectively captured because it brings clarity to the term as opposed to relying on the respondent's own interpretation. Secondly, the use of screening reduces the possibility that inactivity responses will be bundled with activity responses in employment status questions. This unbundling also improves the measurement of economic participation.

3.3 Measurement of primary economic activities – the role of the recall period

The measurement of the primary employment status is influenced by the recall period used. The recall period can make a significant difference to the measured participation in a typical African country where the majority of the people eke their living out of rain fed agriculture while participation in self-employment activities and casual wage work is seasonal. Under these circumstances, the use of shorter recall periods potentially understates participation in specific economic activities. For example, if an interview takes place

in a month after the agriculture season, a person who primarily works on a family farm during the agriculture season could accurately report being inactive or a non-farm main activity if asked their main employment in the 7 day recall but would report the farm activity in the 12 month recall. Consequently the 7 day recall questions can underestimate overall participation (if the person reported no economic activity during the post harvest season) as well as producing a distorted picture of the categorization by activity type.

Do the differences between the answers to the 7 day and 12 month recall matter? In the UNHS 2009/10 respondents were asked if their primary activity in the past 12 months was the same as their 7 recall primary status, 7 day recall secondary activity or some other activities. As shown in Table 6 about 10 percent of the respondents reported a 12 month main employment status different from the 7 day recall main employment status. Below, we consider how in practice different recall periods actually affect the analysis of employment outcomes, at the aggregate level (e.g. structure of employment) and the individual level.

Table 6: Differences between the 12 months and 7 day recall main employment activity among the working population in Uganda, 2009/10

	Gender		
	Female	Male	Uganda
Different 12 month and 7 day recall primary activity	9.0	10.7	9.8
Same 12 month and 7 day recall activities	90.0	89.3	89.2

Source: Author calculations based on the UNHS 2009/10. Note: Based on a direct question on whether the 7 day main activity is the same as the main activity in the past 12 months for ages 15 to 65.

Our first finding is that in a 12 month survey (such as an LSMS), recall period does not affect the aggregate indicator on structure of primary employment. As Table 7 shows, the structure of employment calculated from primary employment status data shows that at an aggregated level, and when the data were collected over a 12 month period, the individual differences between activity reported during the 7 day recall and the 12 month recall wash out. *Thus, for macro analyses, using aggregates, the difference can be ignored*. Note that the UNHS is collected in a 12 month period thereby minimizing the effects of seasonality, and the country has a largely bimodal rainfall season which shortens the duration between agriculture seasons. In other country settings and with surveys of shorter duration, the result might be different. For example, in the case of a Labor Force Survey (LFS) which typically is *not* conducted over 12 months, these differences could matter even in the aggregate. For this reason, it is extremely important for LFS and other surveys not carried out over 12 months to include, and rely on, the 12 month recall data.

Table 7: Comparison of primary employment status in 2009/10: 7 day vs 12 months recall period

	UNHS 2009/10	
	7day recall	***12 month recall***
Primary Employment Status		
Private wage - non agriculture	11.8	11.7
Private wage – agriculture	5.7	5.8
Government wage	3.3	3.2
Own account - non agriculture	13.1	13.1
Family worker- non agriculture	2.2	2.1
Family worker - agriculture	61.7	59.2

Unclassified[c]	2.1	4.9
Total	100.0	100.0

Source: Authors calculations based on the UNHS 2009/10. Notes: a) Unadjusted rates based only on screening questions for employment and inactivity and corresponding recall period self reported main employment status b) The composition of primary employment is based on the unadjusted employment status responses for the working age bracket (15-65 years), with the exception of the inclusion of domestic servants in the private wage category. c) Refers to people who reported a economic activity in the respective recall screening questions but have missing responses in the main employment status questions.

When analyzing individual or household behavior however, *the difference between reported status by recall period matters a lot more*. The differences in primary employment by recall period are not uniform across employment statuses. Table 8 shows that differences are disproportionately larger for those who reported a 7 day recall primary status in (i) agricultural wage activities and (ii) non-wage (e.g. self employed) non-agricultural activities. This means that in analyses such as the decomposition of changes in poverty by primary activity of the head of household (Ravallion and Huppi, 1991), or in the very common consumption regression analyses, if the 7 day recall variable is used, measurement errors are introduced which will result in biased or inconsistent estimates on the role of household economic activities in poverty status.

Table 8: Comparison of primary employment status in 2009/10: 7 day vs 12 months recall period, UNHS, individual

	Same primary activity in 12 month recall?	
Primary activity - last 7 days	***No***	***Yes***
Private wage - non agriculture	8.6	91.36
Private wage – agriculture	10.3	89.71
Government wage	4.5	95.54
Own account - non agriculture	8.1	91.06
Own account – agriculture	5.8	94.23
Family worker- non agriculture	8.3	91.72
Family worker – agriculture	5.3	94.68
Total	100.0	

Source: Authors calculations from the UNHS 2009/10. Note: Table based on a direct question asking the main 12 month activity which has 3 responses, i) the same as the 7 day main activity, 2) 7 day secondary activity and 3) other activity.

3.4 The recall period in the measurement of multiple economic activities

The measurement of multiple activities of individuals and households is influenced by the recall period. Evidence from the various surveys shows that the difference between the short (7-day) and long (12 month) recall period in the number of economic activities measured is significant, with the shorter recall period consistently understating reports of secondary economic activities (see Table 9). A comparison of the share of the working age population reporting secondary activities by recall period a lower proportion captured with the 12 month period in all surveys, with the gap widest in 2005/6, where the proportion reporting a secondary activity in the 7 day recall question was *19 percentage points* lower than the 12 month recall in 2005/06. This is despite the longer recall question being improperly framed in 2005/6 so the responses are not fully comparable (response options included inactivity ones). A similar result was found in the 2009/10 UNPS survey, when the questions were worded exactly the same. It is possible that the extent of secondary activities varied over time in Uganda, with a peak in 2005/6, but we can't really be sure.

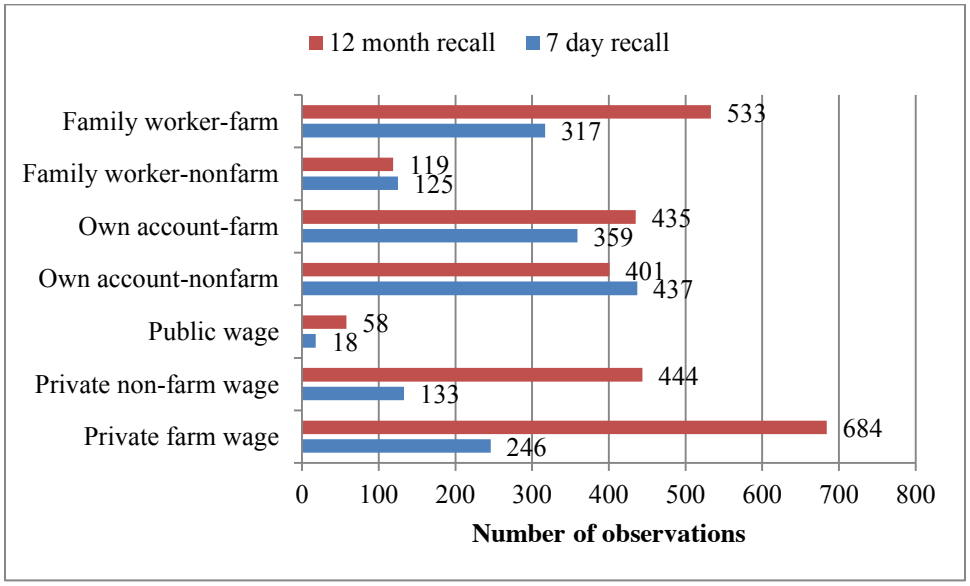
Table 9: Reports of secondary activities by recall period, 1992/92 – 2009/10[a]

	Survey IHS 1992/93	UNHS 2005/06	UNPS 2009/10[16]
Short Recall (7day recall)	25.3[b]	17.4	21.0
Long recall (12 month)	30.7	36.7	33.7
Difference	5.4***	19.3***	12.7***

Source: Authors calculations from the IHS 1992/93, UNHS 2005/06, UNPS 2009/10. Note: a) ages 15 – 65, unweighted b) Based on 30 day recall period, c) *** indicates significant at 0.1% level of significance (LOS).

The symmetry of the questions across recall period in the Uganda National Household Panel survey (UNPS), 2009/10 allows a comparison of which secondary jobs are missed in the 7 day recall. Missing jobs were mainly wage work (see Figure 1), both farm and non-farm. In both cases, the missing activity in the 7 day recall could be daily labor, meaning that both demand and supply can vary by week, month or season. Not surprisingly, family farm work (reported either as own account or contributing family member) was also under reported in the 7-day recall.

Figure 1: Comparison of reported secondary jobs by activity type: 7 day versus 12 month recall, 2009/10



Source: Authors calculations based on the UNPS 2009/10.

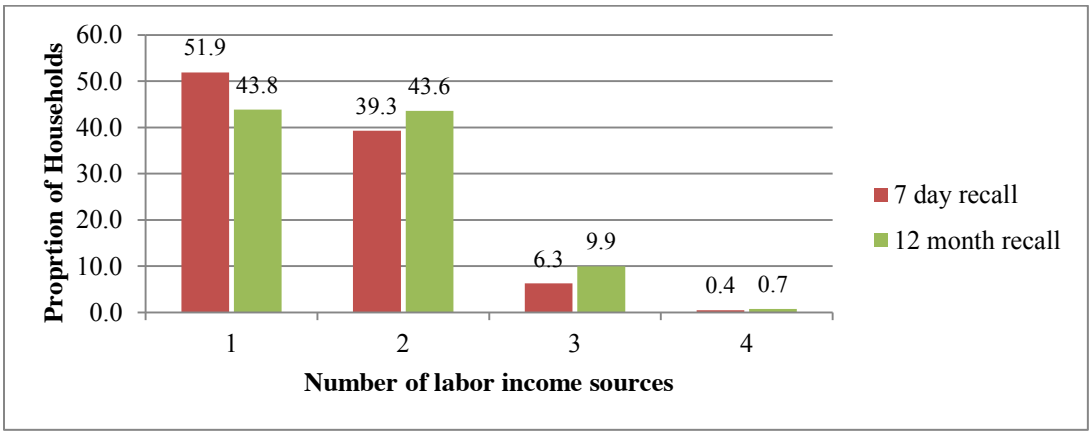
Using only short recall affects other indicators and analyses as well. One common indicator prepared as an alternative to the ILO defined unemployment measure is *underemployment*. This concept is used

[16]The Uganda National Panel Survey (UNPS) 2009/10 is a follow up of a randomly selected sub-sample of the UNHS 2005/06. It is the latest survey with available data that captures multiple activities in different recall periods. However, the national representativeness of this data is compromised by high attrition, which could be non-random from the perspective of employments characteristics of those who attritted. While the structure of primary employment [weighted with attrition adjusted weights] from the data is similar to that in the UNHS 2009/10, the paper uses unweighted statistics from the survey and only provides a sample interpretation of the statistics to analyze how respondents in the sample varied their responses to questions with different reference periods.

because few people in a low income country can afford to be completely without work, although they may not have a steady wage and salary job that "employment" in an OECD country usually connotes. They may only be able to work part of the year owing to lack of rainfall or water management, or lack of demand for the goods and services that they are selling in their primary economic activity. Often the proportion of time in which the individual actually worked in an economic activity is used as a measure of underemployment. Failure to consider secondary activities could lead to an overestimation of the extent of underemployment. Not only would this result in the wrong diagnosis, it could also lead to the wrong policy prescription.

Failure to capture second jobs even affects aggregate statistics on livelihoods. Analysis of household livelihoods is very important when household production is the main income source for 70 percent or more of the population, and the type of activity the household is engaged in may vary over the year. Livelihood analysis measures the *portfolio* of income sources in the household. Given that the primary employment status is an incomplete representation of all the economic activities undertaken by an individual, it follows that failure to account for secondary employment would mis-measure the structure of livelihoods in the country (see Fox and Pimhidzai, 2011). Figure 2 confirms this. By missing some secondary jobs, the 7 day recall understates people's livelihood diversification because *even at an aggregate level* it fails to pick up the income diversification of about 8 percent of households who have more than one source of income, but are identified as having a single labor income source when the 7 day recall period is used.

Figure 2: Number of household labor income sources by recall period, UNPS 2009/10



Source: Authors calculations based on the UNPS 2009/10

The underreporting of both activity types and their number using a 7 day recall period means that indicators based on these data are plagued with measurement errors that would affect any micro level analysis. The measurement of earnings is one such example. Individual earnings will be understated in the 7 day recall period by virtue of missing out other income sources of the individual over the year. This measurement error is unlikely to be random. As Figure 1 shows, certain types of activities including temporary private agriculture jobs and casual non-agriculture jobs tend to be missed the most, and these are mainly performed by the poor. Furthermore, the shift in economic activities over the year by individuals and households will make it difficult to extrapolate 7 day recall earnings to annual earnings or to analyze variations within groups (Schaffner,

2000).[17] It is difficult to tell whether those households who had high incomes or engaged in a particular activity in one month would also have high incomes or would they have been engaged in similar activities in different months. Measurement errors also occur with occupational choice, since the individual level 7 day recall status can deviates from the 12 month recall status. These measurement errors will result in inconsistent estimates in regressions using earnings at both the individual and household level as well as analysis involving individual employment variables.

In sum, while the recall period has a negligible impact on aggregate indicators of primary employment, it matters a lot for indicators capturing multiple activities both at aggregate and household/individual level analysis. A short recall period biases any kind of individual and household level analysis because it understates participation in secondary activities, and by implication gives a misleading picture of livelihoods and overstates underemployment. It also makes it difficult, if not impossible, to measure annual earnings for important segments of the labor force – likely the poor. It follows then that indicators based on the 12 month recall period are the more reliable, especially for earnings and livelihood level analysis[18].

4. Suggestions on minimizing measurement error when calculating LFPR in low income countries

The Uganda analysis, as well as other analysis cited above, showed that there is significant variation in the LFPR depending on the questionnaire design. There are categories of respondents – specifically women and those in school - whose labor force participation is systematically understated. The key question is how to construct a meaningful indicator for LFP in a given country, in the presence of non-sampling measurement error traced to the type of questionnaire or the implementation. The first step is to recognize that the bias of the error is to underestimate the labor force. But then, what can the analyst do after the data have been collected?

Guarcello et al (2010) offer one option. Using data from 24 countries, they constructed a model which would predict LFPR for an individual or group based on characteristics observed consistently in all surveys for that country, as well as dummies for type of questionnaire. They used this model to estimate the LFPR for children aged 10-14 in each country. While their predicted ("harmonized") LFPR based on individual/household characteristics showed a more consistent pattern than the one derived from the different types of surveys, they noted that a large share of the variation in estimates across surveys is explained by *the type of survey*, not individual characteristics. They therefore found these results less than satisfactory, since type of survey (CWIQ, DHS, LSMS) is only a weak proxy for the extent of non-sampling error.

[17] The measurement of earnings encounters many difficulties beyond; recall period is only one of these.
[18] See annex 1 of World Bank, 2012 which among other things, discusses the analytical implications of the absence of a 12 month recall secondary activity question in the UNHS 2009/10.

Fox and Gaal (2008) proposed another alternative, arguing that knowledge of the biases by module design (e.g. exclusion of screening questions for economic activity) can be used to make an adjustment in the data and bring better (but not complete) comparability. This method involves recognizing that a satisfactory measure of economic participation should be able to capture participation in any form of economic activity, irrespective of whether the respondent deems such an activity primary or secondary. In the absence of screening questions that capture participation in all economic activities over a 12 month period for example, information from various questions in the labor module (primary and secondary employment status) and other modules where economic activities are separately captured (e.g. a household enterprise module, agriculture module and time use or non-labor market activities module), can be used to get a complete picture of LFP. This search of the questionnaire and the data should also include looking at time use data to pick up some domestic duties technically classified as work such as fetching wood and water.[19] This method was applied in Table 10 to estimate the extent of the non-sampling bias in the Uganda data. The steps are as follows.

a. *Use secondary activity information*. In the absence of screening, users can identify those who are reportedly inactive, or are missing the status variable in the main employment status question. Then, the user can search for a secondary activity response. If they report a secondary economic activity, the person would be recoded as active, and the secondary activity would be recoded as the primary. For example in 1992/93 a woman might report domestic activity (i.e. economically inactive) in the main employment status question but reports helping without pay in a household enterprise or farm in the secondary employment status question. She should be coded as active. This is the same adjustment shown for Uganda in Table 3 which compared LFPR based on primary activity only to the LFPR including secondary activity.
b. *Use additional information from the survey to verify and recode those who still remain inactive*. Where there is a household enterprise module, identify household members working in the enterprise. If they have been coded as inactive, recode them as active, and their primary activity as a contributing family worker in the nonfarm enterprise sector.
c. Use the relationship to head information to code all domestic servants as *economically active, and as wage workers*.
d. For those who are still inactive, look for *time use or household chores modules*. Code as active those who reported economic activities in either the detailed household chores or time use modules (e.g. activities such as agriculture, or fetching water or wood).
e. Identify agriculture households (i.e. where at least one household member reports agriculture work) and code all "inactive" women who reported 'domestic duties' as reason for inactivity as *active*, in the agriculture sector. Mostly likely, they did participate, but were doing unpaid and unrecognized work, and this should be captured.
f. For those still coded inactive or missing use the education, health, and age information to verify reasons for inactivity.

[19] Many surveys check for participation in household chores unpaid but considered "economic" (such as fetching water) in a separate section on household chores, not in the screening section. This approach works as long as the analyst is careful to include this information when constructing indicators of LFP.

Where a survey does not contain this extra information to rely on, we suggest using a rule of thumb: *code as active people who report doing domestic work but live in agricultural households*. Table 10 below uses the IHS 1992/93 to show that while not a perfect adjustment, this rule of thumb on aggregate considerably adjusts for underreporting for participation when compared to a finer adjustment based on the use of information from other sections as stated above. Nonetheless, the rule of thumb is not a panacea, as it overstates men's participation when compared to a detailed adjustment, while restricting it to agriculture household still excludes those in urban areas who carry out household chores that should be counted as economic activities.

Table10: Comparison of unadjusted and adjusted labor force participation rates by gender, 1992/93

	Gender		
Labor force participation rate[a]	**Male**	**Female**	**Uganda**
Unadjusted LFPR	81.8%	75.2%	78.3%
Rule of thumb adjustment	90.9%	82.1%	86.3%
Detailed adjustment (from table 3)	86.7%	90.4%	88.6%

Source: *Authors' calculations based on the IHS 1992/93.* ***Note:*** *The unadjusted LFPR is based only on the main employment status questions, rule of thumb simply codes as active people who reported domestic work but resides in agricultural households where similarly aged people of their gender report agriculture work, and detailed adjustment makes adjustments using information from other questions and modules of the questionnaire (secondary activities, and time use and household enterprise module).*

When the information content differs across surveys, a trade-off between comparability and adequately capturing the concept of economic participation is implicitly involved. As an example, the UNHS 2005/06 and UNHS 2009/10 both have screening questions that form a better base for measuring economic participation (LFPR), but indicators constructed from these will not be comparable with previous surveys. If the trend is to stretch back to 1992/93, it will be even more difficult to create comparable indicators, even if we use both primary and secondary employment status responses for all three years. The suggestions above are simply ways to approach data comparability and minimize non-sampling bias, but obviously user discretion is advised where questionnaires are simply too different.

5. Conclusions: A framework for capturing activities not dreams

While much has been made of the poor quality of national accounts or poverty data in Sub-Saharan Africa, less has been written on the quality of labor force and employment outcome data. Yet the shortcomings in employment data are at least as important, not only because they inform the national accounts data. The main problem has been the use of poorly worded questionnaires. These questionnaires – the "bed" in the title of the paper – are applied without first clarifying the reality they seek to measure. So they end up measuring what is in people's heads as they hear the words of the interviewer - their "dreams". This makes the data difficult to interpret or to use in cross-country or trend analysis.

The analysis in this paper showed the sensitivity of labor outcome estimates to type of questions, their framing, recall periods and computation methods. Using multiple Ugandan cross section household survey data sets and exploiting the variations across surveys, and sometimes within surveys, we illustrate how

sensitive employment outcome variables are to the design and implementation of a questionnaire. For example, if an analyst chooses to use a 7-day recall period employment variable for analysis, erroneous conclusions can be drawn with respect to key outcomes for specific types of individuals and households. There is also the possibility that different analysts, using variables constructed in different ways, could easily come to different conclusions on level indictors using the same data. Trends over time are even more difficult to compute and interpret. We show that measurement bias can affect economic analysis relating these employment outcomes to other key variables at the household, sectoral or macro level. By implication, these results also apply to country comparisons when different countries use different instruments

Inadequate questionnaires mean that analysis of labor force participation and employment outcomes is not a simple prospect in low income countries. In Uganda, older surveys used questionnaires which clearly led to an underestimation of LFP at the household and aggregate level. Many countries still use these formats today. This complicates even the simplest aggregate trend analyses, such as whether the share of the labor force in agriculture is going up or down. Even data from the most recent Ugandan surveys, which include formats specifically designed to measure LFPR more accurately, have to be used carefully owing to multiple economic activities. Although the 7-day recall period is an international standard, and has worked just fine for OECD countries employment statistics, using it in a low income setting will underestimate, and possibly distort, analyses of individual outcomes such as underemployment and household livelihoods.

Some have argued that the concepts of labor force participation and employment embedded within the SNA are not suited to a low income setting. For example, it has been suggested that economic activities which produce output included in the SNA (such as growing food for home use) should not be considered employment as no income is earned.[20] The problem with this argument is that it goes against decades of revisions of the SNA, all designed to include more economic activities, not less. As SSA countries such as Ghana, Nigeria, and Uganda make progress on implementing the 1993 SNA recommendations on including non-market activities such as household production, their employment data needs to catch up too. Calculating national accounts under the 1993 SNA revision, including the Social Accounting Matrix (SAM) used in checking consistency and balance, requires more extensive data on non-wage activities from household surveys. Mismeasurement at the micro level will pose problem for the macro statistics. The argument here is that it is possible to measure employment in low income countries in a manner consistent with the SNA and internationally accepted definitions if the questionnaire design takes into account the reality of how economic activities are performed in the country context. The problems occur when this does not happen. Statistical offices need to recognize this, and raise their game.

The flagship household survey of most SSA countries is their 12 month multi-purpose one (including household consumption), which is used to monitor household welfare outcomes, including poverty. Over time, as these surveys were used for policy analysis, it was recognized that employment outcome variables (including earnings) need to be included in order to discern the growth-poverty linkages and the reasons why growth was or was not shared. As a result, more surveys are adding questions on the economic activities of household members. To be useful, these surveys need to use formats that are adapted to a low income country

[20] This argument was recently made in ILO (2012), which was a draft discussion document.

context. Labor force surveys (LFS) usually do a better job of screening for economic activities, but these surveys do not take place over a 12 month period, so they run the risk of losing seasonal employment unless a 12-month recall period is used. Other surveys such as the CWIQ and the DHS usually include a minimal employment section, with no screening questions. The analysis here shows that this approach will deliver biased results. The value of including a minimal employment section in these surveys is questionable.

The measurement of employment outcomes needs as much effort in survey design and implementation as does the measurement of other outcomes. Effort has gone into researching the measurement of consumption and poverty in low income countries. The same efforts need to be applied to the problem of measuring employment outcomes, including incomes and earnings. Funding for survey research in low income countries is usually not a donor priority, but it is clearly important. The techniques discussed in Section 2 have been shown to produce better results in low income settings. But better research on alternative questions and methodologies in specific country context, including how to translate these concepts into local languages, is still needed.

References

Bardasi Elana, Beegle, Kathleen, Dillon, Andrew and Serneels, Pieter. (2010). "Do Labor Statistics Depend on How and Whom the Question was asked? – Results from a Survey Experiment in Tanzania", Policy Research Working Paper 5192, The World Bank.

Carletto, Calogero, Sara Savastano, Alberto Zezza (2011). "Fact or Artefact: The Impact of Measurement Errors on the Farm Size - Productivity Relationship." Policy Research Working Paper 5908, The World Bank.

Clifton, Jim. (2012). *The Coming Jobs War*. New York, NY: Gallup Press

Dixon-Mueller, R., and R. Anker (1988). "Assessing Women's Economic Contributions to Development. Training in Population", Human Resources and Development Planning Paper number 6, International Labour Organization, Geneva

Ehrenberg Ronald G. and Robert S. Smith (2006). *Modern Labor Economics: Theory and Practice*. 9th Ed. Prentice Hall.

Fox, Louise and Melissa S. Gaal (2008). Working out of Poverty: Job Creation and the Quality of Growth in Africa. The World Bank, Washington, DC

Fox, Louise and Pimhidzai Obert. (2011). "Is Informality Welfare-enhancing Structural Transformation? Evidence from Uganda" Policy Research Working Paper No. WPS 5866. The World Bank, Washington DC

Guarcello, L., I. Kovrova, S. Lyon, M. Mancorda, and F.C. Rosati (2010). "Towards consistency in child labor measurement: Assessing the comparability of estimates generated by different survey instruments", Understanding Children's Work Programme Working Paper series, Rome, Italy.

ILO (1982) Resolution Concerning Statistics of the Economically Active Population, Employment, Unemployment and Underemployment. Adopted by the Thirteenth International Conference of Labour Statisticians, Geneva: International Labour Organization.

ILO (1993) "Resolution Concerning the International Classification of Status in Employment (ICSE), Adopted by the Fifteenth International Conference of Labour Statisticians" http://www.ilo.org/public/english/bureau/stat/download/res/icse.pdf

ILO, (2012). "Towards a revised framework for labour force and work statistics". Document prepared for the Preparatory Regional Meeting of Labour Statisticians in support of the 19th ICLS (26-28 March 2012 –Santiago, Chile)

Jerven, Morten (2012) *Poor numbers: What do we know about income and growth in sub-Saharan Africa?* http://www.cgdev.org/doc/17-NOV-CGD-Poor%20Numbers-Jerven.pdf

M. Grosh and P. Glewwe (eds.), *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study*. Oxford: The World Bank

Langsten, Ray and Salem Rania. (2008). "Two Approaches to Measuring Women's Work in Developing Countries: A Comparison of survey data from Egypt" Population and Development Review, Vol. 34 (2) pp. 283 -305

Ravallion, M. and M. Huppi (1990). "Measuring Changes in Poverty: A Methodological Case Study of Indonesia during an Adjustment Period." World Bank Economic Review, 5, pp. 57-82.

Schaffner, Julie Anderson. 2000. "Employment," in M. Grosh and P. Glewwe (eds.), Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study, vol. 1. Oxford: The World Bank, pp. 217-250 -. 2000b.

Uganda Bureau of Statistics. 2010. *Statistical Abstract 2010*. Kampala, Uganda: UBOS

Uganda Bureau of Statistics. 2010. *Statistical Abstract 2011*. Kampala, Uganda: UBOS

United Nations 2009. *System of National Accounts*. New York, NY: United Nations.

World Bank (2006). "Uganda Poverty and Vulnerability Assessment" October 2006, Report No. 36996-UG, The World Bank, Washington DC.

——— 2009. "Sharing the Growth in Uganda: Recent Labor Market Outcomes and Poverty Reduction Summary" Draft Policy Note April 2009, The World Bank, Washington DC

——— 2011(a). "African Development Indicators". The World Bank, Washington DC

——— 2011 (b). "Defining Gender in the 21st Century: Talking with Women and Men around the World: A Multi-Country Qualitative Study of Gender and Economic Choice." Dataset. World Bank, Washington, DC.

----------2012. *Uganda - Promoting inclusive growth: transforming farms, human capital, and economic geography - synthesis report*. Washington D.C. - The World Bank.