POLICY RESEARCH WORKING PAPER 5671

Reliability of Recall in Agricultural Data

Kathleen Beegle
Calogero Carletto
Kristen Himelein

Abstract

Despite the importance of agriculture to economic development, and a vast accompanying literature on the subject, little research has been done on the quality of the underlying data. Due to survey logistics, agricultural data are usually collected by asking respondents to recall the details of events occurring during past agricultural seasons that took place a number of months prior to the interview. This gap can lead to recall bias in reported data on agricultural activities. The problem is further complicated when interviews are conducted over the course of several months, thus leading to recall of variable length. To test for such recall bias, the length of time between harvest and interview is examined for three African countries with respect to several common agricultural input and harvest measures. The analysis shows little evidence of recall bias impacting data quality. There is some indication that more salient events are less subject to recall decay. Overall, the results allay some concerns about the quality of some types of agricultural data collected through recall over lengthy periods.

Reliability of Recall in Agricultural Data

Kathleen Beegle

Calogero Carletto

Kristen Himelein

1. INTRODUCTION

For most of the population of Sub-Saharan Africa, farming is a main source of both food and household income. Analysis of the Rural Income Generating Activities (RIGA) study shows that, while non-farm activities are increasing in importance, the vast majority of rural African households remain heavily dependent on agriculture, with on-farm sources of income ranging from 59 to 78 percent of total income (Davis et al., 2010). The comparatively higher importance of agriculture for the African countries included in the RIGA study confirms the critical role that agriculture plays in the economic development and improvement of living standards in the region. This relationship has also been the focus of vast quantities of other literature, including the World Bank's 2008 World Development Report, *Agriculture for Development* (World Bank, 2008). However, only rarely is there a focus on issues of underlying data quality and data collection methods. Historically, the data on rural household farming are perceived to be of poor quality, particularly when collected outside the domain of specialized farm surveys, but recently the greater demand for household data on both farming decisions and non-farm activities has led to the expansion of general household surveys to include extensive agricultural modules to capture agricultural production.

Because of the complexity of farming, where salient actions on the farm take place over several months of a season (plot preparation, input application, plot maintenance, harvest and selling), agricultural information would ideally be collected through multiple visits over a farming season to facilitate accurate recall of events. Specialized farm surveys are often designed to visit the household at multiple times, particularly those utilizing resident enumerators (e.g. agricultural extension agents or other Ministry of Agriculture staff). Such specialized surveys, however, offer limited scope to analyze the links between agriculture and non-farm income activities, as well as other socioeconomic outcomes. To meet the demand for integrated data, multi-topic/multi-purpose household surveys (such as a Living Standards Measurement Study [LSMS] survey) are often extended to cover agricultural issues. Yet, cost and logistical considerations usually dictate that the data in these surveys be collected during a single visit to the household. In this case, the household/farmer is asked to report information relating to farming by recalling the details of past events for the last completed agricultural season, often including two or more separate harvests. Depending on the timing of the survey, the interview itself will not necessarily be the most propitious timing for a single visit (i.e. immediately after the main harvest).

Moreover, it is common that such multi-purpose surveys will be fielded over 12 months, to account for seasonality in consumption expenditures. In this case, the recall period will vary across households depending on the month during which the household is interviewed. In the case of surveys from year-long fieldwork, such as the three examined here, there is variation up to 11 months between the time of the harvest and the data collection, and recall periods for input use are several months longer.

Even if all sample households are surveyed once over a very short period just after harvest, events early in the season will be reported with a recall of several months (e.g. fertilizer application). This raises concerns about the reliability of information reported for earlier events. Substantial recall effects resulting from this gap in time could introduce bias or measurement error into data collected further from events of interest, and the problem may be further aggravated by variable recall periods across sample households.

The objective of this paper is to investigate the extent of recall bias, using data from three national household surveys conducted in East and Southern Africa with fieldwork over about 12 months. We use the variation in recall period and random assignment of households to month of interview (and, therefore, recall period) to examine several important agricultural indicators within the three datasets. Specifically, we explore whether we find any differential in reporting of input use and harvest amounts for main crops based on the time elapsed between events and reporting. In general, we find little evidence of any significant recall bias, in terms of under/over reporting. The findings suggest that farmers' reports of harvest, crop sales, and input use are not significantly different when collected more than 8 months later as opposed to just after the harvest. Although this is *not* evidence that agricultural data are not of poor quality, it does address at least one aspect of data collection – length of recall period – which could compromise data quality.

The next section briefly reviews the literature on survey reporting errors, with attention to the types of bias most relevant for agricultural household surveys. Section 3 discusses the data and empirical approach. Results are presented in Section 4. Section 5 concludes.

2. Literature Review

There is a large body of work in the survey literature about reporting errors in survey data; Sudman and Bradburn (1974) summarize the extensive work in this area from the 1950s and 1960s. Retrospective data, in particular, introduce various data quality issues linked to recall errors. Over time, respondents may not be able to accurately recall details of events. Accurate remembrance of events will depend not only on the duration of the recall period, but also on the nature of the event being reported. The literature generally categorizes three main types of recall error in household survey data: telescoping, heaping, and recall decay. Telescoping refers to inaccurately identifying the date of events, either forward or backward into the recall reference period. Heaping refers to the use of estimation by respondents that places large numbers of responses at particular points, such as an expenditure of about $100 or dating an event about 6 months ago. Recall decay refers to forgetting details of events. Since the focus of our study is farm input usage and harvest size, rather than the timing of events, we do not focus on telescoping. However, there are areas within agricultural statistics where telescoping would be relevant, such as dates of fertilizer application or input purchase.

It is also difficult to consider the issue of "heaping" because agricultural data are naturally heaped in that both inputs and products are sold in uniform quantities (e.g. 50 kg sacks, oxcarts, etc.), as opposed to variables such as reported age, where one does not expect to find heaping and can then look for it as evidence of reporting errors. Although there is considerable variation in the units reported across the crops and countries we study here, we nonetheless find that most of the quantities are reported in heaped quantities reflecting the normal patterns one would expect for these measures. For example, in the data used in this study from Kenya on maize harvest, more than 90 percent of reports are in terms of numbers of 50 or 90 kilogram bags. For coffee, more than 97 percent of reports are in kilogram amounts divisible by 10. Likewise, in Rwanda, the vast majority of maize kilogram reports are in multiples of 20 or 50, and sorghum reports are in multiples of 10 and 100. In Malawi, more than 90 percent of maize production reports are in terms of 50 kg bags, 90 kg bags, or ox carts.

Recall decay bias can occur in two ways. First, respondents may altogether forget having performed certain activities, biasing frequency estimates downward. While there is little empirical work on the subject related specifically to agriculture, the topic is covered in guidelines for marketing and social research questionnaires. This literature describes salience as an

important factor as to whether events are likely to be affected by recall decay. Bradburn et al. (2004) identify three factors that determine salience: "(1) the unusualness of the event, (2) the economic and social cost or benefits of the event, and (3) the continuing consequences of the event" (p. 64). Loftus and Marburger (1983) assess the improvement in accuracy of retrospective accounts when surveys use a highly salient landmark event (e.g. the eruption of Mt. St. Helens) to mark the beginning of the reference period. In their study on migration, Smith and Thomas (2003) find that migration events that are of greater salience to the respondent are less likely to be affected by recall decay. Relating specifically to agriculture, Judge and Schechter (2009) test the distribution of the first significant digit of various data to detect abnormalities. Using Benford's Law, they find evidence of the importance of salience in data quality from an agricultural survey in Paraguay, specifically that more accurate information is recorded for crops that represent a larger share of household income.

Second, recall decay can affect the details of events reported by respondents. Much of the research in this area relates to consumption experiments, which focus on the reliability of reports of consumption or expenditure on food and non-food consumer goods. One exception is Coleman (1983) who addresses recall decay in agriculture through comparisons of the mean number of reported labor hours for each day during the week prior to the interview. Using a dataset of 129 households in 12 villages in Benue State, Nigeria from 1979/1980, he takes as a reference point the mean number of hours reported for the day before the survey. With each additional day from the reference point, he finds a consistent and significant over-reporting of labor hours, ranging from 17.5 percent on the day before the reference day to 61.6 percent on the first day of the recall period. His finding with regard to agricultural labor contradicts expectations based on marketing and household consumption data, which would predict increased under-reporting with the passage of time, as respondents forget events. Gibbs et al. (1986), in their study on agricultural innovation discovery in Southern Australia, find that information collected with an aided recall module was generally reliable over a recall period of at least 12 months despite the relatively low salience of the events. In addition, they find the quality of the information does not diminish substantially with increases in the length of the recall period.

Studies of recall in household consumption data assess the reliability of data by the length of time to which the data refer, such as food expenditures over the past 7 days, 30 days, or even as long as the past year. These studies suggest that longer recall periods are associated with

lower aggregate totals of frequent food expenditures. Gibson and Kim (2007) hypothesize that when respondents are asked to recall larger amounts of information, from a greater number of transactions due to larger household size or longer recall period, they will shift from summing the total of individual remembered events to estimating the overall total. In a survey experiment in Papua New Guinea, Gibson (2002) found that the average food expenditure was 26 percent higher with a consumption diary (asking for consumption for each of the 7 days covered) than asking for one reported amount for the entire 7 days. In a survey experiment in Ghana, reported expenditure on frequently purchased items fell by nearly 3 percent for each day added to the recall period, although it leveled off at about 20 percent after 2 weeks (Scott and Amenuvegbe, 1991).[1]

There are also examples of recall decay for analysis of other topics. Bound et al. (2001) provide a general review of this literature, with examples taken from labor, health care, crime and motor vehicle accident statistics. The authors summarize the overall findings as "the greater the length of the recall period, the greater the expected bias due to respondent retrieval and reporting error" (p. 3743).

Relating these findings to agricultural data and the concept of salience, we would expect that large scale, unusual or expensive events for farm households are less likely to suffer from recall decay than smaller, less-salient events. We would expect to find less recall decay in cash crops than staple crops. Cash crops are sold, generally by weight at a fixed price, while at least a portion of staple crops are kept for home consumption. Moreover, cash crops are critical sources of cash income to subsistence farmers. With respect to inputs, we would expect to find more evidence of decay in labor usage as opposed to fertilizer use, as labor can be used sporadically throughout the growing season while fertilizer purchase and application are generally singular events. We would also expect to find less evidence of recall decay in relation to less common events, such as labor use in places where outside labor is uncommon (such as Rwanda). Finally, when both are collected in a single visit, we expect to find more evidence of decay related to input rather than harvest data, as the recall period is several months longer.[2]

---

[1] There is a large literature on reporting errors in consumption, and, to a lesser extent, in income surveys, primarily from the United States and other developed countries. Reviews of some of the evidence can be found in Gibson (2006), Deaton and Grosh (2000), and Scott and Amenuvegbe (1991).

[2] In the data used here, farmers self-report production and input use. An alternative to self-reported production (which is subject to recall bias among other reporting problems) is to measure production through crop-cutting procedures conducted by field staff. Studies comparing crop-cut estimates to "whole-plot" harvests in Africa suggest that crop-cutting itself is not free from error and can result in large over-

3. Data and Empirical Approach

To investigate potential recall bias in agricultural harvest estimates, we make use of three nationally representative multi-topic household surveys from Sub-Saharan Africa, the 2004/2005 Malawi Integrated Household Survey (IHS), the 2005/2006 Kenya Integrated Household Budget Survey (KIHBS), and the 2001 Rwanda *Enquête Integrale sur les Conditions de Vie des Menages* (EICV). These surveys were chosen because of their 12-month fieldwork calendar and because interview locations were randomized across regions throughout the data collection.[3]

The 2004/2005 Malawi Integrated Household Survey (IHS) was collected over 13 months from March 2004 to March 2005, and covered all districts of Malawi excluding Likoma Island (Malawi NSO, 2005). Fieldwork was conducted concurrently in the three main agricultural regions of the country (north, central and south) to prevent regional averages from being distorted by seasonal bias. The sample was selected using a two-stage stratified sampling design based on a frame from the 1998 census, and was structured to be representative at the district level. The total number of households interviewed was 11,280. Of these households, the analysis was restricted to those households involved in rain-fed agriculture. Households reported agricultural information with respect to the most recently completed agricultural seasons (2002/03 or 2003/04).

The 2005/2006 Kenya Integrated Household Budget Survey (KIHBS) was collected over 12 months from May 2005 to April 2006, covering all eight provinces of the country concurrently (Kenya NBS, 2007). The KIHBS sample was also selected using a two-stage stratified sampling

---

estimates of production, as concluded in the review of evidence by Fermont and Benson (2011). The same study notes that farmer reports were closer to actual production and of lower variance than crop cuts.

[3] In the three surveys studied, households are clustered in enumeration areas (EAs, also referred to as primary sample units), with all households in the EA interviewed over the course of several days. Therefore, the randomization necessary for our study is at the EA level. The survey documentation does not explicitly state that the enumeration areas were randomly allocated (within strata) over time, although it is implicit given the overall survey objectives (poverty and well-being measurement over a 12-month period). Two of the authors worked directly on the Malawi survey and sample design, and are able to confirm this is the case. For all three surveys, we examined the pattern of EAs interviewed within and across geographic areas in each survey to ensure the distributions were even. We also examined whether household characteristics (like landholdings, female headship, and head's education) vary significantly with interview month and they do not. These results are available upon request.

design based on a national sample frame developed from the 1999 Population and Housing Census. The sample is designed to be representative within urban and rural stratifications at the national and provincial levels and at the level of the country's 69 districts. The total number of households interviewed was 13,212, and the questionnaire refers to the most recently completed agricultural season.

The 2001 Rwanda *Enquête Integrale sur les Conditions de Vie des Menages* (EICV) was collected over 21 months from October 1999 to June 2001. The bulk of the data collection was done in the last 12 months, and therefore this paper limits the analysis to this time period (Rwanda Statistics Department, 2002). Data collection was done concurrently in all 12 prefectures of the country. The sample was selected using a two-stage stratified sample design and the sample is designed to be representative at the prefecture level. The total sample size was 5,739 rural and urban households, interviewed from July 2000 to June 2001, and the questionnaire refers to the most recently completed agricultural season.

In the analysis, we regress information on harvest sales (in kilograms and in local currency) and input use (fertilizer and hired labor) on time elapsed between the harvest and the date of the interview. We explore systematic under/over-reporting by length of recall. We examine whether reporting changes between interviews conducted further from the harvest and those completed close to harvest.

Agriculture is the dominant source of income for the rural households in these surveys. In Kenya, 82 percent of rural households were engaged in agriculture, as opposed to 22 percent that were engaged in non-farm enterprise and 31 percent where at least one member was engaged in wage labor. Differences are even larger for rural households in Malawi and Rwanda. In Malawi, 95 percent of rural households were involved in agriculture, with only 30 percent engaged in non-farm enterprises and 16 percent in wage labor. In Rwanda, 99 percent of households were engaged in agriculture, and only 13 percent in non-farm enterprises and 30 percent in wage labor. Most of the farming households are smallholders with plot sizes generally between two and three acres; Appendix Table 1 shows means for the sub-samples studied in this paper.

Both food staple and cash crops are analyzed for evidence of recall decay. Maize, a main staple crop, is examined in all three countries. Sorghum is also included for Rwanda as it is the most

8

widely grown staple in that country. The main cash crops studied are coffee in Kenya and Rwanda and tobacco in Malawi.

In addition to harvest quantity, we also examine the value of cash crop sales (conditional on any sales). On the one hand, we expect sales events to be very salient for poor farmers as they provide cash income, thus suggesting they are subject to less recall. On the other hand, as farmers may sell to several buyers, this measure may be more prone to recall errors, especially as time passes. Although staple food crops are sold, we do not examine this outcome.[4] Farmers may sell their harvest over the course of several months, resulting in an actual increase in sales by month since harvest. Therefore, if we observe an increase in quantity sold for interviews further from harvest, it may not be over-reporting but, rather, reflect the pattern of sales. We expect that the scope of this is small in the context of Kenya, Malawi and Rwanda, since farmers have limited storage facilities and a high demand for cash as these sales are their main source of annual cash income. We explore the extent to which we see an increase in the percentage of growers reporting any sales by time since harvest, and, in fact, find a fairly flat relationship. That is, the percentage of growers who sell does not increase as interviews occur further from the harvest date. Of note is the system of tobacco sales in Malawi. Three-quarters of Malawian tobacco growers sell directly to tobacco auction floors and the remaining sell to an intermediate buyer who then sells on the auction floor. Auction floors operate from April to July each year, so we do not expect to see sales increasing over time. Nevertheless, to be cautious, we only examine cash crops where we think farmers are less likely to store and sell over several months.

With regard to input usage, labor usage varies across both crops and countries. Generally, in our final sample (described below), the use of hired labor for staple crops is most common in the Kenyan households, with about 34 percent of households; lower in Malawi at just below 20 percent; and lowest in Rwanda, where less than 5 percent engaged outside labor. The incidence of hired labor is higher for staple crops than cash crops. For hired labor, we examine only Malawi tobacco (where 18 percent of farmers report hired labor) and Rwanda coffee (8 percent). Labor is less than 4 percent for Kenya coffee farmers, so we exclude this group. Fertilizer use is fairly high for both Malawi and Kenya maize farmers (64 and 70 percent,

---

[4] In Malawi, about 16 percent of maize-producing households sell at least part of their maize crop, compared with 18 percent in Rwanda and 44 percent in Kenya for the last completed cropping season. In Rwanda, 42 percent of sorghum producers also sell part of their crop.

respectively) and 57 percent for tobacco growers in Malawi. Fertilizer usage data are not available for Rwanda and are missing for a large share of the Kenya coffee sample.

In addition to information on harvest and input use, this analysis also requires an estimate of the date of harvest. Since our long recall indicator will refer to only one harvest, we restrict the analysis to crops that are harvested only once during the agricultural calendar. Coffee and tobacco are both single harvest crops. Maize and sorghum, however, can have multiple harvests during the same year. In Rwanda, where the month of harvest information is recorded in the questionnaire for staple crops, we restrict the analysis to households that indicated only one month of harvest, or two consecutive harvest months for maize or sorghum. In Malawi, seasonal rain patterns allow for one main and one secondary (*dimba*) harvest per year. We include only the main harvest from rain-fed plots. In Kenya, the calculation of the harvest date is complicated by high degrees of variability in agricultural zones within the country. The analysis in Kenya is therefore limited to areas and crops with only one harvest per year and reliable information on the harvest month for the crop.[5] This includes only the Rift Valley for maize production, and the Eastern, Central, Western and Rift Valley regions for coffee production. For all three countries, the household interview can pertain to one of two agricultural seasons, depending on the last completed season relative to the month of the interview.[6] Thus, the results are not specific to only one agricultural season.

For the harvest date assigned to households, we would ideally use the actual harvest month for each household (a second best option would be the village-specific harvest month). In the Rwanda EICV, we have and use the direct reporting of the household's harvest month for maize and sorghum. In Malawi and Kenya, and for coffee in Rwanda, we have no information on the month of the harvest (or reported harvest window) for either households or villages in the surveys. In these cases, we assume it is based on other sources. The harvest date is constructed based on input from local agronomists on the normal month of harvest by region to account for spatial variation in production seasons. The harvest month for maize in Malawi is assumed to be April for the southern region, May for the central region, and July for the more arid northern region. The harvest months for tobacco are assumed to be one month earlier. In

---

[5] In Kenya, the design of the survey instrument makes the exclusion of irrigated plots more difficult. The incidence of irrigation in the Rift Valley, however, is quite low, with less than 5 percent of farms using irrigation on maize crops. These observations remain in the dataset, although their exclusion does not markedly change the results.

[6] In Malawi, the data pertain to the 2002/03 and 2003/04 rainy seasons. In Kenya, it is the 2003/2004 and 2004/2005 rainy seasons. In Rwanda, it is the 2000/2001 season.

Kenya, the harvest date for maize in the Rift Valley is assumed to be November. For coffee, the sample is expanded to include the Eastern, Central and Western provinces, with coffee harvest dates assumed to be December in the Central, Western and Rift Valley provinces, and May in the Eastern province. The harvest month for coffee in Rwanda is assumed to be June.

To measure the variation in recall period, we construct a simple binary variable to indicate whether the household was interviewed 8-11 months since the harvest as opposed to 0-3 months since the harvest.[7] In the analysis, we are therefore not using households interviewed in the interim months (4-7 months from harvest). The same binary variable is used for both input and harvest analysis, despite input use being several months before the harvest. As the planting date would be approximately the same number of months before the harvest date in all cases, using the harvest date should not affect the estimations. It is possible that although the mean of the variable of interest does not change with longer recall, the variance might change. Respondents may be less accurate in their reporting further from the event of interest, but this inaccuracy does not necessarily vary systematically in direction. Unfortunately, because we are using only a binary indicator (and not a variable like months between harvest and interview, due to the concerns noted in footnote 7), we are not able to test for this.

The final sample sizes used are smaller than the complete dataset due to restricting the analysis to households interviewed just after harvest and those interviewed further from the most recent harvest (i.e. just before the next harvest). In Malawi, the sample size is 4,435 households for maize, and 620 households for tobacco. In Kenya, for maize production in the Great Rift Valley province, the sample size is 956 households, and for coffee farming in the Eastern, Central, Western, and Great Rift Valley regions, we have 286 households. In Rwanda, the sample size is 1,018 for maize, 633 for sorghum and 362 for coffee.

For the empirical approach, there are two ways in which recall bias may impact the quality of the data. Respondents may inaccurately remember details, such as the amount of fertilizer used or

---

[7] Ideally we would include a variable that measured the number of months from the harvest to the interview, ranging from 0 to 11. This would also allow us to explore partial linear regression results to allow for non-linear patterns in the time effect. However, our calculations of the minimum detectable effects to be able to detect changes at 90 percent power and a 5 percent significance level showed that the monthly specification would yield significance on time-since-harvest only for very large recall effects. With our chosen specification (a binary 0/1 for long versus short recall), on average across outcome variables, the calculations show that the minimum detectable is less than a 10 percent change per month of added recall from the mean. For some outcomes, however, the change is closer to 15 percent -- as is the case for the amount (in kgs) of fertilizer.

crops harvested, or they may forget events altogether, such as the use of fertilizer or the hiring of labor. To that end, we examine both the quantity of harvest in kilograms and local currency (assuming that farmers do not forget the harvest all together) and the application and quantity of inputs. We estimate the following specification:

$$Y_h = \alpha + \beta T_h + \lambda X_h + \gamma D + \varepsilon_h \qquad \text{(Eq. 1)}$$

where Y includes outcome variables (harvest, sales, input application), T is an indicator for whether the interview of the household took place 8-11 months since harvest (or 0-3 months since harvest), X is a vector of household characteristics (landholdings, household head gender, age, education, an indicator for head being main decision maker in household, a dummy variable for agricultural season, and, for harvest regressions, dummy variables for the unit in which the harvest was reported). D consists of dummy variables for geographic region and ε is the stochastic error term, which is randomly distributed across households.

This informs on the overall average extent to which there is under/over reporting with longer recall. It is possible, however, that recall bias, while on average not statistically significant, varies by some key characteristics. Larger farmers, for example, may have more recall bias, since they have higher production, or less recall bias because they keep better accounting records or have better unobservable entrepreneurial skills. We extend the basic specification above by introducing interaction terms of our "long recall" dichotomous variable with a subset z of explanatory variables X and estimate:

$$Y_h = \alpha + \beta T_h + \lambda X_h + \beta_2 (T_h * z_h) + \gamma D + \varepsilon_h \qquad \text{(Eq. 2)}$$

where z includes landholdings, female headship, and head's education. In the case of landholdings and head's education, both continuous variables, we interact the "long recall" variable with $z_h - \bar{z}_h$, where $\bar{z}_h$ is the vector of sample averages for the continuous variables being considered. This specification provides results that are easier to interpret, since β alone then measures the impact of longer recall duration at the mean level of X (Wooldridge, 2001).[8]

4. Results

---

[8] We thank an anonymous referee for suggesting the approach.

Starting with the harvest of staple crops, Table 1 shows partial results of the regressions of harvest amounts for both staple and cash crops in the three countries. We estimate four specifications, each presented in the four rows in Table 1. In the base model with no interaction term, the binary variable for long recall is regressed on quantities harvested of each crop and includes a number of control variables at the household and geographic level. We estimate three additional models in which the long recall indicator variable is interacted with farm size, education and gender of the household head to explore possible heterogeneity in recall accuracy across different groups of respondents. The reported coefficients in Table 1 refer only to the elapsed time variable and its interactions.[9]

Across both countries and crops, the results consistently reject the presence of recall bias in harvested quantities; the long recall variable is not significant in any specification, suggesting that the reported amount of harvest of each crop is not statistically lower (or higher) for households interviewed further from the harvest date. Other covariates included (not reported in the table) generally have coefficients in the expected direction. The size of the (insignificant) coefficient for long recall for staple crops is small in relation to the mean of the harvest, whereas it is larger for the cash crops where there is more variation in harvest levels.

As we introduce different specifications with an interaction term between the recall variable and specific household characteristics (land, female headship and education), in very few cases are the interaction terms significant. For example, larger landholdings in Rwanda are associated with greater under-reporting of maize harvests *vis a vis* smaller holdings for interviews further from harvest. Households with one acre above the mean landholdings report, on average, about 7 percent less maize output from the mean level of almost 100 kilos when that household is interviewed further from the harvest.

Similar results are found when we interact our recall variable with education and female headship, with only one coefficient in each being significant. In the case of education, only for sorghum in Rwanda do we find that more-educated households tend to report lower production compared with the average education level. For sorghum in Rwanda, male-headed households tend to underestimate quantities reported when interviewed 8-11 months after the harvest compared with an interview just after harvest; we find no significant difference in reporting of

---

[9] The full set of results is available from the authors upon request.

13

sorghum output for female-headed households in Rwanda. We do not find any further instances of significance for level of education or for female-headed households.

Turning to recall of cash crop sales, we assume that farmers will report cash crop sales more accurately than quantities, and in Table 2 we find some evidence of recall effects, although limited to tobacco sales in Malawi. The direction of the bias, however, goes against our initial intuition, with respondents reporting higher sales for longer recall periods. This finding holds even when we drop outliers in sales. Introducing the land interaction terms suggests that such bias among tobacco farmers in Malawi is driven by smaller landholders. Since we found no impact on the kilograms produced for Malawi tobacco (Table 1 results), this finding could be the result of an overestimate of either the recall of the portion of total harvest sold or it may reflect differences in the recall of prices received from sales. Regressions using the amount sold, measured in kilograms as opposed to local currency units, show similar results, therefore indicating that the source of this bias in the Malawi tobacco recall data is at least partly tied to reports of the quantity sold (results not presented). No differences from the base model are detected after introducing the other two interaction terms in Malawi. Male-headed households growing coffee in Kenya also seem to overestimate sales when interviewed further from harvest.

Turning now to inputs, using a probit specification, we explore the recall bias in reporting any hired labor for staple and cash crops (Table 3). One might expect that hiring labor would be an event of high salience, especially for low-income smallholders for whom there is some payment (cash or in-kind) necessary for this input, resulting in less recall effects. However, if little is expended on hired labor (that is, few days of hired labor particularly in relation to the household's own labor input levels), it might be a decision of low salience and hence there will be recall bias in reporting. There is some evidence of recall bias in reporting for maize in Kenya and Malawi, but with different signs. In the case of Kenya, longer recall is associated with approximately a 32 percentage point increase in the probability of reporting any hired labor. This is an increase of nearly 100 percent from the mean. Conversely, in Malawi, longer recall is associated with a decrease in the probability of reporting any hired labor of about 36 percentage points, an even more drastic decline from the mean. This recall decay for maize in Malawi is offset with education level of the head, but it grows for larger landholders. There are generally no significant recall effects for Rwanda sorghum and coffee, although small coffee farmers under-report when asked further from the event. There is a small effect for female heads hiring labor for maize plots in Malawi (who underestimate with longer recall).

With regard to fertilizer usage, we employ three different estimations: a probit model for fertilizer usage, a Tobit model for the amount of fertilizer application in kilograms, and an OLS model on non-zero values of fertilizer purchases (Tables 4, 5, and 6, respectively). Rwanda is excluded from these regressions since the EICV survey did not collect this information. For fertilizer use (Table 4), the coefficients on the recall variables are not significant, suggesting the lack of recall bias, which is consistent with fertilizer application being a decision of high salience for farmers. Even after introducing the interaction terms, again, we find little evidence of recall bias overall. A few exceptions are by gender of the head in Malawi. There are recall effects for fertilizer use on maize plots for female-headed households and on tobacco plots for male-headed households, with underestimation of use when the household is interviewed further from the harvest (whereas in Table 3 both male and female-headed households under-report hired labor for maize plots when interviewed further from the harvest date).

The level of fertilizer use is presented in Table 5. For maize in Kenya, although not significant, the results suggest that smaller landholders are more likely to under-report with recall duration. The result is the opposite for Malawi tobacco, where under-reporting is statistically significant for larger landholders; there is also more under-reporting among male-headed households. For maize in Malawi, overall fertilizer use levels decline with recall duration, although the results are not statistically significant. This recall decay is more concentrated among female farmers and is offset with the education of the head and smaller landholdings.

With respect to the value of fertilizer and the OLS results (Table 6), Kenyan farmers report larger values when interviewed further from harvest and the value is associated with smaller landholdings. Combined with the kilo quantity results in the previous table, this suggests that smaller landholders are over-estimating the price paid for fertilizer when interviewed further from the purchase itself. This is also the case for female heads growing maize in Malawi; they report higher expenditure on fertilizer when interviewed further from harvest (although a lower quantity is used in the previous table).

As a robustness check, we repeat the above analysis excluding the control variables (including only indicator variables for season and geographic area). Given that households are randomly assigned across interview months, we expect that including these control variables will not impact our results. This is what we find when we re-estimate Tables 1-6 excluding the controls.

It could also be hypothesized that what we record as recall bias is actually caused by interviewer error. Due to the geographic distribution of fieldwork and the construction of the time elapsed variable, if certain interviewers did not accurately collect harvest or input information, this could be correlated with recall bias. To test for interviewer effects, enumerators' fixed effects were added to the Malawi regressions (results not presented).[10] The coefficient on the main variable of interest, time elapsed since harvest, does not change, and remains not statistically significant.[11]

Finally, it is important to recognize that the harvest date is assigned to households with some noise, since we do not have household specific reports on month of harvest (with the exception of Rwanda sorghum) and there can be variation within regions as to the specific harvest month of households. By using the simple binary indicator for a longer or shorter recall duration, we have tried to avoid this potential noise. However, this may still be problematic for households interviewed near the cutoff to the completed season. Specifically, this is where we assign time elapsed as 0-1 months when it should be 11 months (or 11 months where it is actually 0-1 month). We do robustness checks for this by excluding observations where time elapsed is less than 2 months or more than 10 months.

5. Conclusion

Agriculture is critical for development, particularly in Sub-Saharan Africa, where a large share of households depend on farms as their main source of livelihood. Despite this, data on agriculture are often perceived to be both lacking and, when available, of poor quality (UN Statistics Commission, 2010). Little research has been undertaken to date on the quality of agricultural statistics. Where data are available that link agriculture with non-farm activities, information on farming activities may be collected as part of a larger, multi-purpose survey based on a single interview. At best, households are interviewed over a short period of time just right after the

---

[10] Interviewer identification variables are not available for Kenya or Rwanda.
[11] As noted earlier, it is possible that recall bias does not vary linearly with time. For example, a farmer could accurately remember production details up to a certain point following the event, then begin to forget as time passes. Alternatively, the slope could increase and then decrease over time, producing no slope on average. That is, the speed of decay might change as time from harvest increases. Unfortunately, due to small samples and the variance in our outcome measures, with the revised specification (a simple binary variable for longer versus shorter recall period), we are not able to explore this issue.

harvest, which will minimize recall bias for crop production estimates. However, even in this case, input application may have occurred several months before and the harvest may have been conducted over several months. Adding to this, survey data are often collected over several months, often twelve or more. These data potentially suffer from recall bias since both farming decisions and data collection occur over the course of several months, leading to long and variable recall periods across households. This paper studies the extent of recall bias in agricultural statistics, using data from three national surveys conducted over a year such that the recall period has sufficient random variation across sample households. This work does not speak to concerns about the overall quality of data from smallholders. However, it does address at least one aspect of data collection – length of recall period – which could compromise data quality, and assesses whether differences in recall periods consistently affect the reported values.

Using various specifications on key input and crop production measurements, we generally find little evidence of recall bias for the average farm household. With regard to harvest estimates and fertilizer usage, we find only isolated incidents of statistically significant results for some specifications including interaction terms. Findings are similar for cash crop sales in Kenya and Rwanda and labor use in Rwanda. The most evidence of recall bias in production data is found for tobacco sales in Malawi, where, interestingly, the direction of the bias is generally positive, contradicting the initial hypothesis that the forgetting of events as time passes would lead to under-reporting and suggesting, instead, other possible sources of bias. We also find statistically significant evidence of recall bias in labor use for maize crops in Kenya and Malawi. These results represent substantial changes in terms of magnitude with regard to the mean, but they are in opposing directions, with over-reporting in Kenya and under-reporting in Malawi as recall durations are increased.

Finally, we find some evidence supporting the salience hypothesis. Hired labor usage requires payments and is not common, suggesting some salience. Yet, we find recall decay in reporting any hiring in Kenya and Malawi. We find less evidence of decay in labor in Rwanda, where it is a comparatively much rarer event than in Kenya and Malawi. Although not very common in any country, hiring labor may be an event of small relevance, especially in relation to the household's own labor contribution. Fertilizer usage is a single, more expensive event for farm households. In view of the prohibitive costs of fertilizers and the poor welfare of most farmers in our samples, the lack of recall bias in financially taxing events such as fertilizer purchase and

application is not surprising. However, we do not seem to find differential recall bias between staple and cash crop harvests, although this might be partially explained by the fact that many households also sell staples, which would make information on staples equally salient as the data on cash crops.

Overall, the results seem to suggest that the majority of agricultural data collected through single visits over several months do not suffer from the large recall decay initially hypothesized, at least for the types of events and in the contexts studied here. Given the lack of studies on the topic, further research is warranted since this work is limited to specific countries and crops. Moreover, data with community or household specific harvest dates would offer a more refined measure of time elapsed from the harvest to the interview.

Table 1: OLS Harvest Amount (in kilos)

	Staple Crops				Cash Crops		
	Kenya Maize	Malawi Maize	Rwanda Maize	Rwanda Sorghum	Kenya Coffee	Malawi Tobacco	Rwanda Coffee
n	956	4,435	1,018	1,713	286	620	362
Long recall	82.45	-59.83	19.76	-13.29	159.53	112.14	4.57
	(181.09)	(70.53)	(22.41)	(12.08)	(167.26)	(84.88)	(6.67)
Long recall	178.53	-52.58	20.46	-14.03	158.29	100.63	2.92
	(186.90)	(84.86)	(22.39)	(12.08)	(162.47)	(110.14)	(6.31)
Long recall x (land – mean land)	-78.11	21.90	**-6.52****	6.78	33.26	13.76	2.51
	(124.77)	(75.41)	(3.23)	(4.87)	(59.20)	(77.63)	(1.87)
Long recall	120.30	-53.68	19.26	-15.64	156.66	109.68	5.34
	(245.93)	(70.16)	(22.38)	(12.23)	(165.53)	(79.41)	(6.69)
Long recall x (head's education – mean head's educ)	-6.11	18.56	-1.84	**-7.26****	6.84	3.35	2.55
	(33.06)	(15.02)	(3.19)	(3.69)	(24.15)	(21.58)	(2.25)
Long recall	144.94	-53.80	19.31	-23.04	246.83	116.86	5.11
	(190.70)	(80.64)	(26.33)	(14.32)	(194.98)	(82.32)	(8.46)
Long recall x female head	-271.41	-23.77	1.49	**28.70****	-231.60	-64.44	-1.93
	(412.71)	(70.41)	(20.53)	(14.34)	(158.11)	(153.51)	(10.03)
Mean of harvest amount	1276.76	663.55	91.37	118.76	554.85	467.67	46.62
Standard error	107.86	32.07	7.44	5.07	53.60	47.90	3.65
Median value	630	350	50	70	300	221	30

Note: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Samples are restricted to households that reported growing each of the crops. Robust standard errors are in parentheses. "Long recall" is a binary variable that indicates that the interview took place 8-11 months after harvest. The omitted category is 0-3 months after harvest. Other covariates included but not presented are acres of land, female headship, age and age squared of head, years of education of head, household head is main decision maker, geographic and season dummies, and the original unit in which the harvest was reported.

Table 2: OLS Amount of Crop Sales (in local currency units)

	Kenya Coffee	Malawi Tobacco	Rwanda Coffee
n	270	498	362
Long recall	2,764.26	2,543.84*	-5,093.12
	(3,821.65)	(1,378.03)	(5,081.49)
Long recall	2,958.60	3,743.56***	-5,053.94
	(3,737.30)	(1,098.12)	(5,014.92)
Long recall x (land – mean land)	1,465.13	-2,491.21***	-50.55
	(1,182.63)	(677.00)	(2,065.05)
Long recall	2,880.11	2,696.19*	-4,853.44
	(3,780.22)	(1,405.35)	(5,018.99)
Long recall x (head's education – mean head's educ)	-167.33	-165.11	793.73
	(398.42)	(254.48)	(1,423.65)
Long recall	5,566.68	2,511.70*	-7,363.24
	(4,179.17)	(1,437.95)	(6,544.00)
Long recall x female head	-7,019.12***	328.68	7,970.89
	(2,591.88)	(1,815.00)	(10,359.69)
Mean of sales amount	7,598	10,174	14,811
Mean standard error	1,112	441	2,264
Median value	3,563	8,348	6,000

Note: *** p<0.01, ** p<0.05, * p<0.1. Samples are restricted to households with any sales. Robust standard errors are in parentheses. "Long recall" is a binary variable that indicates that the interview took place 8-11 months after harvest. The omitted category is 0-3 months after harvest. Other covariates included but not presented are acres of land, female headship, age and age squared of head, years of education of head, household head is main decision maker, geographic and season dummies, and the original unit in which the harvest was reported.

Table 3: Labor Use Probit

	Staple Crops				Cash Crops	
	Kenya Maize	Malawi Maize	Rwanda Maize	Rwanda Sorghum	Malawi Tobacco	Rwanda Coffee
n	953	4,380	1,221	1,713	541	401
Long recall	**0.32****	**-0.36*****	0.01	0.02	0.10	-0.03
	(0.14)	(0.10)	(0.02)	(0.02)	(0.36)	(0.03)
Long recall	**0.39*****	**-0.39*****	0.01	0.02	0.15	**-0.05***
	(0.14)	(0.10)	(0.02)	(0.02)	(0.32)	(0.03)
Long recall x (land – mean land)	-0.05	**-0.15*****	-0.00	0.01	-0.04	**0.02****
	(0.03)	(0.05)	(0.01)	(0.01)	(0.14)	(0.01)
Long recall	**0.32*****	**-0.37*****	0.01	0.02	0.23	-0.01
	(0.14)	(0.10)	(0.02)	(0.02)	(0.38)	(0.04)
Long recall x (head's education – mean head's education)	0.03	**0.05*****	**0.01***	-0.00	-0.07	-0.01
	(0.02)	(0.02)	(0.01)	(0.01)	(0.09)	(0.01)
Long recall	**0.33****	**-0.32*****	0.03	0.01	0.17	-0.02
	(0.16)	(0.10)	(0.02)	(0.02)	(0.36)	(0.03)
Long recall x female head	-0.03	-0.15	**-0.05****	0.04	-0.67	-0.06
	(0.29)	(0.14)	(0.02)	(0.04)	(0.73)	(0.07)
Hired labor (percent)	34.2	19.4	4.0	9.4	18.9	7.7

Note: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Samples are households that report growing the crop. Marginal probabilities are presented from the probit estimation. Robust standard errors are in parentheses. "Long recall" is a binary variable that indicates that the interview took place 8-11 months after harvest. The omitted category is 0-3 months after harvest. Other covariates included but not presented are acres of land, female headship, age and age squared of head, years of education of head, household head is main decision maker, and geographic and season dummies.

Table 4: Fertilizer Use Probit

	Kenya Maize	Malawi Maize	Malawi Tobacco
N	872	4,380	598
Long recall	-0.04	-0.06	-0.44
	(0.20)	(0.10)	(0.33)
Long recall	-0.06	-0.07	-0.06
	(0.20)	(0.11)	(0.38)
Long recall x (land – mean land)	0.03	-0.02	-0.54
	(0.03)	(0.04)	(0.37)
Long recall	-0.07	-0.06	-0.44
	(0.20)	(0.10)	(0.34)
Long recall x (head's education – mean head's educ)	-0.05	0.01	-0.04
	(0.03)	(0.02)	(0.07)
Long recall	-0.11	-0.01	**-0.71****
	(0.22)	(0.11)	(0.28)
Long recall x female head	0.26	**-0.20***	**2.05*****
	(0.28)	(0.11)	(0.64)
Fertilizer use (percent)	70.5	63.9	56.8

Note: *** p<0.01, ** p<0.05, * p<0.1. Samples are households that report growing the crop. Marginal probabilities are presented from the probit estimation. Robust standard errors are in parentheses. "Long recall" is a binary variable that indicates that the interview took place 8-11 months after harvest. The omitted category is 0-3 months after harvest. Other covariates included but not presented are acres of land, female headship, age and age squared of head, years of education of head, household head is main decision maker, and geographic and season dummies.

Table 5: Fertilizer Amount Tobit (in kilos)

	Kenya Maize	Malawi Maize	Malawi Tobacco
N	889	4,437	600
Long recall	-5.76	-137.51	-56.52
	(81.22)	(88.19)	(42.19)
Long recall	-97.80	-140.99	8.29
	(136.30)	(90.09)	(59.67)
Long recall x (land – mean land)	79.40	-13.75	**-54.29***
	(100.62)	(23.71)	(27.91)
Long recall	-8.25	**-134.50***	-61.34
	(81.69)	(87.45)	(46.33)
Long recall x (head's education – mean head's educ)	13.06	13.44	7.65
	(18.99)	(10.54)	(14.86)
Long recall	-11.09	-88.79	**-78.14****
	(88.44)	(84.10)	(37.87)
Long recall x female head	22.87	**-190.38****	320.93
	(109.20)	(92.26)	(200.09)
Mean of fertilizer amount	103.80	83.17	134.33
Standard error	20.82	17.93	15.80
Median value	25	0	50

Note: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Samples are households that report growing the crop. Robust standard errors are in parentheses. "Long recall" is a binary variable that indicates that the interview took place 8-11 months after harvest. The omitted category is 0-3 months after harvest. Other covariates included but not presented are acres of land, female headship, age and age squared of head, years of education of head, household head is main decision maker, and geographic and season dummies.

Table 6: Fertilizer Value OLS (in local currency units)

	Kenya Maize	Malawi Maize	Malawi Tobacco
N	578	2,022	463
Long recall	-263.84	464.73	744.75
	(651.18)	(323.82)	(1,282.60)
Long recall	**1,272.86***	469.62	1,474.85
	(650.09)	(326.07)	(1,265.28)
Long recall x (land – mean land)	-1,311.30*	85.06	-584.53
	(784.28)	(215.96)	(728.63)
Long recall	-352.99	424.35	694.18
	(649.18)	(317.16)	(1,320.98)
Long recall x (head's education – mean head's educ)	176.75	**101.08****	172.77
	(111.29)	(49.67)	(443.19)
Long recall	74.15	279.12	806.56
	(505.31)	(298.04)	(1,316.36)
Long recall x female head	-1,446.06	890.60*	-482.51
	(1,439.42)	(504.76)	(1,977.92)
Mean of fertilizer	3850.04	2925.76	6610.44
Standard error	417.07	122.34	591.57
Median value	1,750	1,800	3,400

Note: *** p<0.01, ** p<0.05, * p<0.1. Samples are households that report growing the crop. Robust standard errors are in parentheses. "Long recall" is a binary variable that indicates that the interview took place 8-11 months after harvest. The omitted category is 0-3 months after harvest. Other covariates included but not presented are acres of land, female headship, age and age squared of head, years of education of head, household head is main decision maker, and geographic and season dummies.

Appendix Table 1: Sample Means

	Kenya		Malawi		Rwanda		
	Maize	Coffee	Maize	Tobacco	Maize	Sorghum	Coffee
Long recall since harvest (months)	5.15	6.68	8.99	9.22	4.92	5.67	6.31
Household characteristics	5.99	5.01	4.68	5.24	5.06	5.01	5.52
Household size	1.29	2.24	2.33	3.65	1.84	2.29	2.46
Landholdings (acres)	0.17	0.15	0.28	0.22	0.13	0.16	0.15
Non-farm enterprise (percent)	46.67	54.12	43.82	40.66	44.04	43.74	48.10
Age of household head	0.24	0.27	0.25	0.10	0.30	0.29	0.26
Female headed (percent)	6.12	6.60	4.02	4.85	2.87	2.92	2.66
Years of education of head	5.15	6.68	8.99	9.22	4.92	5.67	6.31
Number of households	956	286	4,435	620	633	1,018	362

Note: See text for description of the samples and datasets. Sample sizes vary in regressions depending on sub-samples for different agricultural outcome variables.

REFERENCES

Bound, John, Charles Brown, and Nancy Mathiowetz. 2001. "Measurement Error in Survey Data." In J. Heckman and E. Leamer (eds.), *Handbook of Econometrics Vol. 5*. Amsterdam: North-Holland, Elsevier Science.

Bradburn, Norman, Seymour Sudman, and Brian Wansink. 2004. *Asking Questions: The Definitive Guide to Questionnaire Design*. San Francisco: Jossey-Bass.

Coleman, Gilroy. 1983. "Analysis of Memory Bias in Agricultural Labor Data Collection: A Case Study of Small Farmers in Nigeria." *Journal of Agricultural Economics* 34: 79-86.

Davis, Benjamin, Paul Winters, Gero Carletto, et al. 2010. "A Cross-Country Comparison of Rural Income Generating Activities." *World Development* 38(1): 48-63.

Deaton, Angus, and Margaret Grosh. 2000. "Consumption." in Margaret Grosh and Paul Glewwe (eds.), Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study. Washington, D.C.: World Bank.

Fermont, Anneke, and Todd Benson. 2011. "Estimating Yield of Food Crops Grown by Smallholder Farmers: A Review in the Ugandan Context." IFPRI Uganda Strategy Support Program Working Paper no. USSP 05.

Gibbs, M, R.K. Linder, and A.J. Fischer. 1986. "Reliability of Two Survey Techniques: A Study of Innovation Discovery by Farmers." *The Statistician* 35: 429-39.

Gibson, John. 2002. "Why Does the Engel Method Work? Food Demand, Economies of Size and Household Survey Methods." *Oxford Bulletin of Economics and Statistics* 64(4): 341–60.

_____. 2006. "Statistical Tools and Estimation Methods for Poverty Measures Based on Cross-Sectional Household Surveys." In Jonathan Morduch (ed.), *Handbook on Poverty Statistics*. United Nations Statistics Division.

Gibson, John, and Bonggeun Kim. 2007. "Measurement Error in Recall Surveys and the Relationship between Household Size and Food Demand." *American Journal of Agricultural Economics* 89(2): 473–89.

Judge, George, and Laura Schechter. 2009. "Detecting Problems in Survey Data Using Benford's Law." *Journal of Human Resources* 44(1): 1-24.

Kenya National Bureau of Statistics. 2007. Basic Report of Well-Being in Kenya. Nairobi: Kenya National Bureau of Statistics.

Loftus, Elizabeth, and Wesley Marburger. 1983. "Since the Eruption of Mt. St. Helens, Has Anyone Beaten You Up? Improving Accuracy of Retrospective Reports with Landmark Events." *Memory and Cognition* 11(2): 114-20.

Malawi National Statistics Office. 2005. Malawi Second Integrated Household Survey 2004-05 Basic Information Document. Zomba, Malawi: National Statistics Office.

Rwanda Statistics Department. 2002. *Integrated Household Living Conditions Survey in Rwanda Final Report.* Kigali: Ministry of Finance and Economic Planning – Statistics Department.

Scott, Christopher, and Ben Amenuvegbe. 1991. "Recall Loss and Recall Duration: An Experimental Study in Ghana." *Inter-Stat* 4(1): 31–55.

Smith, James, and Duncan Thomas. 2003. "Remembrances of Things Past: Test-Retest Reliability of Retrospective Migration Histories." *Journal of the Royal Statistical Society* 166(1): 23-49.

Sudman, Seymour, and Norman Bradburn. 1974. *Response Effects in Surveys: A Review and Synthesis.* Chicago, IL: Adline.

UN Statistics Commission. 2010. "Global Strategy to Improve Agricultural Statistics."

Wooldridge, Jeffrey M. 2001. Econometric Analysis of Cross Section and Panel Data. The MIT Press. Cambridge, Massachusetts and London, England.

World Bank. 2008. *World Development Report 2008: Agriculture for Development.* Washington, D.C.: World Bank.