



## COMPARATIVE ANALYSIS OF CROSS SECTIONAL AND PANEL DATA COLLECTION FOR AGRICULTURAL STATISTICS IN MALAWI: DOES AN INTEGRATED VEHICLE PROVIDE AN EFFECTIVE FRAMEWORK?

PANGAPANGA Innocent Phiri

Agricultural and Resource Economist  
Lilongwe University of Agriculture and Natural Resources,  
Bund College of Agriculture, P.O. Box 219, Lilongwe, Malawi.  
[www.luanar.mw](http://www.luanar.mw). [Phiriinnocent@gmail.com](mailto:Phiriinnocent@gmail.com).

PANGAPANGA Lucy Thangalimodzi

Research for Development-Smart Projects  
P.O. Box 452, Nkhukwa, Kalonga Village, Lilongwe, Malawi.  
DOI: 10.1481/icasVII.2016.f34b

### ABSTRACT

This paper conducts a comparative analysis between cross sectional and panel component data collection methods for agricultural statistics in Malawi. The paper uses meta-analysis and / or revelatory case study approach to characterize a stand over cross sectional and panel data collection in Malawi. In this paper, we adopt a conclusion statement made by Deaton, Solon and Ashenfelter (1986) that, while there are genuine difficulties, there are good arguments for collecting panel data. Collecting panel data in Malawi is a sensible enterprise where both the dynamics and annual level statistics are generated. Despite advantages of a panel component in Malawi, they are very expensive as compared to cross section survey. We recommend a panel that is integrated in cross sectional data collection as is the case in the current Integrated

Household Survey design. This will not only improve and address data demands but also become cost effective for agricultural statistics stakeholders as would have been realized through a fully-fledged panel.

**Keywords: Malawi, Cross Sectional, Panel Data Collection, Agricultural Statistics.**

## **1. Study Context and Problem Statement**

Agricultural development is an essential engine of growth in Malawi (GoM, 2012). It is also indicated as effective mechanism for combating food poverty. It often results in greater benefits accruing to the poorest segments of the population (National Statistical Office (NSO), 2014; Ravallion and Chen, 2007). It contributes about 40% to the Gross Domestic Product (NSO, 2016). Approximately 90% of households in Malawi depend on agricultural for their welfare security (NSO, 2015).

Success made in agriculture can only be tracked and attributed if consistent data is made available over time and in space. Prevailing data collections have produced conflicting agricultural statistics. Climatic change and weather related effects have led to high demand of agricultural statistics. World Bank (2010) argues that rapidly changing nature of agriculture in less developed countries especially in the errors of climatic change and the emergence of new issues make the available data and methods obsolete. Conversely, agricultural statistics are becoming expensive to collect, compile and analyze. Limited public finance and dwindling donor support of data collection exercises in agricultural sector is another obstacle that NSO continue to encounter. In other words, statisticians have to innovate new ways of achieving statistical demands in the agricultural sector in order to provide timely, relevant and consistent agricultural data (Binswanger, 2008). Innovations should at a minimum disaggregate agricultural data into hard to reach population such as women, rural/urban and many others (International Development Committee, 2013).

Existence of serious weaknesses in agricultural statistics also persist throughout sub-Saharan Africa of which Malawi is not an exception (Carletto et al, 2010). Compliance to statistical standards in most of these countries remain low (FAO, 2008). Knowledge about agriculture and its impact on welfare and equity is limited by the lack of available, high quality, and consistent data on rural and gender disaggregated households. Although studies have attempted to understand intra household dynamics, data is lacking to validate cases in terms of intra-household decision-making within agricultural activities (Koirala, Mishra and Sitienei, 2015).

On other hand, agricultural programming would be effectively realized and implemented if policy makers exploit data evidence based decisions. Of which data for informed decisions have to meet quality, timely, relevant and coordinated scopes (Binswanger, 2008). The Malawi National Statistical Office as a government department has this mandate and to exploit best practices that are innovative to ably capture agricultural data in a very cost effective manner and that meet quality dimensions aforementioned. One way would be to must partner with international organizations such as the World Bank, International Food Policy Institute (IFPRI and other institutions to have quality data in the agricultural sector. These institutions would provide an automated technical assistance that most government owned statistical offices lack.

In Malawi, NSO has embarked on this agenda through integrating household surveys with a huge agricultural component where most indicators are tracked. A data quality assurance framework has been developed and declares the needs of adopting a 10 year integrated household survey programme that streamline data collection activities (NSO, 2015). This would achieved both cost effectiveness and efficiency. Similarly, NSO has also reduced the time between panels to facilitate quick and comparable dynamic data of households' welfare in the country. Nonetheless, a rush decision to adopt a survey programme without a verified research would yield more costs than benefits. In this paper, we attempt to provide empirical responses conducting a comparative analysis of cross section, panel and integrated data collections in terms of their value for money, effectiveness and feasibility?

## 2. Rationale

Agricultural and rural development policies have gained in importance since independence (Moreddu, 2011). In Malawi, more complex information has been demanded to evaluate them (Matchaya, 2014). These new types of information that are at once more local, complex, multidisciplinary and integrated (Moreddu, 2011). Rural development and rural statistical indicators require information that goes beyond the agricultural sector, that is available at a disaggregated level, and which is comparable across area such as agricultural versus non-farm activities, males versus female plot managers (FAO, 2015). This requires significant resources sometimes to the detriment of basic agricultural information. The huge outcry of most Central Statistical Offices to become autonomy is not an issue of professional independence but of financial independence. This means existing data collection approaches may be finance demanding. In other words, a properly design data collection vehicle is critical for most National Statistical Offices for their continued existence and role in the economy.

### 3. Research Methodology

In this paper, we first adopt a meta-analysis (desk research) of cross section and panel data collection. In other words, we adopted a revelatory case study approach following Fitzgerald and Fitzgibbon (2014). We basically review publishable articles of the living standards measurement study team of the World Bank and World Class Universities that have provided a thorough discourse on data collection approaches. Survey reports that have been reviewed include Integrated Household Survey and Agricultural Production Estimate Surveys. Second, we have tried to interact with technocrats in agricultural statistics so that we derive experiential based understanding of cross section and panel data collection methods. We have lastly tried to relate such data collection approaches to Malawi.

### 4. Value for Money

Agricultural surveys, such as Agricultural Production Estimate Survey, Cost of Production and other, are determined based on national priorities within the limitations of cost and other constraints (World Bank, 2010). They capture quarterly information on a variety of topics, such as crop areas, yields, production, livestock, pesticides, etc. Carletto et al (2010) points out that such surveys usually collect information on structural characteristics; economic characteristics; labor force characteristics and share of income that goes to household. In the case of Malawi, an agricultural production estimate survey is, a traditional cross sectional survey, conducted quarterly to capture the whole agricultural value chain that is from land preparation, planting and harvesting (MoA, 2016). Nonetheless, it provides data on production, area and yield quantities.

On the other hand, an Integrated Household Survey (Malawi IHS) or Integrated Surveys on Agriculture is more complex than traditional cross-sectional data. This is both analytically and in terms of actual data collection (NSO, 2012). It combines both a cross section and panel data collections. Panel survey sample component are regularly refreshed to maintain the national representativeness of the data (NSO, 2016). It collects a wide variety of indicators as compared to a traditional agricultural production estimate survey.

In other words, an integrated household survey fits well in the needs of the country. On average a panel embedded survey in Malawi collects almost all FAO minimum core indicators in agricultural sector. Agriculture questionnaires collect information on a core set of indicators that have been identified through a consultative process with several experts. Additionally, an Integrated Household Survey

programmesupports the National Statistical Office (NSO) to generate nationally representative, household panel data with a strong focus on agriculture and rural development through technical assistance, hands on trainings and others.

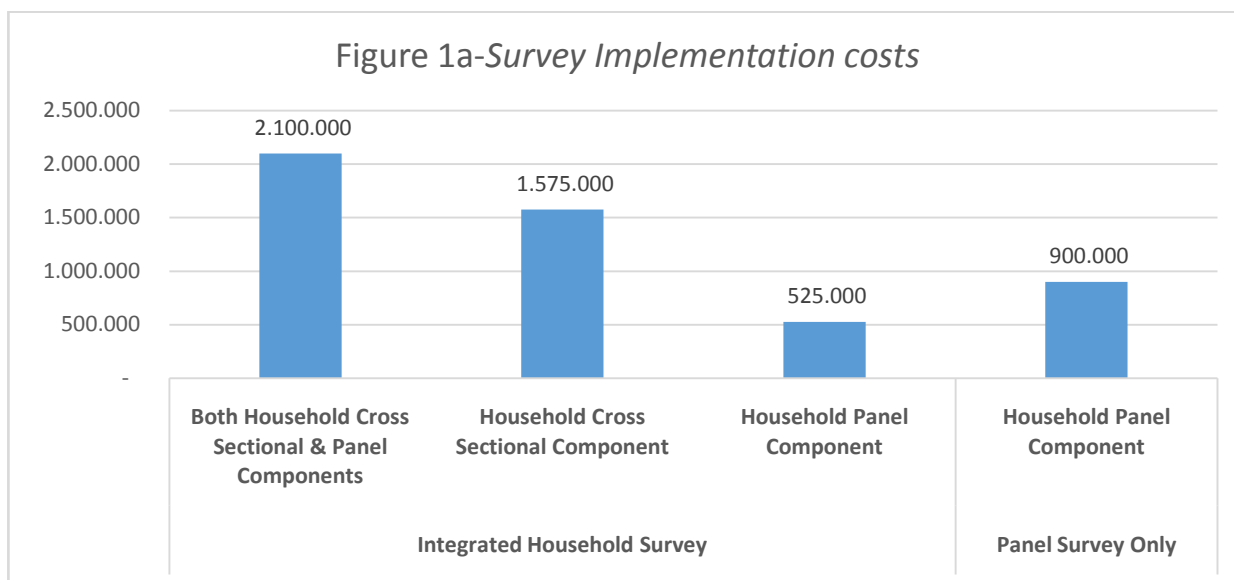
Table 1: Core minimum indicators stipulated by UN FAO

Indicators	FAO minimum core indicators	Integrated Surveys on Agriculture	Agricultural Production Estimate Survey
Geo-variables	√	√	
Garden and Plot Details (rainy/dry season)	√	√	√
Input (Coupon) Use (rainy/dry season)	√	√	√
Crop Cultivation (rainy/dry season)	√	√	√
Crop Sales (rainy/dry season)	√	√	√
Crop Storage (rainy/dry season)	√	√	√
Tree Crop Cultivation and Sales	√	√	√
Livestock	√	√	√
Access to extension services	√	√	√
Household demographics, Education, Health	√	√	X
Housing, Electricity and Sanitation	√	√	X
Food Consumption and Expenditure	√	√	X
Food Security and Anthropometrics	√	√	X
Non Food Expenditures	√	√	X
Assets and Household Ownership	√	√	X
Household Enterprises and Employment	√	√	X
Income, Credit/Loans	√	√	X
Safety Nets, Shocks and Copping Strategies	√	√	X
Internal and International Migration and Remittances	√	√	X

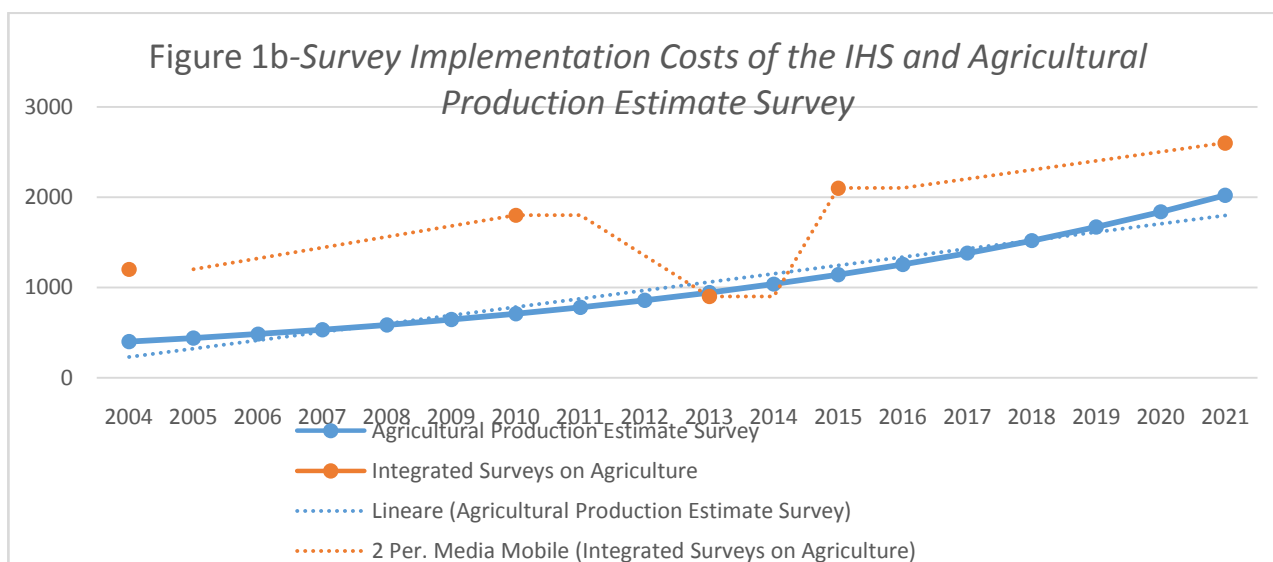
Sources: Ministry of Agriculture (2016); NSO (2014); FAO (2011); FAO (2008)

The Integrated Household Survey is expensive as roughly compared with the Agricultural Production Estimate Survey (APES) (see Fig 1a& 1b). A stand-alone household panel survey component costs more

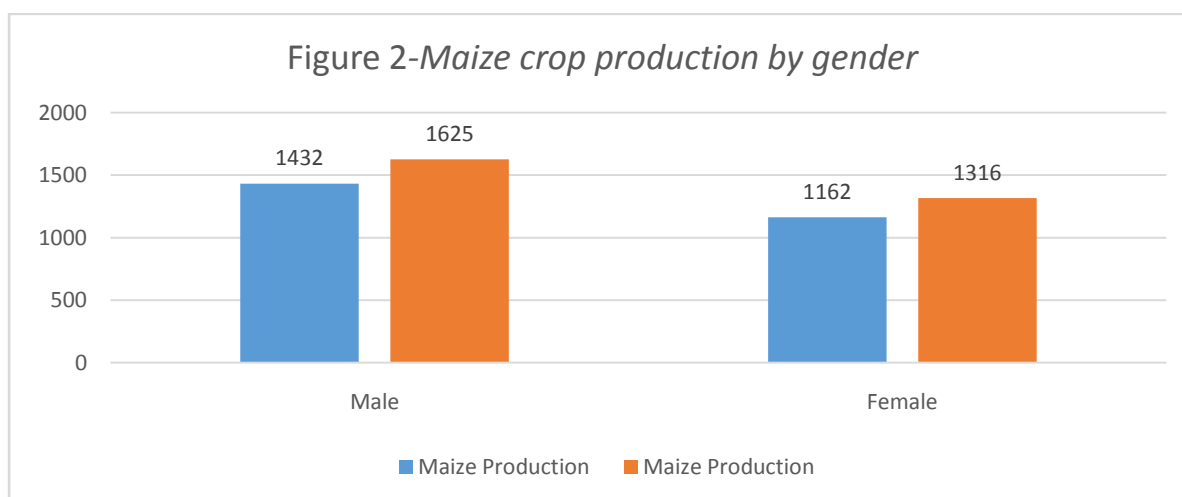
than an APSS. However, if we adopt an integrated household survey approach, the cost of a panel in an integrated household survey is just one quarter of the total cost of the survey.



In terms of value for money, an integrated household surveys collects data on all types of household characteristics, not just households with agricultural holdings (Carletto et al, 2010). This allows for comparison between the different economic sectors and other important disaggregation.



Importantly, an Agricultural Production Estimate Survey supports annual decision with regards to areas under cultivation, production and yield. However, this does not permit the analysis of the relationship between the holding characteristics and the characteristics of the household and its members (MoA, 2016; NSO, 2014).



Agriculture is critical if countries are to achieve the poverty targets set forth by Development Goals within the agreed timeframe (World Bank, 2007). In Malawi, the majority of people suffering from food poverty are rural dwellers and other vulnerable groups, including women, who rely heavily on farm activities (GoM, 2015)). Thus, efforts to fight poverty must focus on rural areas and agriculture, and must be gender-sensitive (Carletto et al, 2010).

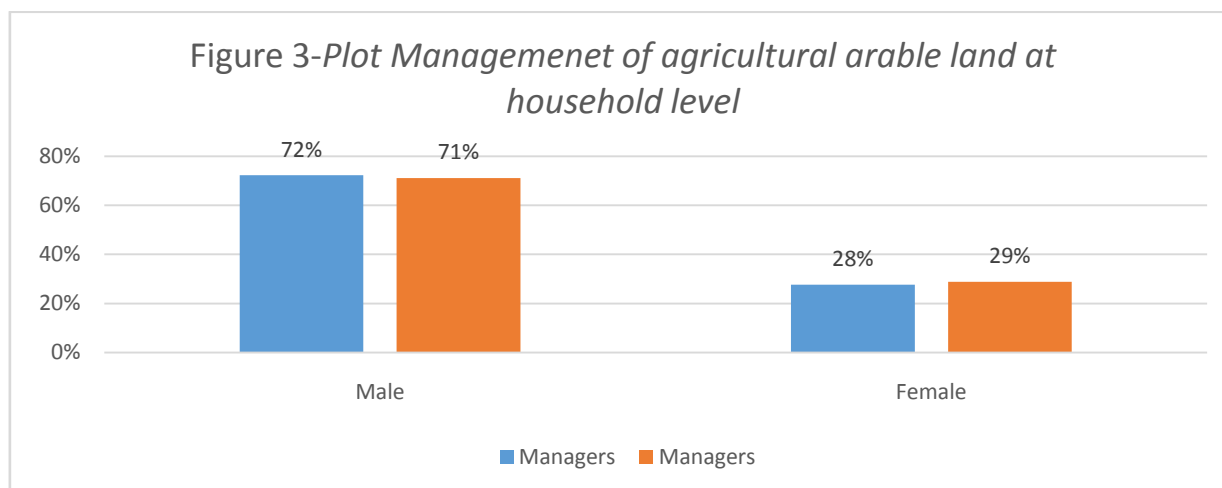
Table 2: *Agricultural characteristics by gender of household heads*

	<i>Female Heads</i>	<i>Male Heads</i>
<i>land area in Ha</i>	0.97	1.15
<i>land value in Mk</i>	48154	53200
<i>land rent in Mk</i>	3369	3700
<i>seed in kg</i>	14	20

Source: Koirola, Mishra and Sitienei (2015)

An Integrated Household Survey approach is implicitly therefore designed to meet this global development goals' call as it collects individual-level data by gender on control of household resources, decision-making and participation in agriculture and off-farm activities (see Figure 2 & 3). This provides researchers with a platform to conduct gender-specific analysis related to agriculture and off farm activities. From an

Integrated Household Survey approach we can assess the changes in household management between genders.



## 5. Advantages

First, research based on cross-sectional data helps to describe snapshots, and with repeated cross-sectional data, to measure broad trends at the macro level. However, they do not assist in understanding the dynamic aspects of the population change at the individual level (Wijesekere, 2009). This limitation portrays lack of capacity to discuss structural change and non-existence of data that could describe across individual and household characteristics. In 2013, a household panel in Malawi provided dynamics over various variables that could be correlated at individual and /or gender level (NSO, 2014).

Second, as the panel surveys have both cross-sectional and time-series elements; each wave is similar to a cross-sectional survey and when data for more than one wave have been collected then it will become a time series. Thus, panel surveys allow cross-sectional analysis of a particular issue of policy relevance as well as providing time-series analysis to assess trends, at the individual level, as opposed to aggregate-level analysis that is possible with cross-sectional data. More importantly, panel surveys provide opportunity to examine transitions between states—flow data to analyze change.

Third, like panel surveys, repeated cross-sectional surveys can collect information on the target population at different points in time, but without the assurance that the subsequent surveys will include the same



population covered in the previous rounds. The advantage of panel surveys, therefore, is that they cover the same persons at different points in time, including split families or households, and add new members 'born into' the sample when they become in-scope according to the criteria used for sample management. This means a panel would still achieve dynamisms in human society (Wijesekere, 2009).

Panel data can further provide valuable information for policy-makers on an issue of policy significance that cross-sectional data cannot provide. At times panel data provide information that is quite different to that derived from cross-sectional analysis. In Malawi, for example, there are large differences in the estimates of the number of farm households between the Ministry of Agriculture (3.4 million farm households) and the National Statistics Office (2.47 million rural households), which in turn affects the accuracy and effectiveness of planning for the subsidized input program (School of Oriental and African Studies, 2008).

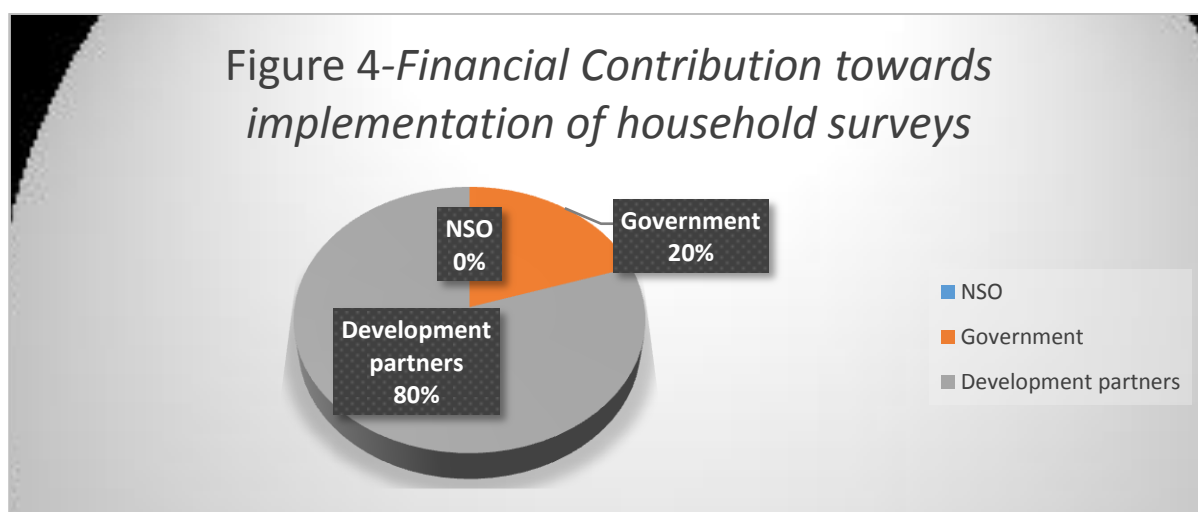
It is well known that cross-sectional data cannot resolve the issue of ambiguity in correlation and, more importantly, cannot confidently demonstrate the direction of causality (Davies, 1994:28). As panel surveys interview the same individual over different points in time, and panel data have a time order of measurement, they are suitable for assessing causality between variables (Hsiao, 2014). Panel data have the capacity to identify stability and change at the individual level. The causal association can be determined when the cause precedes the effect (or outcome of interest). Panel data have the advantage of convincingly identifying this. A detailed discussion of advantages of panel data collection can be found through work of Hsiao, 2014.

The current statistical system provides a fresh cross section of data each year for a new set of (randomly selected) agricultural holders. This format is very good, however, it does not enable policymakers to empirically assess the effectiveness of agricultural interventions over time or investigate medium- to long-term agricultural dynamics. For this, they need a subsample of agricultural holders to be surveyed repeatedly over a number of years (IFPRI, 2011).

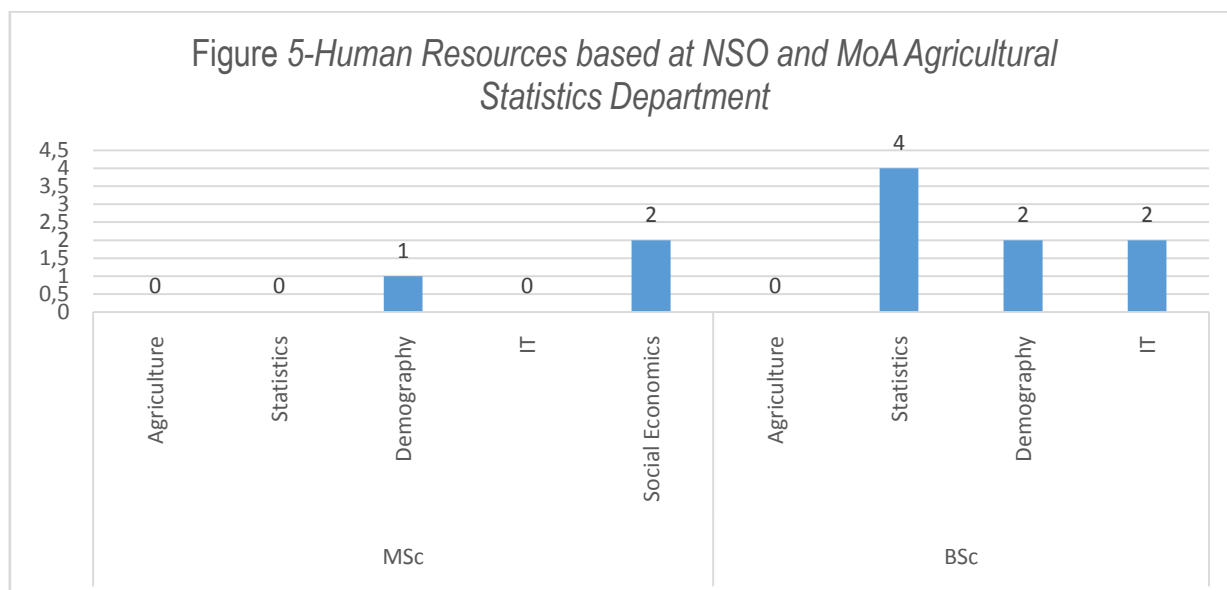
Generally, an Integrated Household Survey improves the quality of household survey data; increases the capacity of statistical institutes to perform household surveys; improves the ability of statistical institutes to analyze household survey data for policy needs and provides policy makers with data that can be used to understand the determinants of observed social and economic outcomes (NSO, 2015).

## 6. Limitations

Panel data collection demands financial and human capacity. Malawi, for which agriculture is a critical source of livelihoods, lacks the financial resources to generate survey data related to agriculture and off-farm activities (see Figure 4). It is noted that most agricultural survey collections are financially supported by development partners. This creates a threat for continued panel data collection in Malawi. For example, in 2015, an Integrated Household Survey failed to take off because of frozen development partner support. Even with sufficient financial resources, the NSO still lacks agricultural driven human resources to collect and analyze such data in a cost-effective and sustainable manner (see Figure 5).



Agricultural data are often collected in institutional isolation, with little coordination across sectors and little analytical value-added beyond the sector. For example, the Ministry of Agriculture collects agricultural data using a production estimate survey while the NSO collects agricultural data using an HIS or National Agricultural Census or Welfare Monitoring Survey. Carletto et al (2010) reports that there is disparity between number of households and yield reported by the MoA and the NSO due to institutional isolated survey designs and implementation. Lastly, there is also lack of analytical capacity that has created a vicious cycle of poor analysis undermining the demand for high-quality data. For example, Figure 5 shows that there is only one personnel with a Masters in Economics and one staff with Masters in Demography. These persons do not have agricultural background.



Author: Calculations

While panel data have numerous advantages over cross-sectional data they too are subject to limitations. The major limitation is the cost. Depending on the methods used to choose the samples, the cost of the first wave of a panel survey is expected to be not much different from that of the cross-sectional survey. Conducting subsequent waves in a Panel survey is expensive because it involves tracking all the original sample members who are in-scope for subsequent surveys.

Panel surveys are considered not as good as cross-sectional surveys at giving cross-sectional estimates if the subsequent waves of panel surveys are not representative of the population or subject to a high level of coverage errors, which are likely to accumulate over time (Deaton et al, 1986). Additionally, the response rates of subsequent waves could also be lower than those observed in cross-sectional surveys, partly because of dropouts from later waves and difficulty in tracing sample members. Obtaining exactly the right balance between overlaps and new households requires more detail on preferences than most investigators could be expected to provide for a multi-purpose survey, but the point remains that a rotating design of some sort will generally be better than either a pure panel or independent cross-sections.

## 7. Conclusion and Recommendations

The paper adopted a conclusion statement made by Deaton, Solon and Ashenfelter (1986) that, while there are genuine difficulties, there are good arguments for collecting panel data. Collecting panel data in Malawi is a sensible enterprise through generating both the dynamics and annual level statistics. However, they are very expensive. It would therefore be important for the panel component to be integrated into cross section household surveys as in the current Integrated Household Survey design so that there is reduction in costs of implementing a fully-fledged panel.

## 8. References

- Ashenfelter O., Deaton A., and Solon G. (1986) Collecting Panel Data in Developing Countries: Does It Make Sense? The World Bank, Washington, D.C. USA.
- Carletto G., Beegle K., Himelein K., Kilic T., Murray S., Oseni M., Scott K., and Steele D. (2010) Improving the Availability, Quality and Policy-Relevance of Agricultural Data: The Living Standards Measurement Study – Integrated Surveys on Agriculture. The World Bank, Washington, D.C. USA.
- FAO/World Bank/Global Donors Platform for Rural Development (2008) Tracking results in agriculture and rural development in less-than-ideal conditions- a sourcebook of indicators for monitoring and evaluation
- Fitzgerald G., and Fitzgibbon M. (2014) Comparative Analysis of Traditional and Digital Data Collection Methods in Social Research in Least Developed Countries. Preprints of the 19 World Congress, The International Federation of Automatic Control, Cape Town, South Africa.
- Hsiao C. (2014) Panel Data Analysis-Advantages and Challenges. Sociedad de Estadística e Investigación Operativa. University of Southern California, USA.
- IFPRI. (2011) Agricultural Production Survey. NW, Washington, D.C. USA.
- International Development Committee. (2013) Post 2015 Development Goals, Eight Report of Session 2012-13. House of Commons, London.
- Koirala K.H., Mishra A.K., Sitienei I. (2015) Farm Productivity and Technical Efficiency of Rural Malawian

- Households: Does Gender Make a Difference? In contributed paper presented at Southern Agricultural Economics Association Annual Meeting. Atlanta, Georgia.
- Malawi Government (GoM). (2012) Malawi Growth and Development Strategy. Ministry of Economic Planning and Development. Lilongwe, Malawi.
- Matchaya G., Nhlengethwa S., and Chilonda P. (2014) Agricultural Sector Performance in Malawi. Regional and Sectoral Economic Studies. South Africa.
- Ministry of Agriculture. (2016). Agricultural Production Estimate Survey Report. Lilongwe, Malawi.
- Moreddu C. (2011). Enhancing Data for Complex Agricultural Establishments. International Expert Workshop Group, Niagara, Ontario, Canada.
- NSO. (2016) Integrated Household Survey 4 (IHS4). Zomba: National Statistical Office of Malawi.
- NSO. (2015). Welfare Monitoring Survey. Zomba: National Statistical Office of Malawi.
- NSO. (2014) Integrated Household Panel Survey (IHPS). Zomba: National Statistical Office of Malawi.
- NSO. (2012) Integrated Household Survey 3 (IHS3). Zomba: National Statistical Office of Malawi.
- Paris21 (2002) "Final Report: Seminar on a New Partnership to Strengthen Agricultural and Rural Statistics in Africa for Poverty Reduction and Food Security". Task Team Food, Agriculture and Rural Statistics. Paris 21, Paris, France.
- Wijsekere W. (2009) The role of household panel surveys in evidence-based policy. a revised version of the paper presented to the Australasian Evaluation Society International Evaluation Conference, 2009, 31 August–4 September 2009, Canberra Australian Capital Territory.
- World Bank. (2010) Source of Agricultural Data. The World Bank Institute, Washington, D.C. USA.
- World Bank (2008) "World Development Report 2008: Agriculture for Devt." Washington, D.C. USA.