STEG WORKING PAPER

INCOME DIVERSIFICATION AND HOUSEHOLD WELFARE IN TANZANIA 1991 – 2018: EVIDENCE FROM A SYNTHETIC PANEL

Vincent Leyaro and Josephat Hongoli

FEBRUARY 2024
STEG WP085
SRG 1457

Income Diversification and Household Welfare in Tanzania 1991 – 2018: Evidence from a Synthetic Panel

Vincent Leyaro and Josephat Hongoli
University of Dar es Salaam School of Economics

Abstract

For over a quarter century (1991 – 2018) Tanzania has been experimenting with structural adjustment reforms that have seen impressive macroeconomic performance and sustained growth in one hand but a sluggish poverty reduction in the other hand. Using four rounds of successive Tanzania National Household Budget Surveys (1991/92, 2000/01, 2011/12 and 2017/18), we construct synthetic (pseudo) panel data to investigate the role of sources of income (labour diversification) on household welfare (consumption expenditure). The analysis shows that over time the share of households with farm income sources has declined with an increasing share of households with non-farm income (though at a slower pace); which is primary evidence of the signs of the structural transformation in the economy. Consistent with the 'diversification as means of accumulation' hypothesis and in line with previous studies in the region and in Tanzania, the study findings show that household income diversification significantly improves household welfare. However, there are substantial differences in the impacts, with a wider impact for rural than urban households. Our results are robust to alternative measures of household income diversification, the share of workers in farm and non-farm activity. Clearly, if the government is serious about addressing poverty in general and rural poverty, it has therefore to enhance rural infrastructure that enhances the returns of both on-farm and off-farm activities.

JEL Classification: C13, I39, O12, O55
Keywords: Synthetic panel, income diversification, household welfare, Tanzania

Acknowledgments
The research is undertaken as a part pf of the project *Long-Run Analysis of Income Diversification and Household Welfare in Tanzania 1991–2021* funded by CEPR (STEG_LOA_1457_Leyaro) under the Small Research Grants (SRGs). We are grateful to CEPR for the funding.

1. Introduction

Following over a quarter century of experimenting with economic reforms, Tanzania presents a notable conundrum for those following growth and development in Africa. On the one hand, the country has reported impressive macroeconomic performance with sustained economic growth of more than 6 percent since 1995 from an average of 3.2 percent in 1985 – 1995, and 7 percent since 2000 up until the intrusion of Covid19 pandemic in early 2020 (Masenya, C., *et al.,* 2018, NBS, 2019). As a result of these reforms, the recovery of most economic sectors began in the mid-1990s and early 2000s, reflecting a lag effect of the policy reforms. During this period, economic growth was strongly influenced by two major activities, namely – services, industry, and construction; as services activities contributed to about 45 percent of gross domestic product (GDP) growth, while industry and construction activities contributed to about 30 percent. Besides real GDP growth rates, other macroeconomic indicators such as double-digit inflation, negative real interest rates, huge fiscal deficits, an overvalued Tanzanian shilling, and a thriving parallel foreign exchange market – in addition to the rationing of foreign exchange, improved substantially during this period (Masenya, C., *et al.,* 2018; NBS, 2019).

While several important non-monetary welfare indicators such as infant mortality, life expectancy, asset ownership, school attendance, and availability of public health services, to mention some, have shown a positive trend (Arndt, C., *et al*., 2017); on the other hand, despite conditions conducive to increases in consumption across a broad swath of the population, consumption poverty reduction and income inequality have not followed suit (NBS, 2019). According to the Tanzania House Budget Surveys (HBS), poverty has fallen marginally. Between 1991 and 2007, poverty fell only by about 5.2 percentage points, from 38.6 percent in 1991/92 to 33.4 percent in 2007; and further by about 5.2 percentage points, from 33.4 percent in 2007 to 28.2 percent in 2011/12; that about 10.4 percentage point for the past 20 years (equivalent to 0.52 percentage points decrease each year). Poverty has fallen even more marginally between 2011/12 and 2017/18 by 1.8 percentage points from 28.2 percent in 2011/12 to 26.4 percent in 2017/18. At the same time, consumption growth became less pro-poor, and inequality as measured by the Gini coefficient has increased; the Gini coefficient went up from 38.5 in 2007 to 39.5 in 2018 (NBS, 2019; World Bank, 2020).

Thus, this apparent contradiction – solid per capital GDP growth, slight increase in inequality, and yet less consumption poverty reduction than desired over a quarter century (i.e. since 1990) has generated considerable debate; which has raised questions: is it due to the underlying determinants of economic growth and poverty or is it due to how growth, poverty, inequality, and other welfare indicators are measured, reconciled, and compared across space and time?

Furthermore, trends in poverty rates differ according to the sector in which the household or the head of the household is employed or geographical location, where studies have shown that there is considerably higher poverty amongst agricultural (especially food crops) households compared to those in manufacturing or services; and those in rural areas compared to those in urban or sub-urban areas. However, sector-level analysis is limited because 'it does not consider the composition of household income', except perhaps for some consideration of the main sector of household head economic activity (Kappel *et al.*, 2005). To capture the broader monetary complexities of poverty and gauge any signs of structural transformation (mobility away from farming to non-farming activities and increase in household productivity), we are considering five sources of household income in this analysis: self-employment in agriculture (farm income), self-employment in non-agricultural activities (off-farm income), wage income (public or private), remittances and rental income.[1] A source is allocated to households if at least one member engages in the activity and if remittances are received; income diversification is a count of the number of sources the household receives.

Existing empirical literature abound have established a positive relationship between income diversification (especially households' engagement in non-farm activities) and household welfare in most low-income countries, including those in sub-Saharan Africa (SSA) (Reardon *et al.,* 2007; Davis *et al.,* 2010, 2017; Loison 2015; Davis *et al.,* 2017; Asfaw *et al.,* 2019; Van den Broeck and Kilic 2019). Other studies have emphasized the important role of income diversification in enhancing economic growth (Carswell, 2002; Niehof, 2004; Anderson, 2016,); its ability to reduce overall inequality (including ensuring gender equity) among rural populations (Winters et al., 2010) and there are those households that diversify their income to address food insecurity and

[1]The income categorization here is as adopted from and collected in the national HBS. While we would like to have a separate wage income for agriculture and non-agriculture, given that in the literature those two are very distinct income types, that has been difficult as the way data has been collected does not allow for that. Even though, the added advantage of these categorization is that it differentiates between rent and remittances, that widen the analysis.

nutrition, especially in rural areas (Atuoye *et al.,* 2019; Etea *et al.,* 2019; Kidane & Zegeye, 2019; Salifu, 2019; Kassegn & Endris, 2021), and invest in innovations to sustain agriculture (Barbieri & Mahoney, 2009; Asravor, 2017).

Two studies on income diversification and household welfare have been done as far as Tanzania is concerned. One study is by Khan & Morrissey (2020) who used three waves of Tanzanian National Panel Surveys (NPS) to construct a panel to explore the effect of income diversification on household welfare measured in terms of food consumption. Another study by Dimova, *et al.* ,(2015, 2021) looks at what drives livelihood diversification among predominantly rural households in developing countries and how can welfare-enhancing patterns be established and sustained in the long run, using the Kagera region in Tanzania as their case; testing whether income diversification is a means of survival or a means of accumulation.

Building on these studies, in this research we first construct a (pseudo) panel of national representative households using successive four waves of National Household Budget Surveys for Tanzania (1991/92, 2001/02, 2011/12, and 2017/18) to investigate the drivers of household welfare (consumption expenditure) over a quarter century to help to understand the dynamics of household welfare, especially the role of sources of income (labour diversification). Unlike Khan & Morrissey 2020 that have used the three waves of NPS (2008/09, 2010/11, and 2012/12) that have looked at the same issue in the short run that allows only 5 years, we have looked at the same thing using a long run analysis over a quarter century, 1991 – 2018. Furthermore, normal panel data have several drawbacks that include non-response and attrition that can cause sampling and selection issues in addition to measurement error, in addition, it is difficult to get nationally representative panel data over a long period. Thus, using individuals and households repeated cross-sections data that suffers less from non-response, attrition, and measurement error; and so less likely to suffer from selection issues. Dimova, et al., (2015, 2021) study that is only based on one region (i.e. Kagera) that is not national representative, the National Household Budget Surveys for Tanzania (1991/92, 2001/02, 2011/12 and 2017/18) used in this study are nationally representative over a larger population and covers a long period. In line with Khan & Morrissey (2019) for the case of Uganda, and 2020 for the case of Tanzania, we first present estimates based on data from the individual surveys pooled, and thereafter to capture dynamics and allow for endogeneity, we provide estimates based on a pseudo-panel fixed effect.

The remaining of the paper is organized as follows; Section 2 briefly reviews the literature on diversification and welfare. Section 3 provides a conceptual framework and empirical strategy deployed, while Section 4 describes the data as well as how income diversification is measured, and the pseudo panel constructed. Discussion of the key findings of the study is provided in Section 5 while Section 6 presents a summary of the study and its policy implications.

2. Diversification and Household Welfare: A Review of Literature

In most developing countries poverty trap is associated with engaging in agricultural activities and residing in rural areas. As a result, the growth and development of the agricultural sector is perceived as a route out of poverty for most people living in rural areas (Ravallion & Datt, 1996; World Bank, 2008; FAO, 2011). Household income diversification, a process through which people develop an increasingly diverse portfolio of activities and assets, both on-farm and off-farm activities, to improve their living standards is seen as one of the pathways to poverty reduction. This is because the diversification of sources of income (including crop diversification) increases total income and managing uncertainties among farmers (Ellis, 2003; FAO, 2011; Njeru, M., 2013; Asfaw *et al.,* 2019; Van de Broeck and Kilic, 2019).

Theoretically, household income diversification is underpinned by agricultural households' model of optimal labour allocation in which households maximize their utility over consumption and leisure time subject to time and budget constraints (Mduma & Wobst, 2005; Wang *et al.,* 2007; Weltin *et al.,* 2017). The households' characteristics in rural areas settings of most low-income countries is that of extreme variability due to factors such as weather variation[2], the incidence of disease and pest infestation, fire, crops price variability, and shocks at global markets that have led to unpredictable and severe income functions for most households. While both risk-pooling and the use of savings and credit transactions (i.e. ex-post means of smoothing consumption) can be perceived as the surest way in which poor households can deploy in the face of large and unpredictable shocks in an attempt to smooth out their consumption, hence little incentives for diversification; there are several reasons why both of them are not attainable, and above all credit markets are not available neither perfect (Dimova, R., *et al.,* 2015, 2021).

[2] Farmers tend to diversify their income with non-farm wage jobs in response to climate volatility (rainfall shocks), leads to additional uncertainty for rural households. The strategy is to diversify into other sectors less likely to be affected by such shocks (Chuang, Yating, 2019).

As a result, to reduce household income fluctuations, the most common method of income – smoothing adopted by poorest households is to diversify the sources of their income (Ellis 2000a; Reardon 1992, 1997; Reardon *et al.,* 1992). Since poorer households have fewer assets that can be sold to smooth consumption and have less access to credit facilities or formal and informal insurance mechanisms, their diversification, therefore, is a coping response to shocks; this is the behaviour that underpins the diversification as survival mechanism. Thus, under this approach, income diversification is mostly undertaken by poor households as a mechanism to smooth consumption in the face of high-income volatility and out of sheer desperation (Barrett *et al.,* 2001; Dercon 2002).

On the other hand, another explanation for household income diversification assumes economies of scope in production, which tend to favour diversification as a means of profit maximization; along with entry barriers to high-return economic activities (Barrett *et al.,* 2001). Since the entry into many activities, both within and outside agriculture, need initial capital or access to land, it is only richer households that can diversify while poorer households are unable to do so; hence diversification in this sense is mostly driven by accumulation motives and thus confined to richer households (Barrett and Swallow 2005). Following this, there are clear theoretical arguments on the causal origins of income diversification, both diversification as survival and diversification as accumulation hypotheses (Dimova, R., *et al.,* 2015, 2021).

There is thus a large empirical literature on whether housed diversification is as the result of a survival strategy or accumulation strategy. The role of income diversification as a form of risk reduction and income stabilization that ensures consistent consumption patterns is noted (Reardon *et al.,* 1992; Reardon, 2001; Eakin, 2005; Anderson and Deshingkar, 2005; Menon, 2009; Biggs, Gupta, Saikia John, 2018). More recently for the case of Tanzania, Khan and Morrissey (2020) find that women in poorer rural households are more likely to enter non-agricultural self-employment, suggesting that among this income group, these activities are tolerated out of necessity.

Income diversification also engenders improvements in quality of life, wealth accumulation, food security for rural households; and as a result, there is strong evidence that income diversification has acted as a means of accumulation in Sub-Saharan Africa, with households with larger holdings of land or access to capital more able to move into high-return activities such as livestock rearing

or non- farm employment (Abdulai and CroleRees, 2001; Barrett *et al.,* 2001; Block and Webb 2001; Dercon 1998; De Weerdt 2010; Mckay, 2016). The benefits of income diversification across industries have also been examined by various works (Ansoms & McKay, 2010; Bird Shepherd, 2003; Ellis and Bahiigwa, 2003; Ellis and Mdoe, 2003; Ellis *et al.,* 2003; Soltani *et al.,* 2012, Yobe *et al.,* 2019).

Furthermore, economic growth studies have emphasized the critical role of income diversification in poverty reduction (CroleRees, 2001; Bezu *et al.,* 2012; Block & 2001; Carswell, 2002; Ellis, 1998, 2000a; Anderson, 2016, Niehof, 2004; Smith *et al.,* 2001). The merits of income diversification are also seen in its ability to reduce overall inequality as well as ensuring gender equity among rural populations (Winters *et al.,* 2010). In terms of poverty reduction, income diversification may occur as a voluntary or involuntary response to the crisis, which can either diminish or accentuate rural inequality depending on the level of access to capital inputs crucial for income diversification (Seratna, 2014). A couple of studies have also looked these issues for the case of East Africa and Tanzania.

Khan and Morrissey (2023) for the case of Uganda have found that diversification became increasingly beneficial for welfare over time in rural areas, particularly for male-headed households, but not for female-headed households that diversified into agricultural wage employment. Diversification was also important for the livelihoods of urban households, but large differences across male and female-headed households reflecting differentials in the returns to non-agricultural employment. Remittances were associated with increasing welfare in the 2000s for all households, although the proportion of households receiving remittances has been declining over time.

Coming for the case of Tanzania, using the three waves of Tanzanian National Panel Surveys (2008/09, 2010/11, and 2012/13), Khan and Morrissey (2020) explore the effect of income diversification on household welfare measured in terms of food consumption and fond that increasing diversification is associated with higher welfare, but there are differences by gender and activity type. Non-agricultural wage employment is beneficial, irrespective of gender, and has had relatively high growth. Non-agricultural self-employment is a welfare-increasing diversification strategy, especially in rural areas (although females benefit more than males in urban areas), but growth has been slow. Furthermore, Dimova *et al.,* (2021) results support the accumulation

hypothesis: richer households engage in more income diversification than poorer households, and the greater diversification of better-off households that was observed in the 1990s persists in 2004. At the same time, households that were originally poorer are found to experience higher incomes by diversifying into off-farm self-employment activities. Factors that explain these improvements include access to a daily market and public transport.

3. Empirical Strategy and Estimation

To estimate the effects of income diversification on household welfare, the overarching objective of this study, we follow a similar specification to that of Khan & Morrissey (2019) for Uganda and Khan & Morrissey (2020) for Tanzania. That, household consumption (welfare) is primarily determined by their earning activities, which are then influenced by assets/capital and household access to different employment opportunities. In the genuine panel data that follows the same household overtime, the specification is as follows:

$$\ln(Cons_{it}) = \alpha + \gamma ID_{it} + X'_{it}\theta + h_{rt} + \tau + \psi_i + \xi_{it} \qquad (1)$$

where $Cons$ is our outcome variable, adult equivalent household consumption level, $\boldsymbol{i}$ denote household and $\boldsymbol{t}$ – survey year (time), ID is the household income diversification measured as the count or number of income sources for the household, X is a vector of time-varying household controls that influence the income diversification choice and household welfare. These include the head's main labor economic activity, household dependency ratio, location dummy (1 if from rural), head's age and age-squared, gender, schooling level, and marital status. h_{rt} denote dummy interaction terms of region of residence and survey year to account for regional differences in prices or inflation across the survey year and τ captures time-fixed effects (with a dummy for the survey years) to capture any circular changes in household welfare. ψ_i captures time-invariant unobservable household heterogeneity that is related to diversification choices such as the members' ability, attitude towards risk, social networks, and other time-invariant factors. $\boldsymbol{\xi_{it}}$ is the mean zero error term capturing the influence of other unobservable factors. $\boldsymbol{\gamma}$ is the coefficient of interest showing the effects of income diversification on household welfare.

One challenge for estimating specification (1) is the endogeneity due to the possible correlation between diversification with time-variant and invariant unobservable household characteristics.

Several empirical studies '*on why households diversify their income sources*' show that household income diversification behaviors are determined by several factors, summarized as either push or pull factors. The push factor includes factors such as risk reduction, response to diminishing factors return, response to land constraints, reaction to the crisis, and liquidity constraints (as means of survival strategy). On the other hand, the pull factors include the need for complementarity in income sources or as means of accumulation (Beegle *et al.,* 2011; Dimova & Sen, 2010; Loison 2015). Normally, the incentives for income diversification are not observable and consequently may lead to self-selection problems with poor households diversifying more (in case of survival strategy) while richer households engaging in more activity (in case of accumulation hypothesis).

Under omitted unobservable factors, a simple pooled OLS estimation of specification (1) might lead to a biased estimate of the coefficient of interest. The γ coefficient may thus only show the correlations between ID and consumption that is biased due to the possible endogeneity of income diversification. In the genuine panel data that traces the same household over time, the endogeneity due to omitted unobservable can be solved by including household or individual fixed effects to wipe out the estimation bias. However, in the absence of actual panel data, like our case in this study, Deaton (1985) suggests the use of pseudo or synthetic panel fixed effects constructed from a series of multiple cross-section survey data to address the endogeneity problem. Thus, in addition to estimating pooled regression of specification (1), we construct pseudo-panel data and estimate the fixed effects model to address the possible endogeneity of income diversification. Following Deaton (1985), we construct pseudo-panel (synthetic panels) data making use of the rich repeated cross-sectional data to deal with the bias. We construct a synthetic panel by grouping households into cohorts based on time-invariant shared common characteristics. In the constructed panel, we aggregate households into cohorts based on unchanging common characteristics that are observed in each survey year as suggested by Deaton (1985) and Verbeek & Nijman, (1992). Then, we treat the cohort as a cross-section unit and averages of the variables of interest within each cohort-year cell as observations (a detailed discussion on pseudo panel data construction is in subsection 4.3).

The pseudo-fixed effects specification is as follows:

$$\overline{\ln(y_{ct})} = \alpha + \gamma \overline{ID}_{ct} + \bar{X}'_{ct}\theta + \psi_{ct} + \tau + \bar{\xi}_{ct} \qquad (2)$$

where *c*-denote cohort and *t*-survey year (time). The dependent variable is the average of the natural logarithm of household monthly consumption per adult equivalent of the cohort at a time *t*. The rest of the control variables including the income diversification index are as defined before, as cohort time averages. The pseudo-panel data are estimated with the assumption that the fixed effects (ψ_c) are time varying. However, following the time-varying fixed effects resulting from measurement error and aggregation bias, the conventional within-estimation model might not solve the bias problem. Such that:

$$\overline{\ln(y_{ct})} = \alpha + \gamma\overline{ID}_{ct} + \bar{X}'_{ct}\theta + \psi_c + \tau + \bar{\xi}_{ct} \quad (3)$$

It is worth pointing out that the synthetic/pseudo panel fixed effect method is not without caveats. Just like many other approaches, there are issues arising with the pseudo-panel method. The most notable ones are the possible sampling errors and aggregation bias (Deaton (1985), Verbeek & Nijman (1992)). The sampling error occurs when cohort sample means are not representative of the underlying cohort population means due to variation in cohort compositions over time. Aggregation bias, on the other hand, occurs due to loss of variation when using cohort-level data. Khan (2018) show that the estimation bias resulting from aggregation bias can be substantially high that it can negate any potential benefit of using pseudo-panel over a simple OLS method. Verbeek & Nijman (1992) argues that the sampling error can be disregarded with large enough cohorts. Heteroscedasticity of error term due to substantial variation in the cell size (number of observations per cohort) is another caveat in estimating pseudo-panel. The heteroscedastic error term may lead to biased standard errors. Deaton (1985) and Verbeek & Nijman (1992) suggest that an efficient estimator can be achieved by weighting each cell with the square root of the number of observations. Given these caveats and with large enough observations, the sampling error is disregarded. To correct the heterogeneity of standard error due to variation in cohort-year size (cell size), the weighted within estimator using the square root of cell size is estimated, as suggested by Deaton (1985).

4. Data, Income Diversification, and Pseudo-panel Construction

4.1 Data

We use four series of Household Budget Surveys (HBSs) collected by the Tanzania National Bureau of Statistics (NBS): 1991/92 (4,823 households), 20001/02 (22,129 households), 2011/12 (10,182 households) and 2017/18 (9,418 households). HBSs are nationally representative surveys covering the population residing in private households in urban and rural Tanzania Mainland. The HBS adopts a two-stage cluster sample design. In the first stage, enumeration areas (primary sampling units – PSUs) are selected from Population and Housing Census National Master Sample Framework, and in the second stage, households are systematically sampled from PSUs. The first round of scientific HBSs that represent urban and rural areas was conducted in 1991. Since then, NBS has completed five rounds of scientific HBS including the 20001/02, 2006/7, 2011/12, and 2017/18 HBS. The HBS collects information both at the individual level and household level. The individual information collected includes demographics; migration; education; literacy; labor market indicators; non-farm household businesses; and individual non-wage income. HBSs also collect household data on consumption and expenditure. Table 1 presents descriptive statistics.

Our key variables of interest throughout all surveys include household welfare and income sources within the household. To measure welfare, we use household-level monthly consumption per adult equivalent constructed by diving household monthly consumption level by adult equivalent. All rounds of HBSs collected information on household consumption levels for 28 days except for 2017/18 which collected for 14 days. We use this information to calculate household monthly consumption. We also use household head demographic information such as gender, marital status, and education, as well as household-level information such as dependency ratio and location. To avoid overemphasis, we restrict our analysis to households with age from 20 to 60 years born in 1931 to 1998.

Table 1: Descriptive Statistics

	1991/92		2000/01		2011/12		2017/18	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age in years	39.07	9.89	39.26	10.24	39.33	10.06	41.09	10.18
Female	0.14	0.35	0.23	0.42	0.23	0.42	0.24	0.43
Rural (0/1)	0.44	0.50	0.33	0.47	0.38	0.49	0.69	0.46
Urban (0/1)	0.56	0.50	0.67	0.47	0.62	0.49	0.31	0.46
Marital Status								
Married	0.80	0.40	0.74	0.44	0.71	0.45	0.76	0.43
Never Married	0.12	0.32	0.12	0.33	0.10	0.30	0.06	0.24
Divorced	0.05	0.22	0.08	0.27	0.11	0.31	0.10	0.30
Widow	0.03	0.17	0.06	0.24	0.08	0.27	0.08	0.27
Education Level								
No Education	0.13	0.34	0.13	0.34	0.10	0.31	0.16	0.36
Primary	0.65	0.48	0.66	0.47	0.66	0.47	0.67	0.47
Secondary	0.14	0.35	0.15	0.36	0.17	0.37	0.12	0.33
Diploma	0.03	0.18	0.03	0.18	0.03	0.17	0.03	0.16
Higher Education	0.01	0.10	0.01	0.11	0.02	0.15	0.02	0.13
Dependency ratio	0.36	0.24	0.36	0.24	0.35	0.24	0.40	0.24
Observations	4240		18615		8546		7501	

Notes: The table reports the summary statistics for the key explanatory variables. All demographic variables such as age, gender, marital status, and education is related to household heads. The dependency ratio is defined as the ratio of household members aged below 18 and above 65 to the total household size.

4.2 Income Diversification Measures

Different approaches have been widely employed in various empirical literature to measure household income diversification (Dimova *et al.,* 2010, 2021; Khan & Morrissey, 2020, 2023). These include participation in off-farm income (dummy), the share of off-farm income (or non-farm labor) on total income, and the Herfindahl index of household income concentration. The other commonly used approach is the use of discrete indicators to show the counts or income portfolios for different types of household income sources (Beegle *et a*l., 2011; Dimova & Sen, 2010; Khan & Morrissey, 2020). However, each approach has its strengths and weakness. Due to the nature of our data, we use the latter approach, the portfolio or counts of income sources within the household as our measure of income diversification. Thus, to measure income diversification, we use both individual-level information on labor market activities and household-level information, including non-farm household businesses and receipt of remittances and rental

income. The sources of income also need to be comparable across surveys, both for urban and rural households to facilitate the national-level analysis (Khan & Morrissey, 2020, 2023).

However, it is difficult to create measures of income sources that remain consistent throughout four rounds of HBSs due to changing nature of the survey questionnaires on household income sources and labor market information. We therefore adopt Khan & Morrissey (2023) strategy to measure income sources that can be tracked across all rounds. The survey questions include individual-level information on primary and secondary labor market activities. The labor market activities are divided into four main categories that can be consistently tracked across all four HBSs rounds; whether an individual in the household is an employee (hold a paid job), self-employed with employees (non-agriculture), self-employed without employees (non-agriculture) or working on own farm. The information on labor market activities is available for the primary and secondary labor market activities. The other income sources that we can track include remittances the household may receive from members who have migrated away and whether the household receives rental income(s). Table 2 reports the income sources, their definition as used in this study and the measurement.

Table 2: Income Sources, Description, and Measurement

Sn	Source	Description	Measurement
1	Farming (agriculture) income	Farming income involves all income generated from agriculture activity with at least one member being self-employed in agriculture; crop production or livestock rearing and/or fishing activity	1 if at least one member of the household is self-employed in agriculture or fishing and 0 otherwise
2	Non-farm income	A non-farm income involves all incomes from self-employment in non-agriculture activity (except for wage activity) including non-agriculture self-employment with or without employees.	1 if at least one member of the household has non-farm income and 0 otherwise
3	Wage income	Wage income involves all income derived from the employer-employee relationship. An activity where at least one member engages in or takes a paid job as the primary activity regardless of the sector of employment (both private and public)	1 if at least one member of the household has non-farm income and 0 otherwise
4	Rental Income	Rental incomes are those from renting household assets such as agricultural machines, houses, and other assets	1 if the household has rental income and 0 otherwise
5	Remittance income	Remittance income is if the household receives remittance from migrated members living outside the country (abroad)	1 if the household receives remittance and 0 otherwise

Notes: The income from individual labour market activity (farm, non-farm, and wage) is defined for only adult members (aged above 18). The rental and remittance income sources are based on household-level information.

The five potential sources of household income that we consider in the analysis include agricultural (farm) income, wage income, non-agricultural self-employment income, remittance, and rental income. Agriculture (farm) income includes all self-employed activities in agriculture and fisheries. All earnings outside agriculture (except for employees) are categorized as non-agricultural self-employment including non-agriculture self-employment with or without employees. Wage income sources are those from employed individuals including private and public employment. However, we are not able to distinguish the sector of employment (such as agriculture and non-agriculture) for wage employment as in Khan & Morrissey (2023) due to inconsistency in the question that traces the sector of employment across the surveys. We thus define wages from the private sector and the public sector as one income source. Remittances and rental income are the other sources of income. To measure the two income sources, we use a dummy for whether the households receive remittance from other households/members residing outside the country and if the household receives rental income for renting part of their assets such

as farms, agricultural machinery, houses, or other assets. However, due to the small proportional of household with remittance and rental income, we combine the two sources of incomes as one income source. For the robustness check, we also use the share of non-farm workers in the households as a measure of household diversification away from agriculture.

After defining the four income sources[3], we finally construct the household income diversification index (ID). Income diversification can be measured in various ways including the share of nonfarm income in the total income, counts of income sources (or categories) of types of income, and the Herfindahl index measure of shares of multiple sources (Khan & Morrissey, 2020, 2023). Only the first three rounds (1991/92, 2006/07, 2011/12) collected information on the amount of income by sources therefore the use of income share and Herfindahl index are not feasible. We adopt two strategies to measure income diversification. First, we use the simple count of income sources (with each income source count once within the household) that the household receives as a measure of income diversification. As Khan & Morrissey (2020) noted, the main weakness of this measure of income diversity is that it may be underestimated as it misses multiple activities by individuals. A household-based simple count function for income diversification can be specified as follows:

$$ID = \sum_{i=j,\ h}^{N\ (4)} I(\chi^j > 0) \qquad (4)$$

where ID is the income diversification score of the household. j denote the income source and N is the total number of income sources under consideration (4 in our case). χ^j is the amount of income that households receive from income source j and $I(.)$ is an indicator function, 1 if the household have a positive income from the source j and 0 otherwise. As the motive for income diversification may differ between rural and urban households, we present a separate analysis for the rural and urban households.

[3] with remittance and rental income being combined as one income source.

Table 3: Distribution of Household by Income Sources

	Rural				Urban			
	1991/92	2000/01	2011/12	2017/18	1991/92	2000/01	2011/12	2017/18
Panel A: Distribution of household by income sources								
Farm Income	0.95	0.86	0.90	0.81	0.28	0.37	0.16	0.24
	(0.22)	(0.35)	(0.30)	(0.39)	(0.45)	(0.48)	(0.36)	(0.43)
Non-Farm Income	0.36	0.32	0.27	0.47	0.90	0.78	0.92	0.90
	(0.48)	(0.47)	(0.44)	(0.50)	(0.30)	(0.41)	(0.28)	(0.30)
Wage Employee	0.20	0.13	0.10	0.10	0.61	0.38	0.50	0.37
	(0.40)	(0.34)	(0.30)	(0.30)	(0.49)	(0.49)	(0.50)	(0.48)
NA-Self Employment	0.18	0.17	0.14	0.41	0.49	0.47	0.52	0.79
	(0.38)	(0.38)	(0.35)	(0.49)	(0.50)	(0.50)	(0.50)	(0.41)
Remittance Income	0.01	0.01	0.04	0.01	0.01	0.03	0.04	0.01
	(0.10)	(0.11)	(0.19)	(0.07)	(0.09)	(0.18)	(0.20)	(0.08)
Income from rent	0.03	0.04	0.03	0.02	0.05	0.10	0.10	0.07
	(0.17)	(0.20)	(0.17)	(0.15)	(0.22)	(0.30)	(0.30)	(0.26)
Panel B: Average number of workers in each income sources								
Farm Income	3.05	1.69	1.86	1.53	0.53	0.62	0.28	0.36
	(2.32)	(1.21)	(1.17)	(1.17)	(1.11)	(0.98)	(0.73)	(0.76)
Non-Farm Income	0.52	0.38	0.30	0.71	1.66	1.10	1.33	1.83
	(0.95)	(0.73)	(0.63)	(0.95)	(1.34)	(0.96)	(0.86)	(1.19)
Wage Employee	0.22	0.15	0.11	0.11	0.83	0.46	0.62	0.45
	(0.49)	(0.42)	(0.37)	(0.36)	(0.88)	(0.66)	(0.75)	(0.67)
NA-Self Employment	0.30	0.23	0.18	0.60	0.83	0.64	0.70	1.38
	(0.79)	(0.59)	(0.50)	(0.87)	(1.18)	(0.84)	(0.82)	(1.10)
Panel C: Share of workers to adult members in each income sources								
Farm income	1.10	0.73	0.82	0.69	0.21	0.28	0.12	0.17
	(0.60)	(0.37)	(0.33)	(0.39)	(0.43)	(0.40)	(0.29)	(0.33)
Non-Farm Income	0.32	0.23	0.20	0.38	1.11	0.74	0.96	1.06
	(0.57)	(0.45)	(0.44)	(0.51)	(0.72)	(0.61)	(0.59)	(0.54)
Wage Employee	0.10	0.07	0.06	0.06	0.38	0.22	0.31	0.22
	(0.23)	(0.20)	(0.19)	(0.19)	(0.39)	(0.33)	(0.37)	(0.33)
NA-Self Employment	0.22	0.16	0.14	0.32	0.73	0.51	0.65	0.84
	(0.41)	(0.30)	(0.30)	(0.40)	(0.47)	(0.38)	(0.34)	(0.34)
Panel D: Income diversification index								
ID	1.36	1.22	1.21	1.35	1.43	1.35	1.31	1.47
	(0.58)	(0.46)	(0.46)	(0.53)	(0.61)	(0.57)	(0.54)	(0.61)
ID (off-farm)	0.41	0.36	0.31	0.54	1.16	0.98	1.15	1.23
	(0.60)	(0.57)	(0.55)	(0.63)	(0.61)	(0.67)	(0.59)	(0.66)
ID - RR	1.32	1.16	1.14	1.32	1.37	1.22	1.18	1.39
	(0.54)	(0.39)	(0.36)	(0.50)	(0.55)	(0.47)	(0.42)	(0.53)
Observations	1851	6161	3262	5181	2389	12454	5284	2320

Notes: The table reports the distribution of households by income sources for each survey year for rural and urban households. The averages do not sum to 100 as one household may have more than one income source. Panel A report the distribution of household in each income source, panel A shows the average number of workers in each income source type, and panel C report the share of workers to total household adult members (those aged above 18 at the time of the survey) and panel D show the income diversification index (average counts of income source). The means and standard deviations (in parenthesis) are reported.

Table 3 shows the distribution of each income source for each survey estimated separately for rural and urban households. *Panel A* presents the distribution of households for each income source[4]. As highlighted, more than 95 percent of rural households depended on farm income in 1991 and 81percent in 2017. Implying that over time, the share of households with farming income is declining for both rural and urban households. On the other hand, the non-farm income including wage income and non-agriculture self-employment income are relatively more important for those in urban than rural. While wage income has relatively declined over time in rural, the overall contribution of non-agricultural self-employment activity is on the rise, from 18 percent in 1991 to 41 percent in 2017. Similarly, for urban households, the relative importance of off-farm income has slightly increased from 88 percent in 1991 to 90 percent in 2017. While wage income has declined, the relative increase in non-farm activity has been triggered by an increase in self-employment in non-agricultural activity. Overall, the declining trends for agricultural income and rising of non-farm income can be regarded as signs of structural transformation in Tanzania, though at a slower pace.

Panel B in Table 3 presents the average number of workers per household (number of adult members aged 18 and above who reported working at least in one of the income activities). Again, the number of workers in agriculture is higher in rural as compared to urban, but with downward trends over time. This shows some evidence of a structural shift away from agriculture in rural and urban areas. Comparable to the rising share of households in non-agricultural (off-farm) activity as in panel A, the average number of workers on off-farm income shows upward trends in rural areas. The rising number of workers in non-farm activity is driven by increased participation in non-agriculture Self Employment than wage employment activity. Panel C presents the share of workers on each income source to the total number of adults in the household (age18 and above), while panel D presents the income diversification index (simple counts of income sources).

4.3 Panel Data Construction

According to Deaton (1985) and Verbeek & Nijman (1992), the construction of a panel requires the presence of data that is consistently collected across all rounds. The construction of a pseudo-panel is normally based on the presence of suitable variables for grouping households into cohorts

[4]Note that shares need not add up to 100 per cent or subcomponents as households can have members in more than one activity.

(that are time-invariant and observable across all survey rounds) and have been used in previous pseudo-panel studies. Therefore, the commonly used variables in the construction of panel cohorts must satisfy at least two characteristics; should be (1) time-invariant and (2) observable in all cross-section surveys. From a review of the literature, the widely utilized variables to construct a cohort include but are not limited to age (based on year of birth), education level of the household head, region of residence, and location (urban or rural) of the household (Deaton, 1985; Verbeek & Nijman, 1992; Khan & Morrissey, 2019).

We follow a similar procedure to that of Khan & Morrissey (2019) in constructing the pseudo-panel data for Uganda. We consider two common shared household characteristics across all waves in constructing the panel cohorts. First, we use the age (based on birth year cohorts) of the household head as a commonly shared characteristic that is not changing over time and is observed across all HBS rounds. Due to the small proportion of individuals born before 1919 and after 1996, we restrict our estimate to individuals born from 1919 to 1996 and therefore we remain with a maximum of 78 birth cohorts. The year of birth was used to construct a birth cohort with 1-year age bands. We then combine birth cohorts (1-age bands) with the region of residence (about 20 regions) as our additional common shared characteristics for cohort construction resulting in 3,187 total cohorts and an average cohort or cell size of 12 observations. After defining our cross-section units (cohorts) for a pseudo-panel setting, we take averages of each variable over individuals in each cell for each survey year. The sampling weights are used in obtaining the averages of variables for each cohort-year cell. In the constructed (pseudo) panel data, the cohorts are used as cross-sections unit, and the within-cohort year averages as the observations.

There are two main challenges in constructing a pseudo panel: the approach to assigning a group of households or individuals into cohorts. First, due to sampling error which occurs when cohort sample means are not representative of the underlying cohort population means due to variation in cohort composition over time, and second, due to aggregation bias because of loss of variation when using cohort-level data. According to Khan (2021), in the presence of sampling error and aggregation bias, the bias from the latter (aggregation bias) can be so substantial that it can negate any potential benefit of using pseudo-panels over simple OLS. Thus, grouping households or individuals into cohorts for causal inferences, the two potential sources of bias need to be

minimized. We follow the approach by Khan (2018) to assess the likelihood of sampling error and aggregation bias in our constructed panel before estimating the pseudo-panel fixed effects.

Respectively, we use the proposed CAWAR and AWAR statistics to assess whether the sampling error and aggregation bias are minimized, and the results are presented in the Appendix, Table A1. The statistics are estimated with varying year bands from 1 to 5. The overall results show that there is a higher trade-off between aggregation bias and sampling error, such that as the year of the band increase, AWAR becomes less than the critical value (0.5) for most variables and CAWAR exceeds the critical value (11). We choose a 1-year band with almost no aggregation bias for most variables but with some sampling error as shown in Appendix Table A1 and therefore our pseudo-fixed fixed effects estimate must be interpreted with caution.

We finally use the constructed cohort to estimate the pseudo-panel fixed effects to account for possible endogeneity of household diversification. The other potential challenge in estimating the synthetic panel fixed effects is that of heteroskedasticity of standard errors due to substantial variation in the cell size. Literature suggests estimation of the weighted least squares (WLS) using the square root of cell size as weights to account for the heteroscedasticity due to substantial differences in cohort size. Throughout our discussion, we estimate and present WLS pseudo-panel fixed effects estimates.

5. Results and Discussion

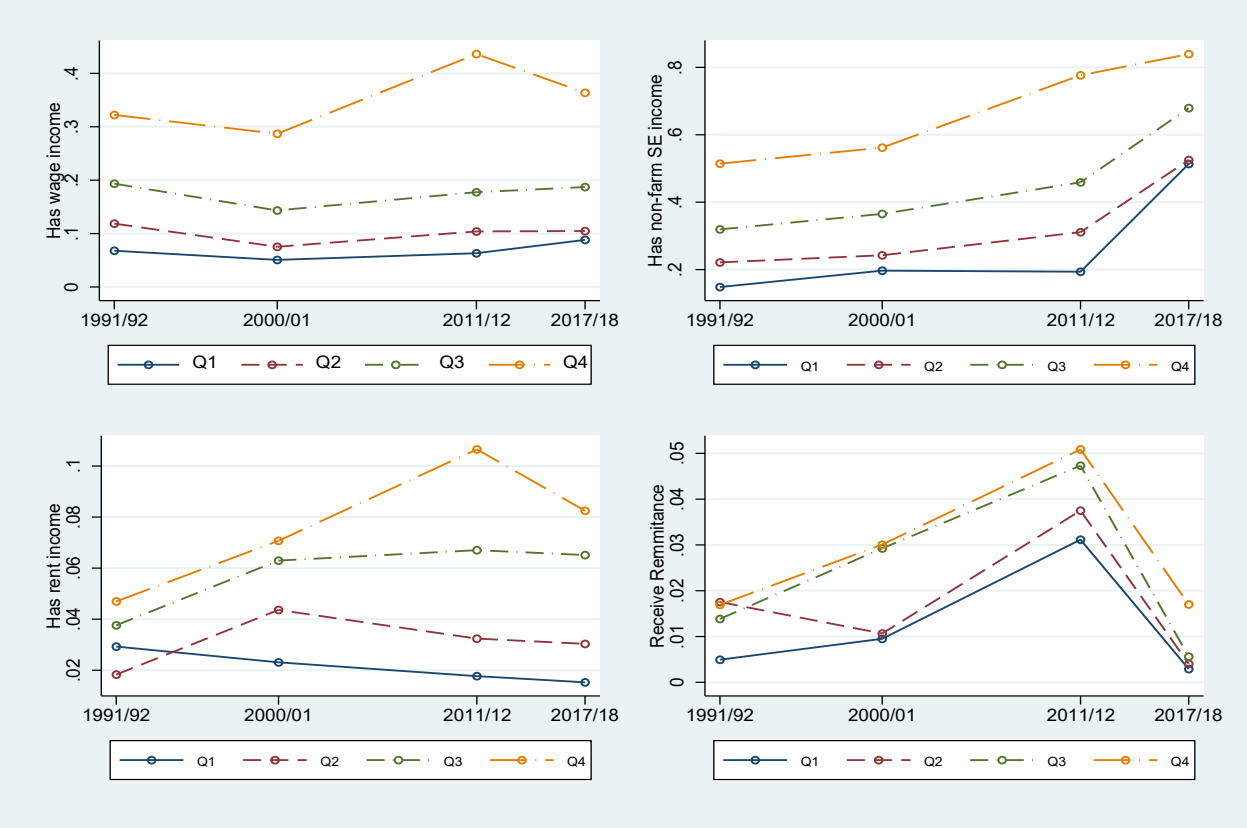
5.1 Preliminary Results

Before turning to the regression results, we begin by plotting income sources and income diversification against the quantile of household consumption to explore the relationship between the level of income and income diversification. We first explore how income activity provides better opportunities to diversify their incomes. Figure 1 shows the correlation between different income sources (wage, non-farm, rental, and remittance) and household consumption per adult equivalent for each survey year. The *top-left* plot presents the correlation between wage income and consumption level. We observe a positive correlation between wage income and consumption, with the upper quartile having a greater share of households with wage income and there has been a relatively high increase in the share of households with wage income at the upper part of the consumption distribution. Similarly, non-farm income (in the *top-right* plot), shows the increase

in the proportion of households with non-farm self-employment income with an increase in consumption level such that households at the upper quartile have a high share of non-farming households. Over time the share of non-farm activity has increased with a higher increase for households in the 3rd and 4th quartile.

A similar conclusion can be drawn from plots at the bottom (on rental and remittances). Households in the 4th quartile have a high proportion of households with both rental and remittance. While the share of households with rental has declined over time, the share for those in the top quartile has increased until 2011/2012 and declined thereafter. The same pattern is observed for remittances income which shows a rising share of those at two top quantiles upm until 2011/12, and declining thereafter. Overall, the results in Figure 1 shows a positive link between household consumption and income sources, such that household at the top distribution of consumption are more likely to have wage income, non-farm self-employment, rental, and remittance income.

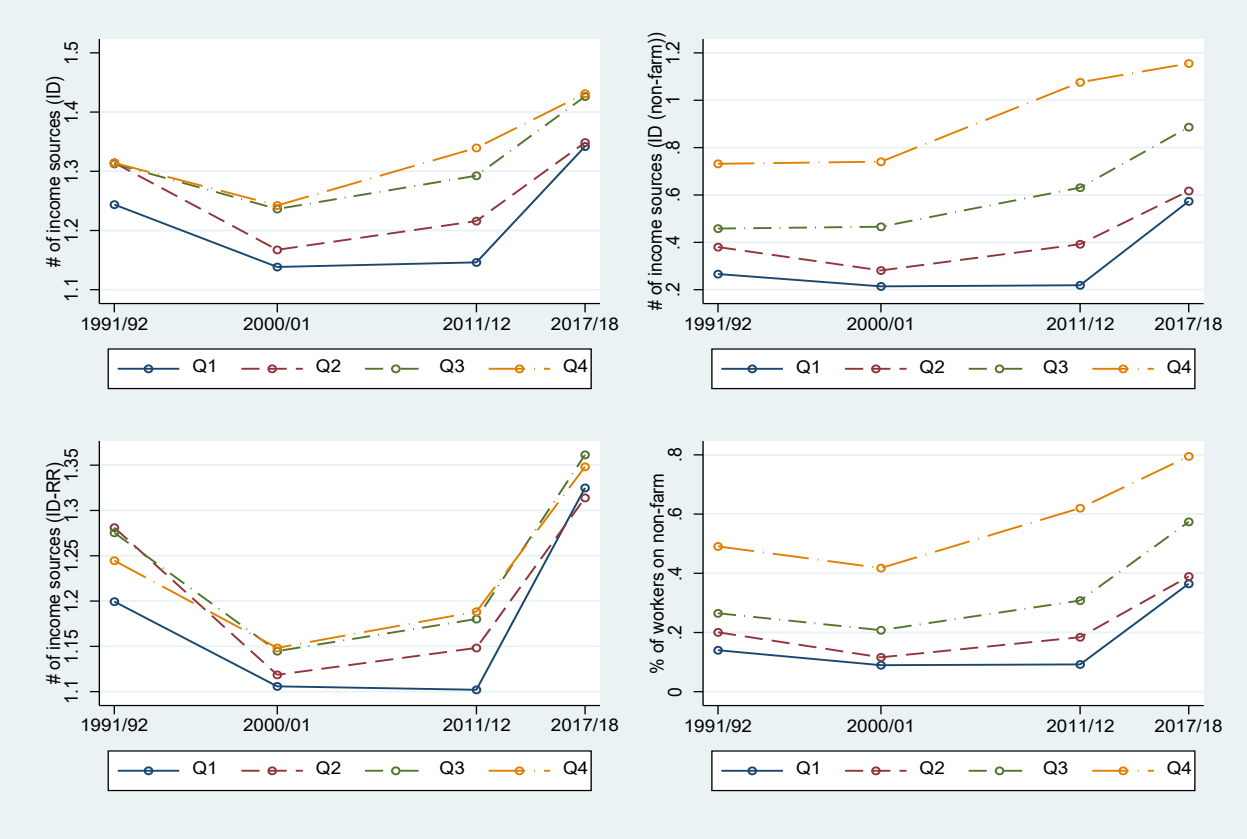
Figure 1: Income Sources and Consumption



Source: Authors' construction from various HBS surveys

It is also interesting to find out how income diversification (aggregate measure of income sources) is related to household consumption level. In turn, Figure 2 plots the household income diversification score measured by counts of total income sources within the households (left plot) and share of households adults in the household on non-farm (self-employment on non-agriculture) versus quantile of consumption per adult equivalent across all four surveys under consideration. Similar to the results in the previous figure, the left plot showing the trends of diversification also suggests a higher correlation between income diversification and consumption. This grants some preliminary support to diversification as means of accumulation rather than a means of survival. The plot shows a higher number of income sources for households in the 4th quartile. Looking at trends over time, the plot show improving income diversification with a relatively higher increase in diversification for the household at the upper tail of the distribution than those at the lower tail suggesting that all households become more diversified over time. The right, with the share of adults in non-farm, reveals a similar pattern, with a non-farm share increasing with consumption level.

Figure 2: Income Diversification and Consumption



Source: Authors' construction from various HBS surveys

5.2 Pooled Regression

Table 4 presents pooled (OLS) regression results of equation (1) for rural and urban households with varying specifications. All estimation account for various demographic controls (such as head gender, dependency ratio, education and marital status, and head economic activity), region-year fixed effects to account for inflation differences and time effects to account for any circular changes in consumption. We only report the estimated coefficients of interest. The results in Panel A show the effects of income diversification (counts) on a household's welfare. Columns 1-3 shows the estimation results of equation (1) for rural and columns 4-6 for urban households. The estimated coefficients of ID in columns 1 and 4 are positive and statistically significant at a 1 percent significant level. This shows a clear and strong positive correlation between household diversification and welfare level for both rural and urban households. However, in terms of magnitude the results show a much higher and stronger effects for the rural household as compared to urban households. Our results in columns (1) and (4) are consistent with other literature in Tanzania showing that income diversifications are more of wealth accumulation (opportunity-led) and not survival-led diversification (Dimova & Sen, 2010; Khan & Morrissey, 2020).

Similarly, results in columns (2) and (5) report the correlation between income diversification (excluding farm in income – ID-off-farm) and household welfare. Resonating with the previous findings, our results show positive and statistically significant effects even after excluding farming income (diversification away from agriculture) with higher impact for rural. In column s 3 and 6 the income diversification excluding rental and remittance income (ID-RR) show a positive and significant effects, though lower in magnitude, for the rural households. The result remains positive and stronger suggesting a strong correlation between income diversification and welfare for those from rural areas. The negative and significant coefficient for urban might be explained by the significance of the two income sources and that urban household diversify their income as means of survival rather than wealth accumulation.

The estimation in Panel B, replaces income diversification indicator variables (ID and ID off-farm) with the respective dummies with one as reference. Columns 1 and 4 show the effects of ID dummies, columns 2 and 5 for the dummies for ID (off-farm) and columns 3 and 6 for ID-RR. Consistent with the earlier results, the estimated coefficients are positive and statistically significant for ID and ID off farm dummies but negative for ID-RR for urban households. It is also

observed that the effects increase with the increase in the counts of income sources within the household. Our results reinforce our earlier finding on the effects of income diversification and welfare.

Table 4: Income Diversification and Household Welfare: Pooled Estimates

	Rural			Urban		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Income Diversification (ID)						
ID	0.126***			0.024***		
	(0.009)			(0.007)		
ID (off-farm)		0.193***			0.154***	
		(0.008)			(0.006)	
ID - RR			0.103***			-0.028**
			(0.010)			(0.009)
Adjusted R-squared	0.804	0.810	0.803	0.772	0.778	0.772
P-value	0.000	0.000	0.000	0.000	0.000	0.000
Panel B: Income Diversification Dummies						
ID = 2	0.116***			0.015		
	(0.011)			(0.009)		
ID = 3	0.293***			0.057**		
	(0.032)			(0.021)		
ID = 4	0.329*			0.222*		
	(0.155)			(0.088)		
ID (off-farm) = 2		0.310***			0.097***	
		(0.023)			(0.010)	
ID (off-farm) = 3		0.484***			0.224***	
		(0.097)			(0.031)	
ID-RR ==2			0.113***			-0.022*
			(0.011)			(0.010)
ID-RR ==3			0.142***			-0.102**
			(0.042)			(0.035)
Adjusted R-squared	0.804	0.805	0.803	0.772	0.773	0.772
P-value	0.000	0.000	0.000	0.000	0.000	0.000
Observations	16455	16455	16455	22447	22447	22447
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Region-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table reports the effects of income diversification and income sources on household welfare. The dependent variable is a natural logarithm of household adult equivalent consumption. Panel A reports the effects of income diversification – counts and panel B report the effects of income diversification dummies on household welfare. Columns 1-3 report the effects for rural and columns 4-6 for urban households. ID is the income diversification measured as a count of household income sources. ID (off-farm) excludes farm incomes and ID-RR excludes rental and remittance income. ID = x or ID (off-farm) =x is the dummy for x counts/ total income sources where x = {2, 3, 4}. All estimation account for various demographic controls, time (survey year) and the regional-year fixed effects. The standard errors are in parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

Overall, the result in Table shows a positive and statistically significant effects of income diversifications for both rural and urban households, suggesting that income diversification is equally important for rural and urban but with substantially higher effects for rural households as compared to urban effects. On diversification excluding farm income (ID (off-farm)), similar evidence is shown, positive and statistically significant for both rural and urban but higher effects are found on rural households. ID (RR) (excluding rental and remittance income) has positive effects for rural households but negative effects for urban. This might be explained by the importance of the two income sources (rental and remittance) for urban households and that the motive for income diversification differ between rural and urban households.

We further explore the impact of each component of income diversification (ID) on welfare. The results are in Table 5. In each estimation, the ID is replaced with four dummies of income sources: wage income, non-farm income, rental, and remittance income. The farm income and head's income sources are omitted due to the possible collinearity with the earnings activity's dummies. Column (8) presents the disaggregated effects of income sources. The finding reveals that having any wage, non-farm, rental, and remittance income sources in the households increases consumption level. The effects of non-farm (self-employment in the non-agriculture sector) though relatively smaller in magnitude, it is statistically significant. Overall, the results show a positive effect of all income sources and statistically significant including self-employment on non-agriculture activity.

Table 5: Income Sources and Household Welfare: Pooled Estimates

	All		Rural		Urban	
	(1)	(2)	(3)	(4)	(5)	(6)
Wage income	0.160***	0.160***	0.265***	0.265***	0.112***	0.112***
	(0.008)	(0.008)	(0.016)	(0.016)	(0.010)	(0.010)
NA Self-employment income	0.005	0.005	0.032**	0.032**	-0.024*	-0.023*
	(0.007)	(0.007)	(0.012)	(0.012)	(0.010)	(0.010)
Farming income	-0.250***	-0.250***	-0.193***	-0.193***	-0.280***	-0.279***
	(0.009)	(0.009)	(0.017)	(0.017)	(0.011)	(0.011)
Rental and Remittance income	0.150***		0.214***		0.130***	
	(0.011)		(0.022)		(0.013)	
Rental income		0.141***		0.244***		0.112***
		(0.012)		(0.027)		(0.014)
Remittance income		0.151***		0.125***		0.160***
		(0.019)		(0.034)		(0.024)
Adjusted R-squared	0.800	0.800	0.814	0.814	0.786	0.786
P-value	0.000	0.000	0.000	0.000	0.000	0.000
Observations	38902	38902	16455	16455	22447	22447
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Region-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table reports the effects of Income diversification and income sources on household welfare separately for rural and urban households. The dependent variable is a natural logarithm of household adult equivalent consumption. ID is the income diversification measured as a count of household income sources. ID (off-farm) excludes farm incomes and ID-RR excludes rental and remittance income. All estimations control for regional-year fixed effects. The standard errors are in parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

In Table 6, we re-estimate equation (1) with three different measures of income diversification (as in panel A of Table 4) to examine whether the effects of income diversification differ across the survey year. Table 6 reports the pooled estimation effects of income diversification on welfare estimated separately for each survey year. Each panel and column represent a separate regression. Each regression controls for region-fixed effects and other controls as in Table 4. Panel A report the effect of ID on welfare across all survey year. The estimated results show that, except for the 1991/92 survey, there is a strong association between diversification and welfare (Panel A). For the diversification score excluding farm income in panel B, the effects are found to be positive across all four-survey years. However, excluding rental and remittance income, in panel C, the effect of income diversification is negative and significant for 1991/92. Though positive the effects on welfare for the other surveys, it's only significant for the 2011/12 survey.

Table 6: Income Diversification and Household Welfare by Survey

	1991/91	2000/01	2011/12	2017/18
	(1)	(2)	(3)	(4)
Panel A: Income Diversification (ID)				
ID	-0.015	0.062***	0.104***	0.054***
	(0.015)	(0.009)	(0.013)	(0.012)
Adjusted R-squared	0.392	0.292	0.490	0.322
P-value	0.000	0.000	0.000	0.000
Panel B: Income Diversification (ID (off-farm))				
ID (off-farm)	0.049***	0.199***	0.181***	0.133***
	(0.014)	(0.007)	(0.011)	(0.011)
Adjusted R-squared	0.394	0.321	0.502	0.336
P-value	0.000	0.000	0.000	0.000
Panel C: Income Diversification (ID-RR)				
ID - RR	-0.030*	0.013	0.099***	0.021
	(0.016)	(0.010)	(0.017)	(0.013)
Adjusted R-squared	0.393	0.290	0.488	0.321
P-value	0.000	0.000	0.000	0.000
Observations	4240	18615	8546	7501
Controls	Yes	Yes	Yes	Yes
Region-year fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes

Notes: The table reports the effects of Income diversification and income sources on household welfare separately for each survey. Each panel and column represent a separate regression. The dependent variable is a natural logarithm of household adult equivalent consumption. ID is the income diversification measured as a count of household income sources. ID (off-farm) excludes farm incomes and ID-RR excludes rental and remittance income. All estimations control for regional-year fixed effects. The standard errors are in parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

In Table 7, we replace the ID with an alternative measure of income diversification (equation 3). We use the share of adult workers (aged above 18) in the household on each income as our measure of the intensity of diversification. Panel A shows the effects based on both rural and urban sample while panel B and C restrict estimation for rural and urban household, respectively. Column (1) of panel A present the effect of share of workers in farming and off-farm activity. The effect is negative and statistically significant for shares in farming but positive for off-farm shares suggesting that an increasing share of workers in agriculture within the household is associated with lower welfare but increases with an increase in off-farm activity share. Column (2) adds the share of workers on non-agricultural activity (off-farm activity – wage and non-agriculture self-employment). The effect of the farming share remains significant and positive for off-farm activity share implying increased welfare with the share of adult members in non-agricultural activity.

Consistently, results in columns (3) to (6), show a significant effect of farm (negative) and non-agricultural activity across all four waves. Restricting the sample to only rural household or urban household shows a similar result.

Table 7: Share of Workers in Income Sources and Household Consumption

	(1)	(2)	1991/92 (3)	2000/01 (4)	2011/12 (5)	2017/18 (6)
Panel A: Rural and urban sample						
Farm	-0.126***	-0.158***	-0.107***	-0.137***	-0.176***	-0.003
	(0.010)	(0.011)	(0.019)	(0.016)	(0.026)	(0.033)
Off-farm	0.195***		0.100***	0.232***	0.207***	0.207***
	(0.007)		(0.014)	(0.011)	(0.015)	(0.024)
Wage		0.247***				
		(0.011)				
NA self-employment		0.131***				
		(0.013)				
Adjusted R-squared	0.800	0.800	0.411	0.344	0.518	0.345
P-value	0.000	0.000	0.000	0.000	0.000	0.000
Observation	38902	38902	4240	18615	8546	7501
Panel B: Rural sample						
Farm	-0.067***	-0.123***	-0.119***	0.027	-0.156***	-0.005
	(0.016)	(0.017)	(0.024)	(0.033)	(0.053)	(0.044)
Off-farm	0.219***		0.115***	0.383***	0.201***	0.180***
	(0.013)		(0.026)	(0.026)	(0.039)	(0.034)
Wage		0.404***				
		(0.025)				
NA self-employment		0.077***				
		(0.021)				
Adjusted R-squared	0.812	0.813	0.226	0.239	0.265	0.211
P-value	0.000	0.000	0.000	0.000	0.000	0.000
Observation	16455	16455	1851	6161	3262	5181
Panel C: Urban sample						
Farm	-0.172***	-0.191***	-0.083***	-0.181***	-0.200***	-0.094
	(0.014)	(0.016)	(0.031)	(0.018)	(0.034)	(0.059)
Off-farm	0.185***		0.094***	0.200***	0.209***	0.221***
	(0.008)		(0.017)	(0.012)	(0.016)	(0.035)
Wage		0.209***				
		(0.012)				
NA self-employment		0.149***				
		(0.016)				
Adjusted R-squared	0.787	0.787	0.294	0.275	0.367	0.330
P-value	0.000	0.000	0.000	0.000	0.000	0.000
Observations	22447	22447	2389	12454	5284	2320
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Region-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table reports the effects of the share of adult workers (aged above 18) on each income source on household welfare. Each panel and column represent a separate regression. The dependent variable is a natural logarithm of household adult equivalent consumption. All estimations control for regional-year fixed effects. The standard errors are in parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

5.3 Pseudo-panel Fixed Effect

As noted earlier, the pooled estimation of specification (1) without controlling for household unobservable heterogeneity might bias the estimated coefficient due to the endogeneity problem. Hence, the results from pooled estimation can only be interpreted as a correlation rather than a causal effect. To address the potential sources of endogeneity due to omitted variable bias for causal interpretation, in this section, we present the results from synthetic panel fixed effects estimation. The fixed effects estimation results with varying specifications are shown in Table 8. The result in panel A report the pseudo-panel fixed effects of income diversification (ID, ID (off-farm) and ID-RR)) on the household welfare (consumption) and panel B uses the dummies for income diversification. Columns (1)-(3) present results for rural and columns (4)-(6) present results for urban household. Each panel-column represents a separate regression. All estimation account for various demographic controls such as the head's age and its squared term, gender, household dependency ratio, marital status, and education level as well as head economic activity (participation in non-farm activity), the region-year, time (survey year), and cohort fixed effects. In panel A, the results on column (1) and (4) on the effects of income diversification (counts) on household welfare show that the effects are positive and statistically significant for the rural household but insignificant though positive for the urban household. Consistent with previous findings, the results suggest a positive causal effect of income diversification on household consumption (hence welfare). Compared with results in Table 4 (column (2)), though closer in magnitude the fixed effect is smaller than the OLS estimate suggesting an upward bias. Also, our estimated effects are closer and comparable to Khan & Morrissey (2020) using panel data but with different income sources mix.

Considering only the off-farm income, the results in columns (2) and (5) show that effects are positive and statistically significant for both rural and urban households. Consistent with the results in column (1), the positive and significant results point out that diversification away from agriculture improves rural households' livelihoods. Controlling for cohort fixed effects, the estimation results suggest that diversification improves household welfare with wider impact for the rural households. Excluding rental and remittance income (ID-RR) in column (3) and (6), the results show that the effect of diversification is positive though insignificant for rural but insignificant for urban households. Column (5) shows the effects of ID-RR (income diversification excluding rental and remittance). Though positive, the effect of ID-RR on consumption is not

significant for unban household. As in Table 4, the insignificant effect might be explained by the significant contribution of the omitted income sources especially rental income in urban areas and remittance in urban areas and that diversification motive may differ between rural and urban household. Overall, our results in panel A provide a piece of clear and supportive evidence on the positive effects of income diversification on consumption/welfare and therefore as a means or opportunity for wealth accumulation rather than a survival strategy.

The estimation in panel B, replaces the income diversification (ID) and off-farm income diversification (ID (off-farm) with its indicator dummy. In column (1), the positive and statistically significant effect of ID=3 but insignificant effects of ID=2 imply that having at least two income sources has a wider effect on overall household welfare for rural households. However, the effects for urban households (column (4)) are not statistically significant. On the contrary, considering only off-farm (non-agriculture) income sources (as shown in columns (2) and (5)) shows that having multiple sources of off-farm income has wider and more significant effects relative to having one or no off-farm income for both rural and urban household. The effects seem to increase with an increasing number of off-farm income sources for rural households. Excluding the rental and remittance income, in columns (3) and (6), the results show that having at least two income sources has a wider effect on overall household welfare for rural households but not for urban households.

Table 8: Income Diversification and Household Welfare: Fixed Effects Estimates

	Rural			Urban		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Income Diversification (count)						
ID	0.063*			0.013		
	(0.033)			(0.031)		
ID (off-farm)		0.149***			0.180***	
		(0.028)			(0.026)	
ID -RR			0.056			-0.028
			(0.038)			(0.036)
Adjusted R-squared	0.941	0.942	0.940	0.917	0.920	0.917
P-value	0.000	0.000	0.000	0.000	0.000	0.000
Panel B: Income Diversification (Dummies)						
ID = 2	0.056			0.007		
	(0.034)			(0.035)		
ID = 3	0.217**			-0.021		
	(0.085)			(0.077)		
ID = 4	-0.828			-0.020		
	(0.606)			(0.166)		
ID (off-farm) = 2		0.189***			0.143***	
		(0.068)			(0.042)	
ID (off-farm) = 3		0.770***			0.095	
		(0.248)			(0.110)	
ID - RR = 2			0.081**			-0.037
			(0.035)			(0.036)
ID - RR = 3			-0.017			-0.109
			(0.126)			(0.094)
Adjusted R-squared	0.935	0.936	0.935	0.905	0.906	0.905
P-value	0.000	0.000	0.000	0.000	0.000	0.000
Observations	2913	2913	2913	2739	2739	2739
Region-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Cohort fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table reports the effects of income diversification and income sources on household welfare using pseudo-panel fixed effects estimation. The dependent variable is a natural logarithm of household adult equivalent consumption. ID is the income diversification measured as a count of household income sources. ID (off-farm) excludes farm incomes and ID-RR excludes rental and remittance income. ID=x or ID (off-farm) =x is the dummy for x counts/ total income sources where x= {2, 3, 4}. All estimation accounts for the demographic controls, regional-year, survey year and cohort fixed effects. The standard errors are in parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

Table 8 reports the fixed effect estimation replacing ID (in equation 3) with each income source dummy. We do so to assess the impact of each component of income diversification on adult equivalent household consumption level (a measure of welfare). The fixed effects results show a higher and more significant effect of non-farm income (wage and rental) on improving welfare of rural and urban household. Though positive, the fixed effect estimate of the NA Self-employment

income for rural household is not statistically significant. On the other hand, the effects of farming income are negative and statistically significant to both rural and urban households suggesting that dependence on income from agriculture is associated with lower level of household welfare.

Table 9: Income sources and Household Welfare: Fixed Effects Estimates

	All		Rural		Urban	
	(1)	(2)	(3)	(4)	(5)	(6)
Income sources:						
Farming income	-0.342***	-0.342***	-0.322***	-0.323***	-0.303***	-0.302***
	(0.030)	(0.030)	(0.053)	(0.053)	(0.039)	(0.039)
Wage income	0.161***	0.160***	0.171***	0.167***	0.135***	0.135***
	(0.027)	(0.027)	(0.045)	(0.045)	(0.035)	(0.035)
NA Self-employment income	-0.014	-0.015	0.021	0.019	-0.017	-0.016
	(0.027)	(0.027)	(0.039)	(0.039)	(0.038)	(0.038)
Rental and remittance income	0.103***		0.070		0.099*	
	(0.039)		(0.059)		(0.052)	
Rental income		0.173***		0.157**		0.161***
		(0.044)		(0.072)		(0.058)
Remittance income		-0.033		-0.076		-0.043
		(0.069)		(0.091)		(0.102)
Adjusted R-squared	0.924	0.924	0.938	0.939	0.912	0.912
P-value	0.000	0.000	0.000	0.000	0.000	0.000
Observations	5652	5652	2913	2913	2739	2739
Region-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
cohort fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table reports the effects of various income sources on household welfare using pseudo-panel fixed effects estimation. The dependent variable is a natural logarithm of household adult equivalent consumption. Wage income is the dummy variable equal to 1 if the household has at least one household member participating in paid/wage activities. Non-farm income equals 1 if at least one household member is self-employed in non-agricultural activity. Rental and remittance are the dummies equal to 1 if the household receives any rental and remittance from migrated members, respectively. All estimation accounts for the demographic controls, regional-year, survey year and cohort fixed effects. The standard errors are in parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

In Table 9, we replace the ID with an alternative measure of income diversification in the pseudo-panel fixed effect model (equation 3). We use the share of adult workers (aged above 18) in the household on each income as our measure of the intensity of diversification. Column (1) and (2) report the overall estimation results while column (3) and (4) report the effects of share of workers on each income sources for the rural and columns (5) and (6) for urban households. Consistent with the previous results, the effect for the share in farming activity are negative and statistically significant suggesting that an increasing share of workers in agriculture within the household is associated with declining welfare. However, the share of workers on non-agricultural activity (off-farm activity) are positive and significant implying that increased welfare with the share of adult

members in non-agricultural activity improves overall welfare. The results also show a significant effect of wage sources and NA self-employment on welfare.

Table 9: Share of Workers in Income Source and Household Welfare: Fixed Effects Estimates

	All		Rural		Urban	
	(1)	(2)	(3)	(4)	(5)	(6)
Share of workers on:						
Farm	-0.109***	-0.114***	-0.107***	-0.114***	-0.139***	-0.123**
	(0.030)	(0.031)	(0.037)	(0.038)	(0.049)	(0.055)
Off-farm	0.227***		0.197***		0.205***	
	(0.023)		(0.035)		(0.032)	
Wage		0.243***		0.281***		0.182***
		(0.040)		(0.081)		(0.049)
NA Self-employment		0.213***		0.152***		0.234***
		(0.037)		(0.053)		(0.056)
Adjusted R-squared	0.924	0.924	0.939	0.939	0.912	0.912
P-value	0.000	0.000	0.000	0.000	0.000	0.000
Observations	5652	5652	2913	2913	2739	2739
Region-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Cohort fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table reports the effects of the share of adult workers (aged above 18) on each income source on household welfare. Each panel and column represent a separate regression. The dependent variable is a natural logarithm of household adult equivalent consumption. All estimations control for regional-year fixed effects. The standard errors are in parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

6. Summary and Implication

Tanzania presents a puzzle for those following growth and development in Africa; since for over a quarter century (1991 – 2018) Tanzania has been experimenting with structural adjustment reforms that have seen impressive macroeconomic performance and sustained growth in one hand but a sluggish poverty reduction in the other hand. In this research, we first construct a (pseudo) panel of representative households using four successive waves National Household Budget Surveys for Tanzania (1991/92, 2001/02, 2011/12 and 2017/18) to investigate the drivers of household welfare (consumption expenditure) over more than three decades. This therefore helps our understanding of the dynamics of household welfare, especially the role of sources of income (labour diversification).

The descriptive statistics have shown that, though falling from 95 percent in 1991/92 to 81 percent in 2017/18, still most households in the rural areas earn their income from farming activities, on average about 88 percent during period while non-farm income sources in rural areas though rising from 36 percent in 1991/92 to 47 percent in 2017/18, on average accounting to about 36 percent. On contrary, the non-farming activities is main sources of income for urban households, on average accounting to about 88 percent, while farming activities account only to about 26 percent during this period. Non agriculture self-employment as sources of income have been increasing both in rural (by about 23 percentage points) and urban (by about 30 percentage points) areas during this period. Other sources, that remittances and rental income have not seen significant changes during this period. Furthermore, the number of workers in agriculture is higher in rural as compared to urban, but with downward trends over time (signs of a structural shift away from agriculture in rural and urban areas).

The preliminary results further shows that, overall there is a positive link between household consumption and income sources, such that household at the top distribution of consumption are more likely to have wage income, non-farm self-employment, rental, and remittance income; implying non-farm share of income inceaesing with consumption level. Consistent with the 'diversification as means of accumulation' hypothesis and in line with previous studies in the region and in Tanzania, the study findings show that household income diversification significantly improves household welfare. However, there are substantial differences in the impacts, with a wider impact for rural than urban households. Our results are robust to alternative measures of household income diversification, the share of workers in farm and non-farm activity. Clearly, if the Government is serious about addressing poverty (and rural poverty), it has therefore to enhance rural infrastructure that enhances the returns of both on-farm and off-farm activities.

References

Abdulai, A., and A. CroleRees (2001). 'Determinants of Income Diversification Amongst Rural Households in Southern Mali'. Food Policy, 26: 437–52.

Anderson, E., and P. Deshingkar (2005). 'Livelihood Diversification in Rural Andhra Pradesh, India'. In F. Ellis and H.A. Freeman (eds), Rural Livelihoods and Poverty Reduction Policies, pp. 55–74. London: Routledge.

Ansoms, A., and McKay, A.' A quantitative analysis of poverty and livelihood profiles: The case of rural Rwanda Food Policy, 35 (6), 584 598.

Arndt, C., V. Leyaro, and K. Mahrt (2014) "Multi-dimensional poverty analysis for Tanzania: First order dominance approach with discrete indicators", UNU-WIDER Working Paper WP/2014/146, Helsinki.

Arndt, C., L. Demery, A. McKay, and F. Tarp (2015) "Growth and poverty reduction in Tanzania." UNU-WIDER working paper 2015/051.

Atkinson, A.B., and M. A. Lugo (2010) "Growth, poverty and distribution in Tanzania", International Growth Centre, London/Oxford Working Paper 10/0831 (November).

Asfaw, S., A. Scognamillo, G. Di Caprera, N. Sitko, and A. Ignaciuk (2019). 'Heterogeneous Impact of Livelihood Diversification on Household Welfare: Cross-Country Evidence from Sub-Saharan Africa'. World Development, 117: 278–95.

Biggs, E. M., Gupta, N., Saikia, S. D., John, M. A. D. (The tea landscape of assam: Multi stakeholder insights into sustainable livelihoods under a changing climate. Environmental Science Policy, 82 9 18.

Bird, K., Shepard, A. Livelihoods, and chronic poverty in semi-arid Zimbabwe. World Development, 31 (591 610.

Block, S., and P. Webb (2001). 'The Dynamics of Livelihood Diversification in Post-Famine Ethiopia'. Food Policy, 26: 333–50.

Barrett, C.B., and B.M. Swallow (2005). 'Dynamic Poverty Traps and Rural Livelihoods'. In F. Ellis and H.A. Freeman (eds), Rural Livelihoods and Poverty Reduction Policies, pp. 16–28. London: Routledge.

Barrett, C.B., T. Reardon, and P. Webb (2001). 'Nonfarm Income Diversification and Household Livelihood Strategies in Rural Africa: Concepts, Dynamics, and Policy Implications'. Food Policy, 26: 315–31

Bezu, S., & Barrett, C. (2012). Employment dynamics in the rural non-farm sector in Ethiopia: Do the poor have time on their side? Journal of Development Studies, 48(9), 1223–1240

Block, S., & Webb, P. (2001). 'The dynamics of livelihood diversification in post-famine Ethiopia', Food Policy, 26(4): 333–350.

Carswell, G. (2002). 'Livelihood diversification: increasing in importance or increasingly recognized? Evidence from southern Ethiopia', Journal of International Development, 14(6): 789–804.

Chuang, Yating (2019), 'Climate variability, rainfall shocks, and farmers' income diversification in India', *Economic Letters,* Volume 174, 55-61

Deaton, A. (1985). Panel data from time series of cross-sections. Journal of Econometrics, 30(1), 109–126.

Dercon, S. (1998). 'Wealth, Risk and Activity Choice: Cattle in Western Tanzania'. Journal of Development Economics, 55: 1–42.

Dercon, S. (2002). 'Income Risk, Coping Strategies and Safety Nets'. World Bank Research Observer, 2(17): 141–66.

De Weerdt, J. (2010). 'Moving out of Poverty in Tanzania: Evidence from Kagera'. Journal of Development Studies, 46(2): 331–49.

Dimova, R., Kristine, S., Nyysolla, M., and K. Sen (2021), 'Long-run rural livelihood diversification in Kagera, Tanzania'. WIDER Working Paper 2021/9

Dimova, R., and K. Sen (2010). 'Is Household Income Diversification a Means for Survival or a Means of Accumulation? Panel Data Evidence from Tanzania'. BWPI Working Paper Number 122. Manchester: University of Manchester, Brooks World Poverty Institute.

Ellis, F. (1998). 'Household strategies and rural livelihood diversification', Journal of Development Studies, 35(1): 1–38.

Ellis, F. (2000). 'The Determinants of Rural Livelihood Diversification in Developing Countries'. Journal of Agricultural Economics, 51(2): 289–302.

Ellis,F., and Mdoe, N. (2003), Livelihoods and rural poverty reduction in Tanzania. World Development, 31(8), 1367 1384.

Ellis, F. & Bahiigwa, G. (2003). Livelihoods and Rural Poverty Reduction in Uganda, World Development, vol. 31, no. 6, pp.997–1013.

Eakin, H. (2005). 'Institutional Change, Climate Risk and Rural Vulnerability: Cases from Central Mexico'. World Development, 33(11): 1923–38.

FAO (2011). The state of food and agriculture 2010–11. Food and Agriculture Organization of the United Nations, Rome, 2011.

Fox, L., and T. P. Sohnesen (2012). 'Household Enterprises in Sub-Saharan Africa - Why They Matter for Growth, Jobs, and Livelihoods'. Policy Research Working Paper WPS6184. Washington, DC: The World Bank. https://doi.org/10.1596/1813-9450-6184

Khan, R., and O. Morrissey (2019). 'Income Diversification and Household Welfare in Uganda 1992-2013'. CREDIT Research Paper 19/05. Nottingham: University of Nottingham, School of Economics.

Kappel, R., J. Lay & S. Steiner (2005), Uganda: No More Pro-poor Growth?, Development Policy Review, 23(1), 27-54.

Khan, R., & Morrissey, I. (2020). Income diversification and household welfare in Tanzania 2008–2013 (WIDER Working Paper, Vol. 2019). UNU-WIDER

Khan, R., & Morrissey, O. (2023). Income diversification and household welfare in Uganda 1992–2012. Food Policy, 116. https://doi.org/10.1016/j.foodpol.2023.102421

Leyaro, V., O. Morrissey and T. Owens (2010), "Food Prices, Tax Reforms and Consumer Welfare in Tanzania 1990 – 2007", International Tax and Public Finance, 17(4), 431-450.

Leyaro, V (2012), “Protection and Determinants of Household Income in Tanzania 1990 - 2007”, Tanzania Economic Review, Vol. 2. No. 2.

Leyaro, V (2013), “Household Consumption Response to Food Price Changes: The Case of Tanzania”, UTAFITI Journal, Vol. 10. No. 1.

Leyaro, V., Baffour, P. T., Morrissey, O., and Owens, T., (2014), ‘Determinants of Urban Labour Earnings in Tanzania’ CREDIT Research Paper 14/03

Masenya, C., Rewete, W., Mogere, D., Temba, V., and Macha, D., (2018), ‘Drivers of Economic Grotwh in Tanznaia’, Bank of Tanzania Working Paper Series No. 14

Menon, J. (2009). ‘Rainfall Uncertainty and Occupational Choice in Agricultural Households of Rural Nepal’. Journal of Development Studies, 46(6): 864–88.

Mduma, J. K., & Wobst, P. (2005). Determinants of rural labour market participation in Tanzania. African Studies Quarterly, 8(2), 32–47.

National Bureau of Statistics (2002). Household Budget Survey 2000/01. Government of Tanzania.

National Bureau of Statistics (2009). Household Budget Survey 2007 – Tanzania Mainland. Government of Tanzania.

National Bureau of Statistics (2014). Household Budget Survey 2011/12 – Tanzania Mainland. Government of Tanzania.

National Bureau of Statistics (2019). Household Budget Survey 2017/18 – Tanzania Mainland. Government of Tanzania.

Njeru EM (2013) Crop diversification: a potential strategy to mitigate food insecurity by smallholders in sub-Saharan Africa. J Agric Food System Community Dev 3:63–69

Niehof, A. (2004), The significance of diversification for rural livelihood systems. Food Policy, 29 (4), 321 338

Osberg, L. and A. Bandara (2012) Why Poverty Remains High in Tanzania: And What to Do About It? Special Paper 12/3, Dar es Salaam, REPOA.

Khan, R. and O. Morrissey (2021), ‘Income Diversification and Household Welfare Uganda 1992 – 2021’, CREDIT Research Paper, No. 19/05

Khan, R (2021), "Assessing Sampling Error in Pseudo-Panel Models," Oxford Bulletin of Economics and Statistics, Department of Economics, University of Oxford, vol. 83(3), pages 742-769, June. <https://ideas.repec.org/a/bla/obuest/v83y2021i3p742-769.html>

Verbeek, M., & Nijman, T. (1992). Can cohort data be treated as genuine panel data? Empirical Economics, 17(1), 9–23.

World Bank (2008). World development report 2008: Agriculture for development. Washington, DC: World Bank.

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

Wang, X., Herzfeld, T., & Glauben, T. (2007). Labor allocation in transition: Evidence from Chinese rural households. China Economic Review, 18, 287–308.

Weltin, M., Zasada, I., Franke, C., Piorr, A., Raggi, M., & Viaggi, D. (2017). Analysing behavioural differences of farm households: An example of income diversification strategies based on European farm survey data. Land Use Policy, 62, 172–184.

Winters, P., L. Corral, and G. Gordillo (2001). 'Rural Livelihood Strategies and Social Capital in Latin America: Implications for Rural Development Projects'. Agricultural and Resource Economics, 2001-6.

World Bank (2015) Tanzania Poverty Assessment, World Bank, Washington DC.

World Bank, (2020). Poverty and Shared Prosperity 2020: Reversal of Fortune. The World Bank, 1818 H Street NW, Washington DC 20433

Ravallion, M., & Datt, G. (1996). How important to India's poor is the sectoral composition of growth in India. World Bank Economic Review, 10, 1–25.

Reardon, T. (1992). 'Income Diversification of Rural Households in the Sahel'. Rural Development Studies, 24: 281–97.

Reardon, T. (1997). 'Using Evidence of Household Income Diversification to Inform Study of the Rural Non-Farm Labor Market in Africa'. World Development, 25(5): 735–48. https://doi.org/10.1016/ S0305-750X(96)00137-4

Reardon, T., C. Delgado, and P. Matlon (1992). 'Determinants and Effects of Income Diversification amongst Farm Households in Burkina Faso'. Journal of Development Studies, 32: 899–912

Senaratna Sellamuttu, S., Aida, T., Kasahara, R., Sawada, Y., Wijerathna, D. (2014). How access to irrigation influences poverty and livelihoods: A case study from Sri Lanka. Journal of Development Studies, 50 748 768.

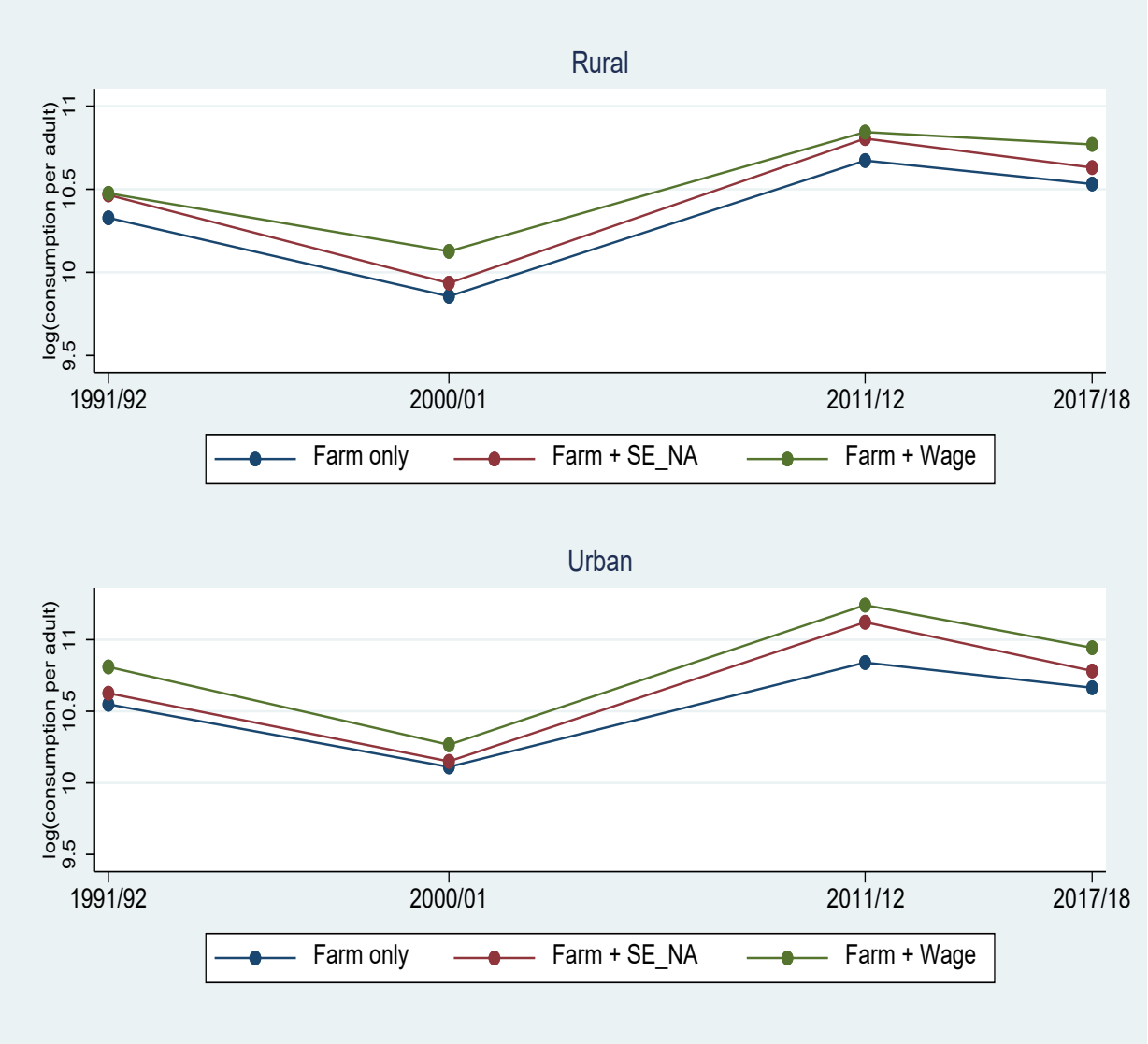
Soltani, A., Angelsen, A., Eid, T., Naieni, M. S. N., & Shamekhi, T. (2012). 'Poverty, sustainability, and household livelihood strategies in Zagros, Iran', Ecological Economics, 79(0): 60–70.

Smith,D. R., Gordon, A., Meadows, K., Zwick, K. (Livelihood diversification in Uganda: Patterns and determinants of change across two rural districts. Food Policy, 26 (421 435.

Yobe, C. L., Mudhara, M., Mafongoya, P. Livelihood strategies and their determ inants among smallholder farming households in KwaZulu Natal province, South Africa. Agrekon, 58 (3), 340 353.

Appendix A: Some More Descriptive Statistics

Figure A1: Income Sources and Consumption level



Notes: SE NA means self-employed on non-agriculture activity.

Table A1: Testing for Sampling Error and Aggregation Bias

	1-year band		2-years band		3-years band		4-years band		5-years band	
	CAWAR	AWAR	CAWAR	AWAR	CAWAR	AWAR	CAWAR	AWAR	CAWAR	AWAR
Panel A: Consumption										
Consumption (log)	59.69	2.21	106.37	2.13	147.08	2.08	191.35	2.07	232.85	2.04
Panel B: Income Diversification and Sources										
ID	4.26	0.59	5.08	0.47	5.82	0.41	6.65	0.39	7.19	0.36
ID (off-farm)	6.36	0.72	9.20	0.63	12.14	0.60	14.87	0.58	17.32	0.56
Farm Income (0/1)	9.14	0.87	14.03	0.77	18.48	0.74	23.17	0.72	28.08	0.71
Non-Farm Income (0/1)	6.82	0.75	9.56	0.64	12.32	0.60	15.00	0.58	17.26	0.55
NA self-employment (0/1)	5.33	0.66	7.53	0.57	9.90	0.54	12.02	0.52	13.22	0.49
Wage Income (0/1)	5.29	0.66	6.95	0.54	8.13	0.49	9.73	0.47	10.89	0.44
Panel C: Share of workers on:										
Non-Farm Income	7.19	0.77	10.52	0.67	14.00	0.64	16.73	0.61	19.85	0.60
Wage Employee	5.48	0.67	7.08	0.55	8.41	0.50	9.88	0.47	11.53	0.45
NA-Self Employment	7.96	0.81	11.57	0.70	15.76	0.68	18.76	0.65	21.98	0.63
Farm income	10.47	0.93	16.04	0.83	21.21	0.79	26.13	0.76	31.01	0.74
Panel D: Other controls										
Dependency ratio	5.42	0.67	7.18	0.55	8.85	0.51	10.38	0.48	11.62	0.46
Married	4.17	0.58	5.03	0.46	5.62	0.41	6.26	0.37	6.92	0.35
Primary	4.29	0.59	5.10	0.47	5.93	0.42	6.62	0.38	7.23	0.36
Secondary	5.00	0.64	5.39	0.48	6.46	0.44	7.21	0.40	7.34	0.36
Diploma	8.18	0.82	7.44	0.56	7.67	0.48	6.87	0.39	5.68	0.32
Higher Education	14.24	1.08	10.35	0.66	10.66	0.56	9.94	0.47	9.16	0.40
Rural (0/1)	9.14	0.87	14.08	0.78	19.18	0.75	23.90	0.73	29.09	0.72
Head: Non-farm activity	10.91	0.95	16.65	0.84	22.82	0.82	28.55	0.80	33.76	0.78

Notes: The table reports the test for sampling error (CAWAR) and aggregation bias (AWAR statistics) based on Khan (2018) for varying age band used to construct the pseudo-panel data cohort.