UNITED NATIONS UNIVERSITY
UNU-WIDER

WIDER Working Paper 2021/9

Long-run rural livelihood diversification in Kagera, Tanzania

Ralitza Dimova,[1] Sandra Kristine Halvorsen,[2] Milla Nyyssölä,[3] and Kunal Sen[3]

January 2021

In partnership with

UONGOZI Institute

United Nations University World Institute for Development Economics Research

wider.unu.edu

This publication results from Sustainable development solutions for Tanzania – strengthening research to achieve SDGs, a collaborative project between the UONGOZI Institute in Dar es Salaam, Tanzania, and UNU-WIDER in Helsinki, Finland, with a main research objective of informing the development and implementation of policies aiming for economic transformation and sustainable development in Tanzania and the East African region. With financial support provided by the Ministry for Foreign Affairs of Finland, the joint project was launched in 2018 with key research questions giving the partners a framework for collaboration and the research work to be undertaken. The project focuses on macroeconomic perspectives, domestic resource mobilization, extractives, industrialization, sustainable livelihoods, and gender as a cross-cutting issue. The project provides local stakeholders with a platform for research and policy discussions on Tanzania and bridges these discussions to the regional and international development debate.

About the UONGOZI Institute

'Uongozi' means leadership in Kiswahili, and inspiring and strengthening leadership is the core purpose of our organization. The UONGOZI Institute is dedicated to supporting African leaders to attain sustainable development for their nations and for Africa. This is done through the provision of high-quality executive education (leadership competencies), facilitation of policy dialogues, action-oriented research, and technical assistance for public and private institutions.

UONGOZI Institute, Magogoni Street, Kivukoni, P.O. Box 105753, Dar es Salaam, Tanzania

Abstract: 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? A large literature has focused on whether income diversification is a means of survival or a means of accumulation, but it remains inconclusive. We first examine the pattern of income diversification for a panel of households in Tanzania from the 1990s—the Kagera Health and Development Survey—with a focus on whether it is primarily driven by survivalist or accumulation motives. We then verify whether this pattern is sustained in the long run using the 2004 wave of the survey while also studying the role that infrastructural improvements and entry into new income generation activities play in the process. Our results support the accumulation hypothesis: richer households engage in more income diversification than poorer households. We also find that 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.

Key words: accumulation hypothesis, income diversification, Tanzania

JEL classification: D12, R20, O55

Acknowledgements: This a is a significantly modified version of Dimova and Sen (2010). The authors would like to thank Joseph Semboja, Oliver Morrissey, and an anonymous reviewer for valuable comments on earlier versions of this paper. The usual disclaimer applies.

[1] Global Development Institute, University of Manchester, UK; [2] Chr. Michelsen Institute, Bergen, Norway; [3] UNU-WIDER; corresponding author: nyyssola@wider.unu.edu

This study has been prepared within the project Sustainable development solutions for Tanzania – strengthening research to achieve SDGs funded by the Ministry for Foreign Affairs of Finland.

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9256-943-3

https://doi.org/10.35188/UNU-WIDER/2021/943-3

Typescript prepared by Joseph Laredo.

United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Finland, Sweden, and the United Kingdom as well as earmarked contributions for specific projects from a variety of donors.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

1 Introduction

One of the most established characteristics of a typical rural household in developing countries is that they obtain their incomes from many different sources (Banerjee and Duflo 2007; Davis et al. 2010; Reardon 1997). Household income diversification is the norm in rural societies, and specialization in a single activity is the exception. The literature has identified a variety of factors that explain income diversification such as risk reduction strategies, responses to household shocks, and asset accumulation strategies that originate from movement into non-farm activities and migration to cities (Barrett et al. 2001a; Ellis 1998, 2000a, 2000b). Household income diversification can be seen as a matter of necessity and survival, where diversification is born out of desperation and driven primarily by the household's poverty status (Ellis 1998), or 'as a matter of choice and opportunity, involving proactive household strategies for improving living standards' (Ellis 1998: 7). The aims of household income diversification driven by survival motives could be

> risk reduction, response to diminishing factor returns in any given use, such as family labour supply in the presence of land constraints driven by population pressure and fragmented landholdings, reaction to crisis and liquidity constraints, high transaction costs that induce households to self-provision in several goods and services, etc. (Barrett et al. 2001a: 315–16).

Household income diversification due to accumulation strategies can involve the 'realization of strategic complementarities between activities such as crop–livestock integration' or 'local engines of growth such as commercial agriculture or proximity to an urban area (that) create opportunities for income diversification in productivity and expenditure-linkage activities' (Barrett et al. 2001a: 316).

We make two contributions to this discussion. First, by examining the pattern of income diversification revealed by the Kagera Health and Development Survey (KHDS), we address the question of whether the income diversification behaviour of a panel of predominantly agricultural[1] households in the Kagera region of Tanzania during the 1991–94 period was characterized mainly by survival or by accumulation motives. One strong manifestation of the 'diversification for survival' view is that poor households diversify more than richer households, while the opposite is true under the 'diversification for accumulation' view.[2] Whether household income diversification is a matter of necessity or choice is of considerable policy importance. If poor households diversify out of necessity, but moving into higher-return yet higher-risk livelihoods endangers their long-term survival, the policy priority should be to reduce the risk of opting for

[1] The clusters are predominantly rural in a global context, even though some are classified as urban in the KHDS. Using agricultural income as a proxy for rural living, only 3 per cent of households in urban clusters in our sample do not have any agricultural income in the study period and 58 per cent have agricultural income as their main source of income.

[2] An alternative way of understanding why households diversify incomes is to relate income diversification to push or pull factors. In theory, survival strategies are driven by push factors while accumulation strategies are driven by pull factors. In practice, however, the difference between push and pull factors on the one hand and survival and accumulation motives on the other is not clear-cut. Not all diversification among the poor is characterized by push factors; nor is all diversification among wealthier households motivated by pull factors. However, in general push factors do characterize the choices made by people who live in poverty or are vulnerable to poverty, while pull factors characterize the decisions of the better-off, as they live in less risky circumstances. In other words, a rich person could bear a risk of falling into poverty and could therefore use a push strategy, but its likelihood is considerably smaller than it is for a poor person.

high-return choices. Alternatively, if diversification is a route out of poverty that is mostly undertaken by wealthier households due to high entry costs, it would be more important from a policy point of view to emphasize public investments in infrastructure (e.g. public transport or electricity), along with the removal of impediments to access to finance and to engaging in high-value agricultural activities by relatively poor households. Earlier literature has failed to provide an unambiguous answer to the 'diversification as survival' versus 'diversification as accumulation' puzzle (Djido and Shiferaw 2018; Ellis 2000b). Moreover, much of earlier literature focuses on diversification within agriculture, examining different crop mixes and so forth, whereas this paper also considers other sources of income. The methodological approach also deals with the endogeneity, as wealthier households tend to have more diversified sources of income (because diversification increases income and spreads risk). The evidence supports the diversification-as-accumulation hypothesis.

Second, we use the extension of the 1991–94 KHDS to 2004 not only to examine how livelihood diversification changed over this 10-year period, but also to discuss the motivations behind these changes and whether the changes turned out to be successful or not. Our main focus is on policy-relevant determinants of diversification, with special interest in how these differ across the different portions of the income distribution. While our results based on data from the 1990s give support to the accumulation hypothesis and inequality in diversification is found to persist in the long run (the rich continue to diversify more than the poor in 2004), there is evidence to suggest that the rural–urban divide in diversification has narrowed. We also find that improvements in public transport and daily markets are related to more diversification, especially among poorer households. Households that were situated in either the poorest or the richest percentile diversified more in 2004 than they did before. This trend was mainly driven by higher return opportunities in non-agricultural employment and non-agricultural self-employment.

The Kagera region in Tanzania is a particularly pertinent context for this analysis, as it tends to be seen as emblematic of a remote, landlocked African rural setting and Tanzania was for long among the poorest countries in the world (De Weerdt 2010; Litchfield and McGregor 2008).[3] The 1990s were characterized by the apparent stagnancy of the rural Tanzanian economy due to slippages in macroeconomic policies and the absence of an enabling environment for households to pursue dynamic strategies of income diversification (Sen 2002). A new wave of reforms started to take place in the 2000s. Although diversification of incomes within rural settings and rural–urban migration can be seen as powerful mechanisms of poverty alleviation in Kagera, there is no unambiguous answer to the questions of whether the government needs to prioritize rural diversification and, if so, precisely how this should be done, or whether rural–urban migration should be a goal in itself (Beegle et al. 2011; De Weerdt 2010; Khan and Morrissey 2020). Access to credit and access to infrastructure are seen as promising routes out of poverty. Yet, there is evidence that granting access to certain types of infrastructure, for instance roads, could in some situations be problematic (see e.g. Dumas and Játiva 2020).

The rest of the paper is organized as follows. In the next section, we discuss how we measure income diversification and describe the conceptual basis of the two perspectives on the causes of diversification. Section 3 discusses the empirical strategy, the econometric methodology, and the data used in the empirical analysis. Section 4 presents patterns of income diversification evident in the data and other relevant descriptive statistics. Section 5 presents the results of the short- and

[3] https://data.worldbank.org/?locations=XO-TZ (accessed 25 September 2020).

long-term regression analyses and provides some further analyses that characterize the findings. Section 6 concludes.

2 Household income diversification: measures and determinants

We begin this section by discussing the conceptual basis of the explanations on the causal origins of income diversification and summarize the findings of previous research on the determinants of household income diversification in the developing country context. We then describe our proposed measure of household income diversification, which we will use in the empirical analysis.

2.1 Why do households diversify?

Perhaps the most important characteristic of household income in rural areas of poor developing countries is its extreme variability. Weather variation; the incidence of disease, pest infestations, and fire; and random shifts in international crop prices can cause farm incomes to fluctuate unpredictably. In the face of large shocks to their income, there are three ways in which poor rural households may attempt to smooth out their consumption. The first of these is the pooling of risk, by which households within a village, kinship group, or social network share each other's risk through institutional arrangements that lead to the most efficient allocation of risk. If such arrangements work well and if shocks or adverse events are idiosyncratic, then any particular household's consumption will track the aggregate consumption of the village, kinship group, or social network and will not be affected by the individual household's income. In this case, there will be little incentive for the household to diversify risks by diversifying the sources of its income. While several empirical studies have documented the existence of risk-pooling mechanisms at the village level (such as Platteau and Abraham 1987 and Townsend 1994 for southern India, and Udry 1994 for northern Nigeria), these studies also show that full risk-pooling is rarely observed, particularly among poorer households. One of the reasons underlying lower levels of risk-pooling is distance. Attanasio and Krutikova (2020) show that the family network acts better as a risk-pooling mechanism if the households in the network maintain close social interactions or are geographically near each other. Thus, informal insurance mechanisms that exist among members of the village, kinship group, or social network will not enable all households to insulate consumption from income fluctuations. This is particularly true if fluctuations in household income are due more to aggregate village-level factors than to household-specific factors (Dimova et al. 2015).

A second way in which rural households may attempt to insulate consumption from large and unpredictable movements in their incomes is by smoothing consumption over time using savings and credit transactions. Households will save in the face of positive shocks to their income, which are expected to be transitory, and dis-save (borrow) in the face of negative shocks to income. By doing so, they will attempt to keep consumption unchanged. While there is a good deal of evidence that households engage in a substantial degree of intertemporal consumption-smoothing using savings and credit transactions (Besley 1995; Deaton 1992), there are strong reasons to believe that rural households in developing countries do not have access to credit markets that might enable them to insulate consumption completely from income shocks. If credit markets are not perfect, some rural households will be constrained in their ability to borrow when faced with a large transitory fall in their incomes, leaving these households unable to cope with income variability.

Both risk-pooling and the use of savings and credit institutions may be seen as ex post means of smoothing consumption. However, if these ex post mechanisms fail (or, more importantly, if households anticipate that these mechanisms will fail), then the preferred strategy for the

household is to smooth consumption ex ante by reducing income fluctuations (Morduch 1995). While households may smooth income by favouring variability-reducing inputs and production techniques and shifting production into more conservative but less profitable modes (Binswanger and Rozenzweig 1993), perhaps the most common method of income-smoothing is to diversify the sources of their income. Thus, income diversification may be seen as a risk-averse household's strategy to minimize the variance of their income by achieving an income portfolio with low covariate risk among its components (Alderman and Paxson 1992; Ellis 2000a; Reardon 1992, 1997; Reardon et al. 1992). Since poorer households tend to be more risk-averse (given the widely held belief that risk aversion tends to decrease with income and wealth), have fewer assets that can be sold to smooth consumption, and have less access to credit facilities or formal and informal insurance mechanisms, they will be more likely to diversify ex ante as a coping response to shocks (Barrett et al. 2001a; Dercon 2002). This theoretical implication of the incomplete markets approach to rural household behaviour underpins the diversification-as-survival perspective in the literature on household income diversification. 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.

Several studies find evidence to support the hypothesis that diversification is driven by income variability linked to survival concerns. Using four years of household data from Burkina Faso, Reardon et al. (1992) find that harvest shortfalls and terms of trade movements drive diversification of income sources out of crops. Using household data for rural districts in Nepal, Menon (2009) finds that when the head of the household is in agriculture, other members of the household are less likely to choose agriculture in districts where rainfall is more uncertain. A similar finding is obtained for Mexico by Eakin (2005). Anderson and Deshingkar (2005), using data from six villages in Andhra Pradesh, India, find that households with lower asset holdings have more diversified income portfolios (though they also find greater diversification among households with asset holdings over a certain level, implying that there is a U-shaped relationship between asset holdings and diversification). Khan and Morrissey (2020) find that in Tanzania 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.

An alternative explanation of household income diversification assumes economies of scope in production, along with entry barriers to high-return economic activities. Economies of scope exist when the same inputs generate greater per unit profits when spread across multiple outputs than when dedicated to any one output (Barrett et al. 2001a). Unlike the presence of economies of scale, which tend to favour specialization in one activity, economies of scope tend to favour diversification as a means of profit maximization. While most empirical studies on Africa or Asia find little evidence of economies of scale beyond a very small farm size, it is likely that diversification across different types of crop (cash crops versus food crops, for example) or across different types of activity (farming in combination with livestock rearing or remittances derived from the migration of some members of the household to cities, for example) leads to significant income enhancement for the household (Barrett et al. 2001a). However, entry into many activities both within and outside agriculture needs initial capital or access to land. Both agricultural and non-agricultural thresholds are evident in most rural economies, where richer households are able to make investment outlays to meet fixed costs in the purchase of cattle and agricultural implements, the setting-up of a non-farm enterprise, or the education of their children for the skilled labour market, while poorer households are unable to do so (Barrett and Swallow 2005). Diversification in this case would be mostly driven by accumulation motives and be mostly confined to richer households. 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. 2001b; Block and Webb 2001; Dercon 1998; De Weerdt 2010).

It is clear from the discussion above that with regard to the causal origins of income diversification, there are strong theoretical arguments for both the diversification-as-survival and the diversification-as-accumulation hypotheses. However, the empirical findings of previous studies remain inconclusive. Additionally, the robustness of the findings of earlier empirical research has been hindered by a lack of adequate attention to problems such as omitted variable bias due to households' attitudes to risk, and reverse causality from diversification either to income or to wealth status of the household. In the next section, we set out our empirical strategy, and show how we test for the necessity versus choice explanations for household income diversification, and we elaborate on how we attempt to address the limitations in the previous literature on household income diversification.

2.2 Measuring household income diversification

A diversified household is generally seen as a household that moves away from only growing crops (that is, being pure cultivators) into non-farm labour such as rearing livestock or into off-farm activities through migration of some members of the household to cities. A variation of this approach makes an additional distinction between crops grown for pure subsistence and commercial (both traditional and high-value) crops. In these studies, diversification is measured using discrete indicator variables for different types of income portfolio that may exist among households (e.g. an income portfolio with no diversification—pure cultivators—will get a value of 1, a mixed income portfolio with both cultivation and livestock rearing will get a value of 2, a mixed income portfolio of both farming and non-farming income will get a value of 3, and so on).[4] Other studies measure income diversification as the proportion of income derived from non-farm sources (Davis et al. 2010; Reardon et al. 1992). While the move from farm activities to non-farm activities would be clearly beneficial to the household in most contexts, measuring diversification only as a transition to more rewarding sources of income or a move away from subsistence agriculture is problematic. First, it becomes a tautological matter that diversification is associated with accumulation if the former is measured as a movement from less productive to more productive sources of income. Second, it is not obvious why a household that derives, say, most of its income from one source should be seen as being more diversified than another household that derives equal shares of income from different sources. For this reason, the use of indicator variables to denote the degree of diversification in different income portfolios is problematic: the construction of such indicator variables is sensitive to the assumptions made about the precise thresholds of income shares used to assign different households to different income portfolio categories.

Two measures of income diversification have been popularly used in literature: the Herfindahl-Simpson (HS) index[5] and the Shannon-Weiner index. Despite the differences in their emphasis, they both measure the richness of used income sources and consider the evenness in their distribution (Barrett and Reardon 2000; Johny et al. 2017; Joshi et al. 2004). In this paper we choose to employ the normalized HS index, as it emphasizes evenness and dominance of a certain strategy rather than rare events or the variety of the strategies available, as in Shannon-Weiner index. The HS index also does not necessitate the arbitrary assignment of households to different income

[4] For example, Abulai and CroleRees (2001) and Dercon and Krishnan (1996).

[5] The HS index is also known as the transformed Herfindahl index, Gini-Simpson index, Gibbs-Martin index, Berry index, Simpson index, and Blau index.

diversification categories (Ellis 2000b).[6] An advantage of the HS measure is that it makes no assumption that a higher degree of diversification is necessarily related to greater household engagement in more remunerative non-farm activities, so by construction, higher values of the measure do not mean greater income accumulation. To explain the used measure with clarity, we begin by presenting the traditional measure of concentration, the Herfindahl index (HI):

$$HI_{i,t} = \sum_{k=1}^{n} IS_{k,i,t}^2 \quad (1)$$

The HI measure is constructed of the sum of income shares $IS_{k,i,t}$ of household *i* from income source *k* in wave *t*:

$$IS_{k,i,t} = \frac{I_{k,i,t}}{I_{i,t}} \quad (2)$$

where $I_{k,i,t}$ is household income from a given source *k*, and $I_{i,t}$ is the total household income from all income sources. In this paper the income diversification index is based on all seven broad categories of income that are present in the data: income from agricultural production (sales and consumption of unprocessed and home-processed crops and livestock); farm employment; non-farm employment; non-farm businesses (self-employment); rents; transfers and remittances; and other non-labour income (pension, insurance, lottery, bride price, and inheritance).[7] This set of income categories is fairly representative of the categories used in the literature (e.g. Davis et al. 2010; Dedehouanou and McPeak 2020).[8]

The HS index is then the complement of the HI, calculated as:

$$HS_{i,t} = 1 - \sum_{k=1}^{n} IS_{k,i,t}^2 \quad (2)$$

A higher value of the index indicates higher levels of diversification. Normalizing the index will render the maximum value to 1. Without the normalization, the maximum value would be 1-(1/*n*). The normalized index is calculated simply[9] as:

$$NHS_{i,t} = \frac{HS_{i,t}}{1-\frac{1}{n}} \quad (4)$$

---

[6] The logic of this income diversification index is broadly similar to those of the income diversification indices used by Davis et al. (2010), which have been constructed at the country level, but unlike these indices, it allows us to measure diversification at the level of the household, which is our unit of analysis, as opposed to the country level.

[7] When defining the categories, we included some sub-income sources within a more general group if their occurrence and income shares were low. An example of this is stipends, which do not form their own income source category but are allocated to transfers, which comprises remittances and other transfers.

[8] Separating crop and livestock incomes into two categories did not change the index value. However, if certain crucial income sources are excluded, the results can change despite the normalization of the diversity index. For example, Djido and Shiferaw (2018) use only three categories (non-farm activities, staple crop, and high-value crop production), which are the most common income groups among the households they study. This means that households tend to have some income from at least two of the three sources, which renders the index value high for all and does not allow much variation in income source portfolios. Using the same number (five) of income sources as in Dedehouanou and McPeak (2020) did not change our measure value statistically significantly (the average NHS using these five income sources is 0.299 using our data).

[9] This form is the one used in Smith and Wilson (1996). It is a simplified form of the ones presented in some recent literature (e.g. Dedehouaunou and McPeak 2020; Djido and Shiferaw 2018).

3 Data and empirical strategy

3.1 Data

We use data from the first five waves of the KHDS, a longitudinal household survey conducted in the Kagera region of Tanzania. There were six waves of the KHDS: 1991/92, 1992, 1992/93, 1993/94, 2004, and 2010. The sixth wave (2010) is not used in this paper, because it omitted the detailed income questions previously included. This region of approximately 1.9 million people is located on the western shore of Lake Victoria, bordering Uganda to the north and Rwanda and Burundi to the West. The population is overwhelmingly rural and mainly engaged in the production of bananas and coffee in the north and rain-fed annual crops (maize, sorghum, and cotton) in the south. The survey was conducted in 51 communities in all 8 districts of Kagera: Bukoba urban, Bukoba rural, Missenyi, Karagwe, Kyerwa, Muleba, Biharamulo, and Ngara.[10] Of the 912 original households, 759 completed all waves in the 1990s and 832 were re-interviewed in 2004. The household questionnaire is based on the World Bank's Living Standards Measurement Survey, assuring representativeness and quality of the data.

3.2 Empirical strategy

To reiterate, the first step of our empirical analysis, based on a four-year panel from the 1990s, focuses on testing whether income diversification is driven by survivalist or accumulation motives. To answer the empirical question, we estimate a regression of the following generic form for the 1991–94 panel:

$$NHS_{ivt} = \alpha + \rho Y_{it} + \sum_k \gamma_k R_{vt} + \sum_k \beta_k X_{it} + d_t + u_i + e_{it} \qquad (5)$$

where $_i$ designates the household, $_v$ designates village in which the household resides, $_t$ designates time, NHS is the normalized HS index of household diversification (larger values imply more income diversification), Y is the logarithm of real household income[11] per adult equivalent (PAE), and R_{vt} and X_{it} are vectors of standard control variables at the village and household level, respectively. The error terms d_t and u_i capture the time-invariant and household-invariant components of the error term, while e_{it} is the white noise component of the error term. The year effects d_t have been included to capture year-specific national-level shocks, such as weather and other macroeconomic shocks that may affect diversification behaviour for all households in a given year. Knowing that agricultural income often varies from season to season, we also control for the season, because the interviews were not done every year at the same time.[12] The household-specific effect u_i captures unobserved household characteristics that are time-invariant and, most importantly from our perspective, unobserved household attitudes to risk. Because the households were drawn from a stratified random sample of households, the standard errors are clustered at village level to account for village-level unobservables.

[10] After the last wave of the KHDS, the district borders were redrawn, so that two new districts were formed in Kagera (namely Kyerwa and Missenyi) and an area from Bukoba rural was reallocated to Muleba. This paper uses the pre-2010 definition of districts in the empirical analysis.

[11] Inflation adjustments are based on the World Bank Consumer Price Index for Tanzania (https://data.worldbank.org/country/tanzania?view=chart, accessed 16 May 2020).

[12] There are three seasons: Masika (heavy rain), Vuli (light rain), and Kiangazi (dry).

A positive sign for the coefficient of the income variable indicates that the diversification-as-accumulation hypothesis holds, while a negative sign indicates that the diversification-as-survival hypothesis holds. Given the plausible assumption that the income portfolios of the richest and poorest strata of the population may be more diversified than those of the middle-income strata (Anderson and Deshingkar 2005), we also test the U-hypothesis, where the relationship between income diversification and income may be characterized by a U-shape pattern, with high levels of diversification at both ends (at low and high levels of income) but a low level in the middle of the distribution. However, this hypothesis was not supported by our data.[13]

Our control variables at the village level are a set of dummy variables that capture whether the village is categorized as urban (such as some villages in the district of Bukoba and Muleba); has access to any formal or informal credit institutions; is electrified; has access to a post office and telecommunications; has public transport; has a daily market; has at least one secondary school; has a health facility. Infrastructure is expected to have a positive effect on household income diversification, as households in villages that have these infrastructural facilities would be more able to diversify by taking advantage of non-farm employment opportunities or trade, or by migrating to cities. The effect of credit facilities on household income concentration could be positive or negative, depending on whether household income diversification is driven by survival or accumulation concerns. If income diversification is driven by survival concerns, then greater access to credit will enable poor households to smooth consumption inter-temporally, rather than by diversifying their income portfolios. In that case, the relationship between access to credit facilities in the village and household income diversification will be negative. On the other hand, if income diversification is mostly a means of accumulation, access to credit will provide households with an easier route into non-farm activities or livestock rearing, which have high entry costs. In that case, the relationship between access to credit facilities in the village and household income diversification will be positive.

Our control variables at the household level are standard demographic variables such as the age, education (dummy equal to 1 if the head has at least completed primary school), and gender of the head of household; the proportion of members of the household who are dependants; and the size of the household. Previous research has found that male-headed households and households with more children are associated with more diversified income portfolios (Dercon and Krishnan 1996). With respect to household size, we expect that larger households diversify more, given that there are more individuals in the household.

We use an Instrumental Variables (IV) method of estimation along with household fixed effects and year effects. The reliance on an IV methodology is driven by the possibility of reverse causality between household income and income diversification. For example, using a Q-squared methodology combining 10-year panel data with qualitative life histories for Tanzania, De Weerdt (2010) finds that households that moved out of poverty were those that diversified their farming activities, growing food crops for their own consumption and cash crops for sale, and keeping livestock. It is not clear whether the positive correlation between diversification away from food crops and movement out of poverty is due to income diversification, or whether household income status drives the diversification behaviour. Our dataset provides a large number of variables that are exogenous in the household income regression and could serve as credible instruments. There is no credible econometric way of assuring the quality of our instruments. Most tests for endogeneity available in the literature assume ex ante that the instrument chosen is at least conceptually appropriate. We experimented with instrumentation (without simultaneously

[13] We tested the hypothesis using centred quadratic regression. The results did not give support to the U-hypothesis, but rather gave an indication of a positive linear relationship between log-income and diversification.

correcting for fixed or random effects) and our final set of instruments passed the Sargan overidentification test. The set of instruments that we finally selected includes different types of village-level shocks (e.g. refugee inflow, epidemic, or natural disaster), the rainfall variability in the district over the preceding year, and an indicator of whether a working member of the household died during the preceding year. These variables impact on diversification via their effects on income and not directly, while also satisfying the exclusion criteria of being included as instruments.

Additionally, we use household fixed effects to account for unobserved household attitudes to risk, which may explain household diversification behaviour, independent of income. For instance, the classical literature invariably conditions income diversification strategies to household (or head of household's) attitude towards risk taking (Dercon 1996, 2002). Yet, the more recent behavioural literature finds high levels of risk preference heterogeneity among households across the different portions of the income distribution (Basu and Dimova 2020; Basu et al. 2020). This suggests that empirical analysis of income diversification must disentangle household innate characteristics, such as attitudes to risk, from those of other households, alongside community and macro variables that may impact on household income diversification. We also include random effects instead of fixed effects in several of the estimates as robustness checks. Given the short time variation in our panel, and the large cross-sectional dimension, random effects estimation may be more efficient than fixed effects, and we experiment with both random and fixed effects in the estimation of equation (5).

Since certain variables are often fixed over time in the short run, we follow the two-step procedure suggested in Pesaran and Zhou (2018) to correct for collinearity between the time-invariant variables in fixed effects models.

In the second part of the empirical analysis we take advantage of the fact that we have a 2004 wave of the KHDS data. The 10-year gap between the last wave in the 1990s and the 2004 wave provides an ideal opportunity for us to examine the long-run patterns and correlates of income diversification for both poor and rich households. First, we examine whether actual income diversification of households in 2004, as captured by the HS index of household diversification, is closely correlated with the predicted measure of income diversification that we obtain from the panel regressions using the 1990s data across the income distribution. This allows us to assess how the actual income diversification patterns differ vis-à-vis the synthetic scenario of the socioeconomic situation of the 1990s remaining unchanged. These differences are explored separately for poor and rich households and between those living in rural and in urban areas.

We next examine whether and to what extent the observed differences between the actual and predicted income diversification patterns are related to households gaining access to infrastructural factors (credit institutions, electricity, telephone and postal service, public transport, daily market, secondary school, and health facility) for different portions of the income distribution through correlation analysis. Following households over long periods is problematic, as household members grow older and enter or exit the labour force, or move in or out of the household, possibly changing the household composition. To link households from the 1991–94 panel with households in 2004, we identify and follow the household head and include only household members who live with the household head. In 38 per cent of the households, the person who was the head in Waves 1–4 is not interviewed in Wave 5, because the person is either no longer alive or has moved out and can no longer be traced. In these cases, we follow the household in which most of the family members remain and where the spouse, son, or parent has taken over as the household head.

Finally, we use a difference-in-differences approach to examine whether entering different types of income activities is motivated by income accumulation or survival. The model follows the standard generic form:

$$Y_{ivt} = \alpha + \beta_1 enter_s + \beta_2 W_{t=5} + \beta_3 (enter_s \cdot W_{t=5}) + \sum_k \gamma_k R_{vt} + \sum_k \delta_k X_{it} + e_{ivt} \quad (6)$$

where *i* indicates the household, *v* indicates the village, *t* indicates the wave (0 if Waves 1–4 and 1 if Wave 5), and *s* indicates an income source. The model is run for each income source separately. Y_{it} is the equivalized log-income, $enter_s$ is a dummy indicating whether a household entered into a certain income activity (for example, it takes the value 1 if a household had no income from employment in Waves 1–4 but had employment income in Wave 5), $W_{t=5}$ is a dummy indicating Wave 5, and R_{vt} and X_{it} are vectors of standard control variables at the village and household level, respectively (head age, female head, household head completed primary school, dependence ratio, household size, season, urban area, district, and log of total household assets). The parameter β_3 is the coefficient of interest and measures the effect of a household's entering a new income-generating activity in Wave 5.

4 Descriptive statistics and patterns of income diversification

Table 1 presents the descriptive statistics of the pooled sample of the first four waves and the fifth wave separately. First, the mean of the normalized index of diversity is roughly 0.30 when pooling all waves with a median of 0.24 and range of 0–0.88. There are no previous results from Tanzania to compare the mean index value with.[14] Diversification is not increasing over time on average. The household characteristics have remained quite similar across the waves. The age of household head has risen by only 5 years in a decade, as some household heads have died between the waves and a younger head has taken over. Households have experienced less death and disasters in Wave 5 than on average in Waves 1–4. Notable improvements have taken place when it comes to access to infrastructure and facilities. In 2004 almost all villages have a health facility, secondary school, and bank.

In Table 2 we explore the variation of the diversification index by key demographic characteristics of the household and infrastructural characteristics of the villages. Diversification is more prevalent in urban areas, during the rainy season, among female-headed households, and among the highly educated. Large households diversify more, but typically when there are not many dependants. When looking at infrastructure, we find that the availability of credit and access to electricity, post and telephone services, public transport, daily market, secondary school, and health facility have a positive correlation with diversification. The observation that better infrastructure, especially in the form of the ability to obtain credit, eases diversification constraints is one of the most common findings in the literature on income diversification (e.g. Menon 2009).

[14] Results from other contexts do exist. For example, Dedehouanou and McPeak (2020) get a mean value of 0.23, suggesting that Nigerians diversify slightly less than the people of Kagera. Our results are robust to using the same income source categorization as in Dedehouanou and McPeak (2020).

Table 1: Descriptive statistics

	Waves 1–4 (n=3,328)		Wave 5 (n=725)	
	Mean	**SD**	**Mean**	**SD**
Diversification index	0.297	0.241	0.294	0.241
HH income (PAE)	200,036	320,719	216,358	526,706
Age of head	49.314	17.107	54.421	16.853
Female head	0.274	0.446	0.309	0.462
Proportion of dependants	0.483	0.243	0.459	0.253
Household size	5.725	3.109	5.266	2.928
Head no education	0.207	0.405	0.258	0.438
Head elementary school	0.741	0.438	0.666	0.472
Head secondary school	0.046	0.210	0.065	0.246
Head university	0.005	0.069	0.010	0.098
Death in family (6–12 months)	0.125	0.331	0.072	0.258
Access to:				
Urban areas	0.212	0.408	0.211	0.408
Credit	0.521	0.500	0.971	0.168
Post and telephone	0.098	0.298	0.760	0.427
Electricity	0.269	0.444	0.684	0.465
Daily market	0.592	0.492	0.472	0.500
Public transport	0.255	0.436	0.699	0.459
Secondary school	0.080	0.271	1.000	0.000
Health facility	0.365	0.481	1.000	0.000
Disaster (last 6–12 months)	0.609	0.488	0.166	0.372

Note: income is reported in real Tanzanian shillings PAE.[15]

Source: authors' calculations based on KHDS data.

Table 2: Income diversification index value tabulated by relevant household characteristics

	Waves 1–4			Wave 5		
	NHS	**SD**	**Obs**	**NHS**	**SD**	**Obs**
Total	0.297	0.241	3,328	0.294	0.241	725
Area type						
Urban	0.414	0.244	704	0.381	0.249	153
Rural	0.266	0.230	2,624	0.271	0.234	572
Difference	0.148***			0.110***		
Household head: gender						
Male	0.290	0.245	2,417	0.285	0.244	501
Female	0.316	0.228	911	0.313	0.234	224
Difference	0.026**			0.028		
Household head: education						
No education	0.248	0.224	689	0.259	0.232	187
Elementary school	0.300	0.240	2,468	0.297	0.241	483
Secondary/University	0.456	0.244	170	0.378	0.250	54
Differences						
(no educ vs elementary)	0.052***			0.038*		
(elementary vs secondary)	0.156***			0.081**		

[15] This paper uses a gender-neutral version of the Tanzanian adult equivalence scale (AES), as in Nyyssölä et al. (forthcoming). The model results are robust to using the official Tanzanian AES as well.

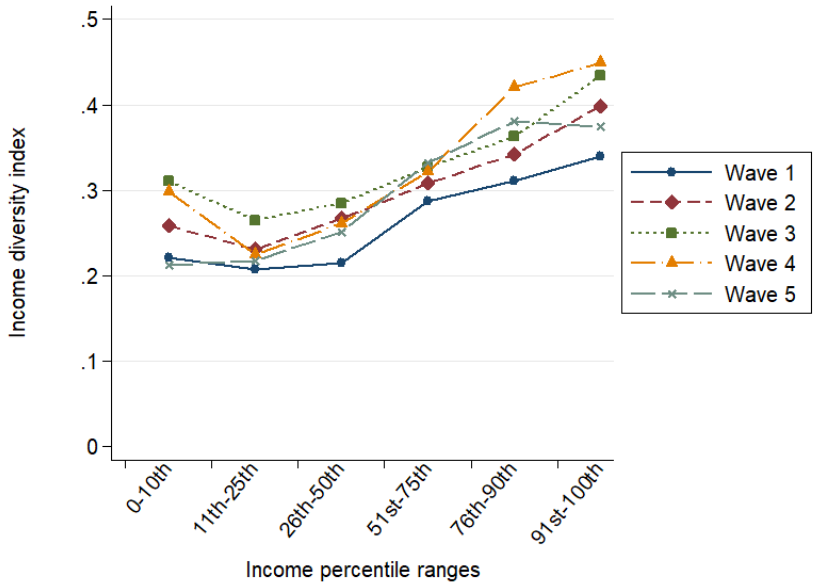
	Waves 1–4			Wave 5		
	NHS	SD	Obs	NHS	SD	Obs
Household size						
1–3	0.273	0.226	852	0.254	0.224	198
4–10	0.298	0.244	2,246	0.307	0.244	501
>10	0.370	0.248	230	0.345	0.278	26
Differences						
(1–3 vs 4–10)	0.025***			0.053***		
(4–10 vs >10)	0.072***			0.038		
Dependants as proportion of household size						
0%	0.293	0.236	358	0.252	0.235	87
1%–50%	0.306	0.245	1,486	0.314	0.246	365
>50%	0.281	0.239	1,483	0.281	0.235	273
Differences						
(0% vs 1%–50%)	0.013			0.062**		
(1%–50% vs >50%)	0.024*			0.032*		
Season						
Vuli (light rain)	0.293	0.241	1,714	0.284	0.228	347
Masika (heavy rain)	0.302	0.236	547	0.319	0.250	296
Kiangazi (dry)	0.301	0.241	1,068	0.244	0.256	82
Difference						
(Masika vs Kiangazi)	0.001			0.075**		
Bank						
Not available	0.290	0.245	1,594	0.445	0.220	21
Available	0.309	0.236	1,734	0.290	0.241	704
Difference	0.025**			0.155***		
Electricity						
Not available	0.268	0.230	2,432	0.247	0.223	229
Available	0.375	0.250	896	0.316	0.247	496
Difference	0.106***			0.068***		
Post and telephone						
Not available	0.290	0.240	3,001	0.252	0.232	174
Available	0.359	0.243	327	0.307	0.243	551
Difference	0.068***			0.056***		
Public transport						
Not available	0.290	0.240	2,478	0.239	0.220	281
Available	0.316	0.243	850	0.318	0.246	507
Difference	0.026***			0.079***		
Daily market						
Not available	0.257	0.233	1,358	0.256	0.233	383
Available	0.324	0.242	1,970	0.337	0.243	342
Difference	0.070***			0.081***		
Secondary school						
Not available	0.290	0.238	3,063	-	-	-
Available	0.372	0.252	265	0.294	0.241	725
Difference	0.092***			-		
Health facility						
Not available	0.282	0.234	2,113	-	-	-
Available	0.322	0.250	1,215	0.294	0.241	725
Difference	0.040***			-		

Note: NHS is the diversification index. ***, **, * indicate p-values of the pairwise t-test of difference in means at the 1%, 5%, and 10% significance levels. Continuously insignificant pairwise comparisons (t-tests) were dropped.

Source: Authors' calculations based on KHDS data.

The main variable of interest for the diversification as a means of accumulation or survival hypothesis is household income. Figure 1 provides a plot of the household diversification index in each year against the household income percentiles. The relationship between income diversification and level of income is positive.[16] This provides some preliminary support for the diversification as a means of accumulation hypothesis.[17] The level of income diversification is lower in Wave 1 (solid blue line) than the level of diversification across all income groups in later waves. In our empirical analysis in Section 5, we control for various types of year-specific shocks that may influence diversification behaviour on a year-to-year basis using a year dummy.

Figure 1: Mean income diversification index by income percentiles



Source: authors' illustration based on KHDS data.

The type of income source varies by income status and over time. Table 3 presents the income diversification strategies by grouping the households into categories of income portfolios. Among the poorer half, the majority is engaged only in agriculture, whereas the richer half tends to have more than one income source. Dependence on agriculture has decreased over time among the richer half, and more households are diversifying in 2004. The poorer half has not become more diversified over time when it comes to the number of income sources.

[16] In Wave 4 there is a difference at the 5% significance level between the 0 and 10th percentiles and the 11th and 25th percentile, suggesting that on average the poorest group (0–10th percentile) was diversifying more than the second-poorest group (11–25th percentile). Such an observation may give support to a U-shape of diversification. However, this is observed in only one of the five waves.

[17] Poor deciles may have liquidity constraints that can stop them from pursuing a more varied set of productive, income-generation strategies. The set of strategies can be geared more towards low-income tasks or cost-saving functions (begging, eating less, taking kids out of school, collecting rubbish), which we cannot capture in our dataset.

Table 3: Income diversification strategies by wave and by income division

	Waves 1–4		Wave 5	
	Below median income	**Above median income**	**Below median income**	**Above median income**
Only one income source				
Self-employment in agriculture	56.5%	38.6%	54.4%	28.5%
Any other source than agriculture	0.9%	4.5%	3.6%	9.7%
Total	**57.4%**	**43.1%**	**58.0%**	**38.2%**
More than one income source				
Agriculture & any other source(s)	34.7%	41.1%	38.2%	46.8%
Any other combination	7.9%	15.8%	3.8%	15.0%
Total	**42.6%**	**56.9%**	**42.0%**	**61.8%**

Source: authors' calculations.

5 Results

5.1 Income diversification in the short run

Table 4 presents the estimates of equation (5), where the key explanatory variable is the equivalized real household log-income using Waves 1–4. Columns 1–3 give the simple ordinary least squares estimates without household fixed effects (FE). Columns 2–3 include household-level controls and year effects, using first fixed effects and then random effects (RE). Columns 4–5 present the estimates using the IV method of estimation with household-level controls and year effects and with fixed and random effects. Finally, columns 6–7 give estimates of the full specification including all village level controls.

The results are fairly consistent across the different specifications. The estimates of the main models (4–7) indicate a strong positive impact of the income variable on the degree of diversification, which provides clear support for the accumulation hypothesis.

While age and education of the household head and dependency ratio do not affect household income diversification in the main models, female-headed households tend to diversify more than male-headed households (only significant in RE models). This is consistent with the descriptive analysis.[18] Household size is positively correlated with income diversification across all specifications.

There are also seasonal effects on diversification. In all specifications, there seems to have been more diversification when the month prior to the interviews was in the dry season, compared with the base season of heavy rain (Masika). This positive finding could be evidence that diversification undertaken during the dry season is based on accumulation motives rather than survival motives. During the dry season farmers tend to engage in different activities than in rainy seasons (de Bont

[18] Most female household heads are widowed or divorced/separated women. Only 8 per cent of female heads on average are married or together with a partner, out of which almost 70 per cent have a spouse living elsewhere. By separating the single and non-single female-headed households in the regression models, we find that non-single female-headed households are more similar to male-headed households in terms of the diversification index, all else being equal. Due to the small fraction of such households, we do not keep this categorization, but instead group together single and non-single female-headed households in the regression models in Table 4.

et al. 2019). Using the dry season as the base category, we find that the estimates for Masika are significant and negative for all the models, suggesting that the farmers who choose to diversify during the heavy rain season may do so out of necessity.

Table 4: Impact of household income on diversification index, Waves 1–4

	(1) OLS	(2) FE	(3) RE	(4) FE + IV	(5) RE + IV	(6) FE + IV	(7) RE + IV
Log HH income (PAE)	0.037***	0.008	0.026***	0.192***	0.220***	0.197***	0.221***
	(0.008)	(0.009)	(0.008)	(0.074)	(0.060)	(0.074)	(0.056)
Head age	-0.000	-0.001	-0.000	0.001	0.000	0.001	0.000
	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)
Female head	0.047***	0.021	0.041***	0.068	0.071***	0.070	0.070***
	(0.013)	(0.031)	(0.013)	(0.045)	(0.021)	(0.045)	(0.020)
Head has primary educ.	0.032**	-0.001	0.024*	0.008	-0.000	0.008	-0.002
	(0.015)	(0.018)	(0.014)	(0.023)	(0.018)	(0.023)	(0.018)
Proportion of	-0.051**	-0.025	-0.050**	-0.037	-0.005	-0.034	-0.002
dependants	(0.021)	(0.039)	(0.022)	(0.045)	(0.035)	(0.045)	(0.033)
Household size	0.012***	0.010**	0.012***	0.025***	0.019***	0.026***	0.019***
	(0.002)	(0.004)	(0.002)	(0.007)	(0.003)	(0.007)	(0.003)
Season							
Vuli (light right)	0.019	0.026*	0.018	0.021	0.018	0.016	0.016
	(0.014)	(0.015)	(0.015)	(0.017)	(0.017)	(0.017)	(0.016)
Kiangazi (dry)	0.031**	0.038**	0.030**	0.042**	0.041**	0.040**	0.042**
	(0.015)	(0.016)	(0.015)	(0.018)	(0.017)	(0.018)	(0.016)
Urban	0.110***	0.148***	0.135***	0.158***	0.065**	0.125***	0.054*
	(0.023)	(0.015)	(0.024)	(0.015)	(0.030)	(0.019)	(0.030)
Credit	0.005					0.032**	0.023*
	(0.017)					(0.015)	(0.013)
Electricity	0.027					0.071***	0.021
	(0.020)					(0.020)	(0.021)
Post and telephone	0.059*					0.027	0.051
	(0.031)					(0.027)	(0.039)
Public transport	-0.003					0.036*	-0.026
	(0.017)					(0.019)	(0.025)
Daily market	0.021					0.003	0.004
	(0.015)					(0.017)	(0.015)
Secondary school	0.065					0.077***	-0.013
	(0.050)					(0.023)	(0.061)
Health facility	0.009					0.010	0.004
	(0.023)					(0.026)	(0.023)
Overall R^2	0.141	0.036	0.121	0.062	0.069	0.063	0.072

Note: number of observations is 3,328 in all models. Number of households is 915 in all models. All models include year effects, which appear statistically significant at 1% level. Log household income is the log of real household income PAE. Season indicates the season prevailing during the month (self-reported by the households) prior to the months when the interviews took place. The omitted season is Masika (heavy rain). Some of the village-level covariates were reported only in Wave 1 and are therefore omitted in the FE regressions. The figures in brackets are cluster robust standard errors. ***, **, * indicate significance at the 1%, 5%, and 10% levels.

Source: authors' calculations based on KHDS data.

To make sure that the regression results are not driven by potential differences between rural and urban areas, an indicator variable was included for urban communities.[19] Not surprisingly, this coefficient is positive, indicating higher diversification in urban areas. Despite significant differences in the descriptive analysis in Section 4, among the infrastructural variables only access to credit has a significant (positive) effect on household income diversification in the main models. This positive effect would support the hypothesis that access to credit in the 1990s may have provided households with opportunities to diversify rather than only allowing them to smooth consumption.

5.2 Income diversification in the long run

As indicated at the outset, one of our main objectives is to examine the difference between the predicted values of the diversification index, based on the 1991–94 sample, and the actual 2004 index values. The actual index averages in 2004 and predicted values of the diversification index, as well as their differences, are reported in Table 5 for all empirical specifications underlying the empirical analysis in Table 4. The predicted values are calculated using the estimated coefficients in Table 4 on the 2004 data. The differences are all statistically significant at the 1 per cent level and positive in Models 2–5 and 7 and negative in Models 1 and 6. There is more variation in the prediction models that include instrumental and infrastructural variables. Models 2–5 and 7 suggest that the actual level of diversification is higher than it would have been if nothing in the economy had changed in the 10-year period. However, models 1 and 6 suggest the opposite: when we include the significant changes in infrastructural factors, actual diversification is higher than it would have been if nothing in the economy had changed in the 10-year period.

Table 5: Comparing predicted and actual income diversification index values

		Income diversification index					Conf. intervals	
		Mean	SD	Predicted	SD	Diff.	Lower	Upper
Model 1	OLS + Inf.	0.294	0.241	0.339	0.107	-0.045	0.331	0.347
Model 2	FE	0.294	0.241	0.261	0.074	0.032	0.256	0.267
Model 3	RE	0.294	0.241	0.228	0.086	0.066	0.221	0.234
Model 4	FE + IV	0.294	0.241	0.143	0.364	0.151	0.116	0.169
Model 5	RE + IV	0.294	0.241	0.089	0.495	0.204	0.060	0.119
Model 6	FE + IV + Inf.	0.294	0.241	0.337	0.378	-0.043	0.310	0.365
Model 7	RE + IV + Inf.	0.294	0.241	0.125	0.406	0.169	0.095	0.154

Note: number of observations (households) is 725. Mean income diversification index is calculated using data from Wave 5 and the predicted income diversification index values are calculated using the estimated coefficients in Table 4 on data from Wave 5, where models 1–7 refer to columns 1–7. Inf. stands for infrastructural variables.

Source: authors' calculations based on KHDS data, Waves 1–4 and 5.

Figure 2 illustrates both the actual (2a, blue bar) and predicted (red bar) diversification index values using Model 6[20] and changes (2b) over the 10-year period by income percentile in 1991–94. Although the actual levels of diversification in 2004 are somewhat higher among richer than among poorer households, relatively poorer households are characterized by very similar actual diversification levels when compared with their 'intrinsic' counterparts. Households above the median, however, are characterized by lower levels of 'real' diversification than the model predicted, given the improvements in observed infrastructure. For this group we observe negative

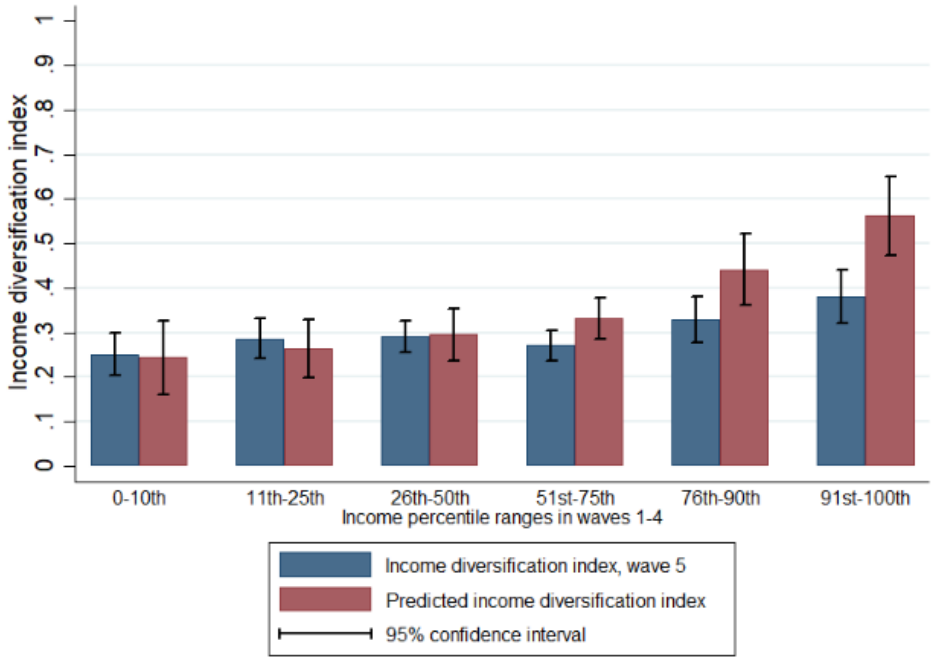
[19] We also estimated the equations controlling for district fixed effects, but this did not affect our results significantly.

[20] Model 6 is the most plausible, as it controls for endogeneity, considers unobserved household attitudes (household fixed effects), and incorporates crucial infrastructural improvements.

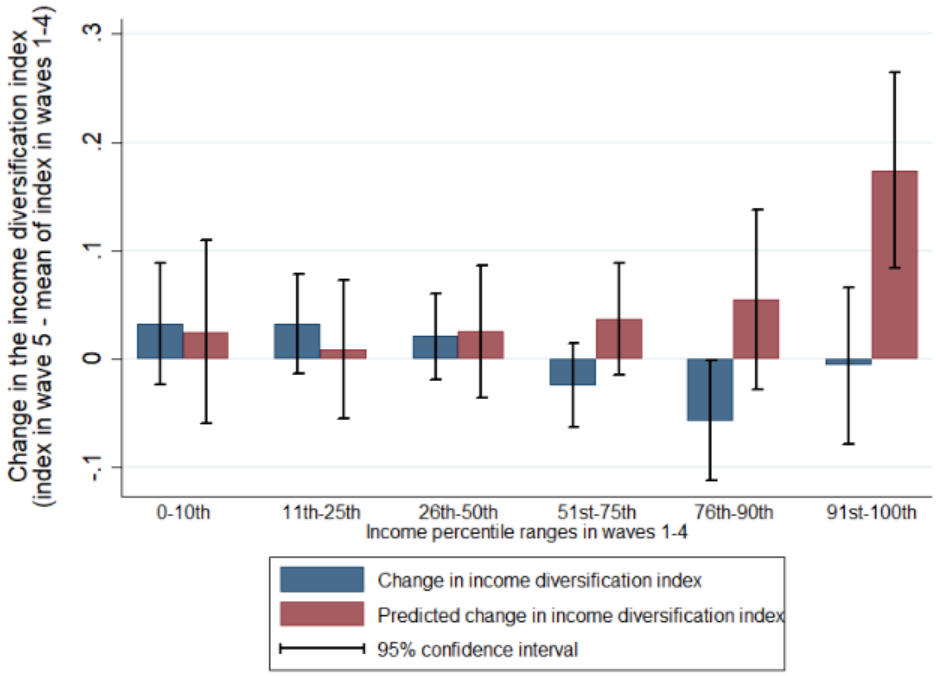
values in the actual changes, while the predicted changes are positive. In sum, and keeping in mind that our baseline analysis indicated a diversification pattern consistent with the accumulation hypothesis, we observe very similar levels of both 'real' and 'intrinsic' livelihood diversification among poorer households, while we observe a failure to fulfil the 'intrinsic' potential in diversification among richer households.

Figure 2: Comparing predicted (model 6) and actual income diversification index values by income percentiles

2a: Absolute value comparison



2b: Change in value comparison



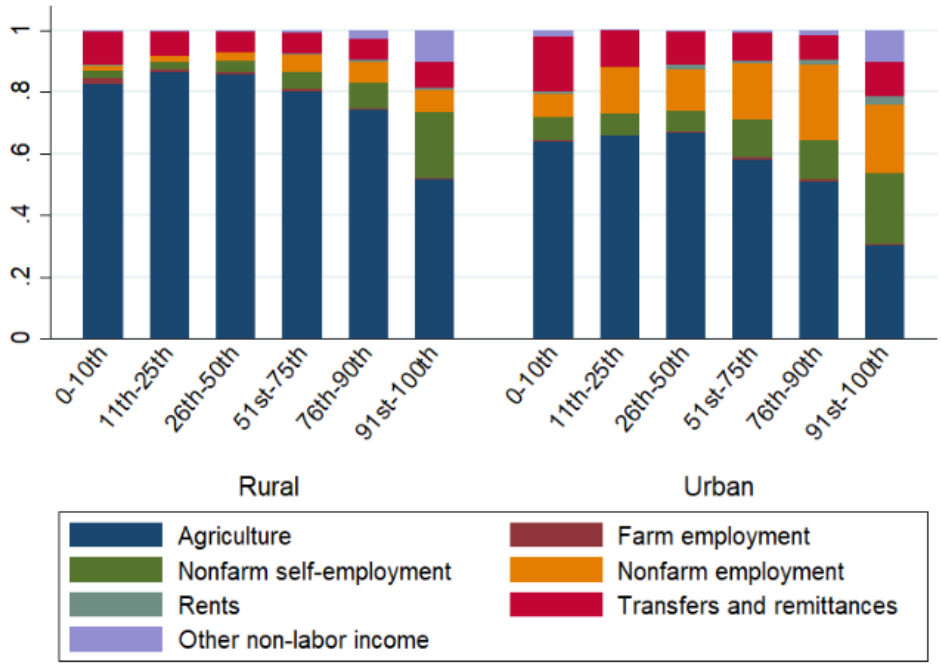
Source: authors' illustrations based on KHDS data, Waves 1–4 and 5.

We continue this sub-section by describing diversification patterns over time and income group to understand developments in the long run. Then we describe how improved access to infrastructure correlates with household income diversification, number of income sources used, and household income. Last, we show whether the household decision to diversify, measured by entering different income-generating activities, is related to either survival or accumulation motives.

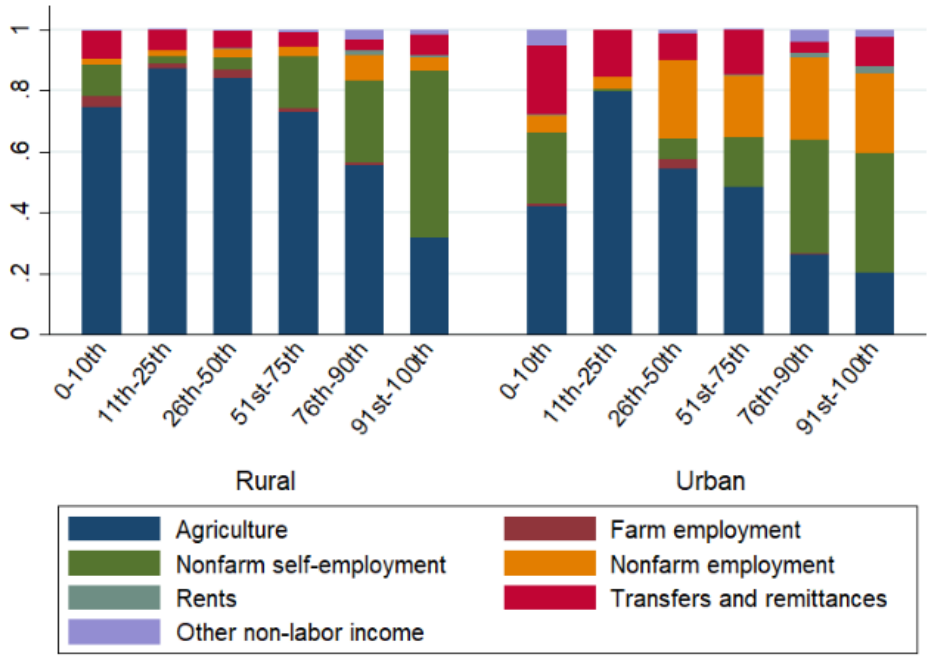
Figures 3a and 3b show diversification patterns over time and income group and across rural and urban areas. The figures show that inequality in diversification across the income distribution has increased. Simultaneously, however, the diversification patterns in rural areas have become more similar to those in urban areas. In general, the agricultural income share (blue) has decreased over time, while off-farm employment and self-employment have increased. In the 1990s, higher-income households in urban areas engaged considerably more in private non-farm businesses (green) and non-farm employment (yellow) than rural households or lower percentiles in urban areas. In 2004, in both urban and rural areas, the better-off half and the poorest 10th percentile were further diversifying into private non-farm businesses (green). The increase in this respect is noteworthy among the top and bottom percentiles: income share coming from private businesses (green) more than doubled in a decade for the top 10th percentile (for rural areas the growth was 32 percentage points (from 21 per cent to 53 per cent) compared with a 16 percentage point increase in urban areas (from 23 per cent to 39 per cent)), while for the bottom percentile it quadrupled (for rural areas the share grew from 2 per cent to 10 per cent and for urban areas from 7 per cent to 25 per cent).

Figure 3: Income shares by income percentile ranges and urban/rural classification

3a: Waves 1–4



3b: Wave 5



Source: authors' illustrations based on KHDS data.

Second, from a policy perspective it is most interesting to consider the factors over which the government has control. In Table 6 we explore how a positive change in access to key infrastructural factors correlates with diversification, number of income sources, and household income in Wave 5. The results are reported for the sample as a whole and separately for poorer households (those that had income levels below the median in the 1990s panel) to understand better the effects on the poor. For both samples, we find a positive and significant correlation between levels of diversification and gaining access to public transport and a daily market, and a negative significant correlation between diversification and gaining access to credit, and secondary school. Access to public transport and a daily market may have facilitated diversification into off-farm employment or self-employment. The negative relationship observed for the pooled sample between income diversification and improved access to credit facilities can be interpreted as evidence of credit being a factor that enables households to smooth consumption inter-temporally, rather than helping them diversify their income portfolios. A possible explanation for the negative correlation between access to secondary school and income diversification (columns 1 and 4) is that the school reform in Tanzania that happened just a few years before Wave 5 may have decreased household child labour and the number of available income sources (columns 2 and 5). Based on these results, infrastructural development appears to have had some effect on household income diversification strategies in the Kagera region over the survey period.

Last, regarding correlations with income, there is a significant positive relationship with improved access to electricity, post and telecommunication, public transport, and health facility. This correlation suggests that the income-generating abilities of the surveyed households improved simultaneously with the introduction of this infrastructure.

Table 6: Correlations between household diversification measures and income and positive change in infrastructural factors

	Pooled sample			Poor households (income < median)		
	Income diversification index	No. of income sources	Log of household income PAE	Income diversification index	No. of income sources	Log of household income PAE
	(1)	(2)	(3)	(4)	(5)	(6)
Gained access to						
Credit	-0.051**	-0.063	0.106	-0.011	0.042	-0.014
	(0.025)	(0.066)	(0.187)	(0.041)	(0.107)	(0.294)
Electricity	0.005	-0.004	-0.134	0.014	0.030	-0.104
	(0.019)	(0.050)	(0.142)	(0.027)	(0.072)	(0.198)
Post and telecom	0.002	0.010	0.079	0.025	0.040	-0.153
	(0.020)	(0.054)	(0.151)	(0.031)	(0.080)	(0.221)
Public transport	0.064***	0.172***	0.277**	0.070**	0.183**	0.109
	(0.019)	(0.050)	(0.140)	(0.028)	(0.073)	(0.201)
Daily market	0.098*	0.304**	-0.109	0.072	0.190	-0.072
	(0.052)	(0.139)	(0.392)	(0.090)	(0.236)	(0.651)
Secondary school	-0.053**	-0.160***	-0.479***	-0.097***	-0.259***	-0.284
	(0.022)	(0.057)	(0.161)	(0.033)	(0.087)	(0.241)
Health facility	-0.010	-0.009	0.094	-0.005	0.025	0.078
	(0.018)	(0.049)	(0.137)	(0.027)	(0.070)	(0.194)
Observations	725	725	725	345	345	345

Note: columns 1–3 show correlations for the full sample and columns 4–6 for the households below median income in the sample. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Source: authors' estimations based on KHDS data.

Third, we consider whether the increase in diversification observed in Figure 3—and particularly entrance into non-agricultural self-employment—among the poorer half of the households was driven by higher return opportunities or out of necessity (pull or push factors). Table 7 presents the results from a difference-in-differences estimation for both the poorer (a) and richer (b) halves, aiming to measure the effect of changes in diversification strategies. For each income activity a dummy variable indicates whether a household entered a new activity that generates income in Wave 5. In other words, the dummy is equal to 1 if in Waves 1–4 a household had on average no or less than 10 per cent of its income from a given source and in Wave 5 more than 10 per cent of its income from this source, and 0 otherwise. While this 'treatment' is endogenous, we do not claim causal inference but instead interpret the estimates as indications of a positive or negative relationship between entering different income activities and household income.

The first row in both tables shows that there is a significant negative time effect in all regressions, indicating that real incomes were lower in 2004 than in the 1990s. The second row indicates the selection bias, that is the income difference of households choosing to enter the various income activities. In Table 7a, column (2) the selection bias is significant and negative. This estimate indicates that households that entered non-agricultural employment were on average poorer than those that did not; whereas, among the wealthier half (Table 7b), columns (3) and (6) have a positive significant estimate: households entering farming and other non-labour income were on average richer than those that did not enter these activities. The third row is the difference-in-differences estimate. This estimate is negative and significant in Table 7a for columns (1) and (4), indicating that households that entered agricultural employment or started receiving transfers or remittances had on average lower income in Wave 5 than other households among the population below the median income.

Table 7a: Impact of households with below median incomes changing their diversification strategies from 1990s to 2004 on their PAE income using DID.

	(1)	(2)	(3)	(4)	(5)
Wave 5	-0.379***	-0.424***	-0.503***	-0.374***	-0.415***
	(0.058)	(0.058)	(0.061)	(0.058)	(0.057)
Enter	-0.110	-0.119	-0.012	0.118	0.247
	(0.137)	(0.102)	(0.067)	(0.106)	(0.262)
Wave 5 * **Enter**	-0.594**	0.328	0.590***	-0.510**	1.155**
	(0.288)	(0.221)	(0.145)	(0.227)	(0.536)
Observations	2,009	2,009	2,009	2,009	2,009
Controls	Yes	Yes	Yes	Yes	Yes
R-squared	0.075	0.072	0.080	0.073	0.075
Enter =	Entered agricultural employment	Entered non-agricultural employment	Entered non-agricultural self-employment	Started receiving transfers and remittances income	Started receiving other non-labour income
% share entering new activity (Enter=1)	3.27	5.62	14.4	5.18	0.83

Note: outcome variable: Log of household income PAE. Changes to diversification in activities where there were less than 15 observations (rents and agriculture) are not significant and not presented. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Source: authors' calculations using DID regression based on KHDS data, Waves 1–5.

Table 7b: Impact of households with above median incomes changing their diversification strategies from 1990s to 2004 on their PAE income using DID.

	(1)	(2)	(3)	(4)	(5)	(6)
Wave 5	-0.776***	-0.812***	-0.776***	-0.887***	-0.779***	-0.790***
	(0.060)	(0.061)	(0.060)	(0.063)	(0.062)	(0.060)
Enter	-0.185	0.087	-0.124	-0.014	0.070	0.257
	(0.200)	(0.132)	(0.292)	(0.081)	(0.103)	(0.201)
Wave 5 * **Enter**	-0.425	0.577**	-0.878	0.843***	-0.086	0.227
	(0.443)	(0.285)	(0.622)	(0.178)	(0.221)	(0.416)
Observations	1,986	1,986	1,986	1,986	1,986	1,986
R-squared	0.169	0.170	0.169	0.179	0.167	0.169
Enter =	Entered agricultural employment	Entered non-agricultural employment	Entered agricultural self-employment	Entered non-agricultural self-employment	Started receiving transfers and remittances income	Started receiving other non-labour income
Share entering	1.50	3.59	0.85	10.0	6.08	1.74

Note: outcome variable: log of household income PAE. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Source: authors' calculations using DID regression based on KHDS data, Waves 1–5.

For the other income-generating activities, there are positive effects, but they are only significant in columns (3) and (5), meaning that households that entered non-agricultural self-employment (private businesses) or received non-labour income such as a pension, inheritance, or marriage payments were on average better off in Wave 5 than other households. For the wealthier half in Table 7b, columns (2) and (4) give a positive significant estimate, suggesting that households entering non-agricultural employment and starting a private business outside agriculture were on average better off than others.

Hence, it seems that for households that diversified their income source portfolio over time, there were two types of trajectory. There were those who were most likely pushed into employment or had to rely on transfers from migrated family members. These households did not experience

income improvements over time from diversifying their portfolio on average. The other trajectory resulted (on average) in success stories among those who managed to enter non-farm self-employment (among both the richer and poorer halves) or start new enterprises outside agriculture (among the richer half).

6 Conclusion

Most rural households in developing countries, and especially in Sub-Saharan Africa, have highly diversified portfolios. While several studies have examined the determinants of household income diversification in developing economies, the causes of household income diversification remain unclear. In particular, it remains an empirical issue whether household income diversification is a consequence of 'push' or 'pull' factors. In this paper, we first examined the determinants of income diversification for a panel of rural households in Tanzania for the period 1991–94. We then described diversification patterns in the long run and showed how improved access to certain infrastructure is linked with more income diversification and higher incomes. Then we studied the diversification push and pull factors using a difference-in-differences analysis.

One of our core findings is that that the diversification behaviour of households in rural Tanzania is on average driven by accumulation motives rather than by survival concerns. In other words, richer households are in a better position to diversify their incomes than relatively poor households. Although the greater income diversification capacity of relatively richer households observed in the 1990s persists in 2004, there are also positive impacts among poorer households. We find that poorer households are motivated to diversify by both push and pull factors, where entrance into off-farm self-employment has a strong positive impact on household income, but entrance into employment or dependence on transfers and remittances is negatively correlated with household income.

We further find that gaining access to infrastructure such as public transport is positively correlated with income diversification. Infrastructural improvements that increase access to electricity, post and telecommunications, public transport, and health facilities are positively correlated with income, suggesting that the income-generating abilities of the surveyed households improved simultaneously with the introduction of the infrastructural improvements.

Our results call into question the pessimistic view of rural Africa that has shaped the academic discourse and policy discussions on household income diversification for many decades (for example, see Bryceson 2005). In contrast to the widely held belief that household income diversification is a symptom of African depeasantization and of a failing agricultural sector, our paper suggests that household income diversification may well be a choice, and not a necessity, as both poorer and richer rural households use the capital that they have generated from agriculture to move into profitable non-agricultural activities. However, our study also suggests that asset and poverty traps may develop among those rural households that are not able to make the transition into such activities. In this case, specific policy measures may be needed to allow rural poor households to generate agricultural income and make the transition to diversified portfolios, including public investment in infrastructure and easier access to rural financial institutions and rural markets.

References

Abdulai, A., and A. CroleRees (2001). 'Determinants of Income Diversification Amongst Rural Households in Southern Mali'. *Food Policy*, 26: 437–52. https://doi.org/10.1016/S0306-9192(01)00013-6

Alderman, H., and C. Paxson (1992). 'Do the Poor Insure? A Synthesis of the Literature on Risk and Consumption in Developing Countries'. Working Paper WPS 1008. Washington DC: The World Bank, Agricultural and Rural Development Department.

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.

Attanasio, O., and S. Krutikova (2020). 'Consumption Insurance in Networks with Asymmetric Information: Evidence from Tanzania'. *Journal of the European Economic Association*, 18(4): 1589–618. https://doi.org/10.1093/jeea/jvaa030

Banerjee, A.V., and E. Duflo (2007). 'The Economic Lives of the Poor'. *Journal of Economic Perspectives*, 21(1): 141–68. https://doi.org/10.1257/jep.21.1.141

Barrett, C.B., and T. Reardon (2000). 'Asset, Activity, and Income Diversification among African Agriculturalists: Some Practical Issues'. Working Paper 14734. Ithaca, NY: Cornell University, Department of Applied Economics and Management. https://doi.org/10.2139/ssrn.257344

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 (2001a). 'Nonfarm Income Diversification and Household Livelihood Strategies in Rural Africa: Concepts, Dynamics and Policy Implications'. *Food Policy*, 26: 315–31. https://doi.org/10.1016/S0306-9192(01)00014-8

Barrett, C.B., M. Bezuneh, and A. Aboud (2001b). 'Income Diversification, Poverty Traps and Policy Shocks in Côte d'Ivoire and Kenya'. *Food Policy*, 26(4): 367–84. https://doi.org/10.1016/S0306-9192(01)00017-3

Basu, A., and R. Dimova (2020). 'Household Behavioral Preferences and the Child Labor–Education Trade-off: Framed Field Experimental Evidence from Ethiopia'. IZA DP 13011. Bonn: Institute of Labour Economics.

Basu, A., R. Dimova, M. Gbakou, and R. Viennet (2020). 'Parental Behavioural Preferences and Educational Progressions of Girls and Boys: Lab-in-the-Field Evidence from Rural Côte d'Ivoire'. Mimeo. Ithaca, NY: Cornell University; Manchester: Manchester University.

Beegle, K., J. De Weerdt, and S. Dercon (2011). 'Migration and Economic Mobility in Tanzania: Evidence from a Tracking Survey'. *Review of Economics and Statistics*, 93(3): 1010–33. https://doi.org/10.1162/REST_a_00105

Besley, T. (1995). 'Savings, Credit and Insurance'. In J. Behrman and T.N. Srinivasan (eds), *Handbook of Development Economics*. Amsterdam: Elsevier Science. https://doi.org/10.1016/S1573-4471(05)80008-7

Binswanger, H., and M. Rosenzweig (1993). 'Wealth, Weather Risk and the Composition and Profitability of Agricultural Investments'. *The Economic Journal*, 103: 56–78. https://doi.org/10.2307/2234337

Block, S., and P. Webb (2001). 'The Dynamics of Livelihood Diversification in Post-Famine Ethiopia'. *Food Policy*, 26: 333–50. https://doi.org/10.1016/S0306-9192(01)00015-X

Bryceson, D. (2005). 'Rural Livelihoods and Agrarian Change in Sub-Saharan Africa: Processes and Policies'. In F. Ellis and H.A. Freeman (eds), *Rural Livelihoods and Poverty Reduction Policies*, pp. 48–61. London: Routledge.

Davis, B., P. Winters, G. Carletto, K. Coarrubias, W. Quinones, A. Zezza, K. Stamoulis, C. Azzarri, and S. DiGiuseppe (2010). 'A Cross-country Comparison of Rural Income-Generating Activities'. *World Development*, 38(1): 48–63. https://doi.org/10.1016/j.worlddev.2009.01.003

Deaton, A. (1992). *Understanding Consumption*. Oxford: Clarendon Press. https://doi.org/10.1093/0198288247.001.0001

de Bont, C., H. Komakech, and G. Veldwisch (2019). 'Neither Modern Nor Traditional: Farmer-led Irrigation Development in Kilimanjaro Region, Tanzania'. *World Development*, 116: 15–27. https://doi.org/10.1016/j.worlddev.2018.11.018

Dedehouanou, S.F.A., and J. McPeak (2020). 'Diversify More or Less? Household Income Generation Strategies and Food Security in Rural Nigeria'. *The Journal of Development Studies*, 56(3): 560–77. https://doi.org/10.1080/00220388.2019.1585814

Dercon, S. (1996). 'Risk, Crop Choice and Savings: Evidence from Tanzania'. *Economic Development and Cultural Change*, 44(3): 484–513. https://doi.org/10.1086/452229

Dercon, S. (1998). 'Wealth, Risk and Activity Choice: Cattle in Western Tanzania'. *Journal of Development Economics*, 55: 1–42. https://doi.org/10.1016/S0304-3878(97)00054-0

Dercon, S. (2002). 'Income Risk, Coping Strategies and Safety Nets'. *World Bank Research Observer*, 2(17): 141–66. https://doi.org/10.1093/wbro/17.2.141

Dercon, S., and P. Krishnan (1996). 'Income Portfolios in Rural Ethiopia and Tanzania: Choices and Constraints'. *Journal of Development Studies*, 55: 1–42.

De Weerdt, J. (2010). 'Moving out of Poverty in Tanzania: Evidence from Kagera'. *Journal of Development Studies*, 46(2): 331–49. https://doi.org/10.1080/00220380902974393

Dimova, R., and K. Sen (2010). 'Is Household Income Diversification a Means of Survival or a Means of Accumulation? Panel Data Evidence from Tanzania'. BWPI Working Paper 122. Manchester: Brooks World Poverty Institute, University of Manchester. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1688433 (accessed 21 December 2020).

Dimova, R., S. Gangopadhyay, K. Michaelowa, and A. Weber (2015). 'Off-Farm Labor Supply and Correlated Shocks. New Theoretical Insights and Evidence from Malawi'. *Economic Development and Cultural Change*, 63(2): 361–91. https://doi.org/10.1086/679193

Djido, A., and B. Shiferaw (2018). 'Patterns of Labor Productivity and Income Diversification: Empirical Evidence from Uganda and Nigeria'. *World Development*, 105: 416–27. https://doi.org/10.1016/j.worlddev.2017.12.026

Dumas, C., and X. Játiva (2020). 'Better Roads, Better Off? Evidence on Improving Roads in Tanzania'. FSES Working Paper 518. Faculty of Economics and Social Sciences, University of Freiburg/Fribourg, Switzerland.

Eakin, H. (2005). 'Institutional Change, Climate Risk and Rural Vulnerability: Cases from Central Mexico'. *World Development*, 33(11): 1923–38. https://doi.org/10.1016/j.worlddev.2005.06.005

Ellis, F. (1998). 'Household Strategies and Rural Livelihood Diversification'. *Journal of Development Studies*, 35(1): 1–38. https://doi.org/10.1080/00220389808422553

Ellis, F. (2000a). 'The Determinants of Rural Livelihood Diversification in Developing Countries'. *Journal of Agricultural Economics*, 51(2): 289–302. https://doi.org/10.1111/j.1477-9552.2000.tb01229.x

Ellis, F. (2000b). *Mixing It: Rural Livelihoods and Diversity in Developing Countries*. Oxford: Oxford University Press.

Johny, J., B. Wichmann, and B. Swallow (2017). 'Characterizing Social Networks and Their Effects on Income Diversification in Rural Kerala, India'. *World Development*, 94: 375–92. https://doi.org/10.1016/j.worlddev.2017.02.002

Joshy, P., A. Gulati, P. Birthal, and L. Tewari. (2004). 'Agriculture Diversification in South Asia: Patterns, Determinants and Policy Implications'. *Economic and Political Weekly*, 39(24): 2457–67.

Khan, R., and O. Morrissey (2020). 'Income Diversification and Household Welfare in Tanzania 2008–2013'. WIDER Working Paper 2020/110. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2020/867-2

Litchfield, J., and T. McGregor (2008). 'Poverty in Kagera, Tanzania: Characteristics, Causes and Contributions'. PRUS Working Paper 42. Brighton: University of Sussex, Poverty Research Unit. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.177.4777&rep=rep1&type=pdf (accessed 14 December 2020).

Menon, J. (2009). 'Rainfall Uncertainty and Occupational Choice in Agricultural Households of Rural Nepal'. *Journal of Development Studies*, 46(6): 864–88. https://doi.org/10.1080/00220380902807387

Morduch, J. (1995). 'Income Smoothing and Consumption Smoothing'. *Journal of Economic Perspectives*, 9: 103–14. https://doi.org/10.1257/jep.9.3.103

Nyyssölä, M., P. Rattenhuber, G. Wright, E. Kisanga, and H. Barnes (forthcoming). 'A Gendered Analysis of the Welfare Impact of the Tax–Benefit Systems in Three Sub-Saharan African Countries'. WIDER Working Paper. Helsinki: UNU-WIDER.

Pesaran M.H., and Q. Zhou (2018). 'Estimation of Time-Invariant Effects in Static Panel Data Models'. *Econometric Reviews*, 37(10): 1137–71. https://doi.org/10.1080/07474938.2016.1222225

Platteau, J-P., and A. Abraham (1987). 'An Inquiry into Quasi-Contracts: the Role of Reciprocal Credit and Interlinked Deals in Small-Scale Fishing Communities'. *Journal of Development Studies*, 23: 461–90. https://doi.org/10.1080/00220388708422044

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.

Sen, K. (2002). 'Economic Reforms and Rural Livelihood Diversification in Tanzania'. LADDER Working Paper 12. London: Department for International Development. Available at: https://assets.publishing.service.gov.uk/media/57a08d4440f0b652dd001884/Ladder-wp12.pdf (accessed 14 December 2020).

Smith, B., and J.B. Wilson (1996). 'A Consumer's Guide to Evenness Indices'. *Oikos*, 76(1): 70–82. https://doi.org/10.2307/3545749

Townsend, R. (1994). 'Risk and Insurance in Village India'. *Econometrica*, 62: 539–91. https://doi.org/10.2307/2951659

Udry, C. (1994). 'Risk and Insurance in a Rural Credit Market: An Empirical Investigation in Northern Nigeria'. *Review of Economic Studies*, 61: 495–526. https://doi.org/10.2307/2297901