zef
Center for
Development Research
University of Bonn

ZEF-Discussion Papers on
Development Policy No. 357

Jonas Guthoff, Chrispinus Mutsami, Martin Parlasca, Thomas Heckelei and
Matin Qaim

Aspirations and Livelihood Diversification in Rural Africa

Bonn, December 2024

The **Center for Development Research (ZEF)** was established in 1995 as an international, interdisciplinary research institute at the University of Bonn. Research and teaching at ZEF address political, economic and ecological development problems. ZEF closely cooperates with national and international partners in research and development organizations. For information, see: www.zef.de.

ZEF – Discussion Papers on Development Policy are intended to stimulate discussion among researchers, practitioners, and policy makers on current and emerging development issues. The papers are not peer-reviewed. They reflect work in progress and should be regarded as preprints.

Jonas Guthoff, Chrispinus Mutsami, Martin Parlasca, Thomas Heckelei, and Matin Qaim, Aspirations and Livelihood Diversification in Rural Africa, ZEF – Discussion Papers on Development Policy No. 357, Center for Development Research, Bonn, December 2024, pp. 64.

ISSN: 1436-9931

Published by:
Zentrum für Entwicklungsforschung (ZEF)
Center for Development Research
Genscherallee 3
D – 53113 Bonn
Germany
Phone: +49-228-73-1861
Fax: +49-228-73-1869
E-Mail: zef@uni-bonn.de
www.zef.de

Aspirations and Livelihood Diversification in Rural Africa

Jonas Guthoff *[1], Chrispinus Mutsami[1], Martin Parlasca[1], Thomas Heckelei[2], and Matin Qaim[1,2]

[1]Center for Development Research (ZEF), University of Bonn, Germany
[2]Institute for Food and Resource Economics, University of Bonn, Germany

December 6, 2024

Abstract

Many people in rural Africa still rely on smallholder agriculture as their main source of livelihood. Yet pursuing other economic activities beyond farming to diversify incomes becomes increasingly important. In this paper, we focus on the role of people's income aspirations in the process of diversification. We first show that in settings where wage employment opportunities are scarce, income diversification is linked to higher levels of income. We test the theoretical prediction that aspirations that are moderately ahead, but not too far ahead of one's current income level, serve as best incentives for future-orientated behavior, such as income diversification. We leverage unique household panel data from four African countries to test this hypothesis, finding strong and robust evidence of an inverted U-shaped relationship between income aspirations and indicators of income diversification. Our findings underpin the importance of behavioral factors as channels for economic development and poverty reduction.

Keywords: Aspirations, Income diversification, Household welfare, Poverty
JEL Codes: D91, I39, O12, Q12.

*Corresponding author: jguthoff@uni-bonn.de. We acknowledge financial support from the German Research Foundation (DFG) through funding for the project "Job Futures" as part of the Collaborative Research Center "Future Rural Africa" (Project-ID 328966760—TRR 228). We also acknowledge funding support from the German Academic Exchange Service (DAAD). We thank Bisrat Gebrekidan, Jessy Amarachi Ezebuihe, Vanessa Amatotsero, Paula Alexiou, Eva Bacud, Hanna Ihli, Jackson Elias and enumerators from the four countries for their support during data collection.

1 Introduction

More than 60 percent of the population in sub-Saharan Africa relies on small-scale agriculture as their main source of livelihood (FAO, 2021). However, when households do not have access to secure and lucrative employment opportunities and limited inputs, technology, or market linkages impede specialization in agriculture, diversification of income sources can increase overall income levels, wealth, and resilience to shocks (Khan and Morrissey, 2023; Block and Webb, 2001). As a result, more and more households in sub-Saharan Africa also derive income from sources other than agriculture (De Janvry and Sadoulet, 2001; Barrett *et al.*, 2001; Musumba *et al.*, 2022). With increasing likelihoods of extreme weather events severely affecting agricultural yields and increasing uncertainty in food production and prices, relying on diverse income sources will likely become even more important in the future (Mehraban *et al.*, 2021; Musungu *et al.*, 2024).

Drivers of income diversification have received much attention in recent decades. In this regard, household-level socio-economic characteristics, particularly education, access to institutions, and location have been shown to be crucial (De Janvry and Sadoulet, 2001; Reardon *et al.*, 2007; Reardon, 1997). Yet, much less attention has been given to behavioral- or psychological constraints to diversification (Kremer *et al.*, 2019). In this paper, we address this research gap and analyze how diversification of income sources of rural dwellers in sub-Saharan Africa is related to their aspirations towards future incomes.

Theory suggests that aspirations serve as good incentives to engage in future-oriented behavior, if they are moderately above the current status (Genicot and Ray, 2017; Ray, 2006). Aspirations that are either too low or too high may lead to fatalism or frustration - in other words aspiration failure. This inverted U-shape relationship suggests that aspirations can be both catalyst or barrier to future-oriented behavior and, thereby, to development processes.

The objective of this paper is to test empirically for an inverted U-shaped relationship between income aspirations and income diversification. To do so, we use panel data from 2,360 households collected in 2019 and 2023 across Kenya, Namibia, Tanzania, and Zambia. We construct an income aspiration gap metric and test for the inverted

U-shaped relationship between three main indicators of income diversification, namely whether a household has diversified income sources (extensive margin), the number of income sources (intensive margin), as well as Simpson's Diversity Index (SDI). We employ various panel data models with household fixed effects to reduce potential issues of unobserved heterogeneity. Taken together, we find strong evidence across all indicators of income diversification that aspirations that are ahead, but not too far ahead, serve as best incentives for future-orientated behavior. Results from semi-parametric estimations suggest that rural dwellers have, on average, too ambitious income aspirations, possibly leading to frustration, inaction, and, consequently, lower than optimal levels of diversification.

This paper adds to the literature on drivers of income diversification, which are often classified as push- or pull-factors (Barrett *et al.*, 2001; Dercon, 2002; Musumba *et al.*, 2022). Pull factors are opportunity-driven and often involve easy-to-enter non-farm activities, such as sales and trade (Nagler and Naudé, 2017). Push factors are survival-led based on experienced or anticipated shocks and require an active process to maintain and continuously adapt a portfolio of activities to improve living standards and secure survival (Ellis, 2000; Alobo Loison, 2015). The success of this process is most often linked to endowments such as education or assets. We find additional evidence for the relevance of behavioral factors such as aspirations.

This paper further complements the recent empirical work on the relationship between aspirations and investment decisions either in agricultural (Bloem, 2021; Villacis *et al.*, 2023; Tabe-Ojong *et al.*, 2023, 2024) or human capital investments (Janzen *et al.*, 2017; Ross, 2019).

With this paper, we make three main contributions to the existing literature: First, we contribute to a set of studies that analyzes determinants of income diversification by considering the role of aspirations in this context. Second, we contribute to the emerging literature that tests the theory of aspirations by leveraging panel data which allows controlling for confounding factors and time effects. To the best of our knowledge, all of the existing papers use cross-sectional data. Third, our unique household panel was collected in four distinct regions across four different countries in sub-Saharan Africa, which offers new insights into potential differences of aspirations across contexts and increases the external validity of our findings, making the results applicable beyond a single country context.

This paper is structured as follows: In Section 2, we detail the conceptual framework and link aspirations to the decision of income diversification. In Section 3, we provide an overview of the data collected, our measures of income diversification and aspirations. Section 4 explains our estimation strategy, in Section 5 we present and discuss our results, before additional robustness tests are presented in Section 6. We conclude in Section 7.

2 Conceptual framework

Genicot and Ray (2017) building on Ray (2006) formulate a model of socially determined aspirations that can serve as individual incentives to invest in future-orientated behavior anchored around goals. The theory suggests that future investments depend on an individual's aspirations relative to their current status. If aspirations slightly exceed the current status, required investments are low. Ray (2006) defines these low investments as aspiration failure. With an increase in aspirations, failure is averted. However, if aspirations are too high above the current status, the required investments become too large and burdensome leading them again to fail. Theory therefore predicts that aspirations that are moderately above the current standard tend to serve as good incentives, while aspirations that are either too close or too far from the current status may lead to frustration or resignation (Dalton *et al.*, 2016; Genicot and Ray, 2017; Ray, 2006). This implies a non-monotonic relationship between aspirations and investments.

Summarized further by Janzen *et al.* (2017) and Villacis *et al.* (2023), Genicot and Ray (2017)'s model conceptualizes aspirations through an inter-temporal utility function with a reference point (Kahneman and Tversky, 2013). Individuals obtain a bonus utility when they realize an outcome above their aspired level and are frustrated if they realize an outcome below their aspired level. This allows the model to be appropriately applied to a future-oriented, long-term investment decisions such as income diversification – in our rural setting in sub-Saharan Africa. Empirical evidence suggests that for most individuals and households, income diversification is superior to classical on-farm or off-farm specialization, particularly in contexts where credit markets that would allow for productivity enhancing on-farm investments are either poorly developed or non-existent, and wage employment opportunities are limited. We will also present strong evidence for this in our setting. Income diversification does not necessarily benefit everyone, but in a context, where secure wage-employment is scarce, it is typically linked with higher levels of wealth. Therefore the decision to invest time, capital, and knowledge into an additional income-generating activity can be seen as a type of future-oriented investment decision.

We test for an inverted U-shaped relationship between aspirations and several indicators of income diversification. In detail, we investigate the relationship between the income aspiration gap, constructed as the distance between current and aspired levels of income, and three indicators of income diversification, namely whether a household has more than one source of income (extensive margin), the number of income sources (intensive margin), and Simpson's Diversity Index (SDI).

3 Data collection and measurement

We leverage primary panel data, collected in two waves across four distinct regions in Kenya, Namibia, Tanzania, and Zambia. Each study represents a different climatic, agroecological, economic, and institutional context to capture some of the diversity of rural settings in sub-Saharan Africa. In Kenya, households were sampled from Baringo County in the Rift Valley, in Namibia, from the Zambezi Region, in Tanzania from the regions of Iringa and Morogoro, and in Zambia, from the Southern part of Western Province. In each country, we relied on a two-stage stratified random sampling process, where in a first step enumeration areas and in a second step households in these areas were randomly selected. The samples from each country are therefore regionally representative.

Baringo County is sparsely populated with around 60 people per km^2, compared to a Kenyan average of 82. Baringo's poverty headcount rate was estimated to be around 47.5, which is above the national average of 38.6 (KNBS, 2019, 2023). The Zambezi region is less dense populated with 9.7 people per per km^2, but more dense compared to the national average with around 3.7 people (NSA, 2024). The region belongs to the poorest three regions in Namibia with a poverty headcount rate of 39.3 compared to 26.9 at national level (NPC, 2015). In Morogoro and Iringa, 45 and 34 people live per km^2, respectively, compared to a national average in Tanzania of 70 (MoFP-PED and NBS, 2024). Morogoro has an estimated poverty headcount rate of 19.3 and Iringa of 24 compared to a national average of 26.4 (URT, 2019). In Zambia, households were sampled from Mwandi, Sioma and Sesheke district which are all situated in the Southern part of Western Province, which with 10.9 people per km^2 is one of the least dense populous provinces in Zambia. The national average is 26.2 people per km^2 (ZamStats, 2024).

Data collection in each country was timed according to the agricultural calendar and took place between May and August in both waves, 2019 and 2023.[1]

Table 1 summarizes the sample across the four countries for 2019 and 2023, respectively. Across all countries, household heads are more likely to be men and married. There is some heterogeneity in the highest completed level of education across the four countries, with the highest education levels observed in Namibia and the lowest in Kenya, where close to 50 percent have no formal education at all. Household sizes average between 4.5 to 6.3 members across all countries and years. The dependency ratio, calculated as the number of individuals under 15 plus those over 65 divided by those of working age, was slightly below one in Tanzania and above one in the other countries. Values above one indicate that there are likely more economic dependent members inside a household, values less than one the reverse. [2] Cropland area owned varied from 0.4 hectares in Kenya to 2.5-2.8 hectares in Namibia and Zambia. In contrast, livestock ownership is highest in Kenya and Namibia. Attrition between the two rounds is modest. In total,

[1]See Section A.1 in the Appendix for details on the sampling procedure and field activities in each country.

[2]For those households, where there was no member of working population age, meaning in between 15 to 65 years we replace the dependency ratio with the sum of those below 15 and above 65.

we were able to re-interview 89% of the 2019 sample households again in 2023. Potential issues of differential attrition are discussed in more detail in Section 6.

We also elicited questions around respondent households' locus of control, which is often described as the belief or attitude about the causal relationship between one's own behavior and the consequences (Rotter, 1966). Broadly, there is a distinction between individuals who belief that their life's outcome is due to their own efforts (internal locus of control) and those who belief that their life depends on external factors, such as luck (external locus of control). Locus of control as other concepts from psychology have increasingly gained interest in many fields of applied economics. In labor economics, in particular, internal locus of control has been linked to human capital investments and a variety of other positive labor market outcomes (Cobb-Clark, 2015). We hypothesize that internal locus of control is also a potential factor that can influence long-term investment decisions, such as the engagement in income diversification.

	Kenya		Nambia		Tanzania		Zambia		Full Sample	
	2019	2023	2019	2023	2019	2023	2019	2023	2019	2023
Panel A: Characteristics of household head										
Age (in years)	45.52	48.46	51.43	53.22	48.85	52.15	47.83	52.51	48.43	51.49
	(15.30)	(15.73)	(17.73)	(15.97)	(15.07)	(15.12)	(16.36)	(17.79)	(16.16)	(16.05)
Male (0/1)	0.75	0.72	0.52	0.56	0.75	0.72	0.66	0.69	0.68	0.67
	(0.43)	(0.45)	(0.50)	(0.50)	(0.43)	(0.45)	(0.47)	(0.46)	(0.47)	(0.47)
Married (0/1)	0.73	0.70	0.56	0.57	0.73	0.70	0.69	0.66	0.68	0.66
	(0.44)	(0.46)	(0.50)	(0.50)	(0.44)	(0.46)	(0.46)	(0.47)	(0.47)	(0.47)
Years of schooling	4.54	4.59	7.38	7.83	5.97	6.10	6.26	5.88	5.99	6.09
	(5.03)	(5.13)	(4.49)	(4.33)	(2.94)	(3.29)	(4.14)	(3.92)	(4.28)	(4.35)
Highest level of education of household head										
No formal education (0/1)	0.47	0.48	0.18	0.14	0.14	0.15	0.21	0.20	0.25	0.24
	(0.50)	(0.50)	(0.39)	(0.35)	(0.35)	(0.36)	(0.41)	(0.40)	(0.43)	(0.43)
Some primary school education (0/1)	0.33	0.31	0.21	0.24	0.77	0.76	0.46	0.49	0.47	0.47
	(0.47)	(0.46)	(0.41)	(0.43)	(0.42)	(0.43)	(0.50)	(0.50)	(0.50)	(0.50)
Some secondary school education (0/1)	0.14	0.13	0.58	0.56	0.08	0.08	0.28	0.28	0.25	0.24
	(0.34)	(0.34)	(0.49)	(0.50)	(0.27)	(0.27)	(0.45)	(0.45)	(0.43)	(0.43)
Some post-secondary education (0/1)	0.06	0.08	0.03	0.06	0.01	0.01	0.05	0.03	0.04	0.04
	(0.24)	(0.27)	(0.17)	(0.23)	(0.11)	(0.12)	(0.22)	(0.17)	(0.19)	(0.20)
Panel B: Household characteristics										
Locus of control index (0-5)	3.76	4.30	3.66	4.12	3.84	4.18	3.24	4.16	3.67	4.20
	(1.10)	(0.99)	(1.19)	(0.98)	(1.09)	(1.06)	(1.48)	(1.04)	(1.20)	(1.02)
Household size	6.34	5.90	5.00	5.06	4.59	4.51	5.64	5.67	5.33	5.20
	(2.87)	(2.64)	(2.45)	(2.37)	(2.19)	(2.12)	(2.52)	(2.70)	(2.60)	(2.49)
Socio-dependancy ratio	1.49	1.38	1.17	1.11	0.95	0.99	1.53	1.19	1.24	1.16
	(1.16)	(1.07)	(1.23)	(1.04)	(0.86)	(0.86)	(1.33)	(0.96)	(1.15)	(0.99)
Cropland area (in ha)	0.42	0.51	2.81	2.58	1.39	1.38	4.81	2.47	2.04	1.62
	(0.45)	(0.58)	(3.10)	(2.98)	(0.98)	(1.02)	(4.48)	(2.03)	(2.87)	(1.99)
Area cultivated (in ha)	0.49	0.44	1.31	1.46	1.55	1.37	5.82	2.65	1.91	1.35
	(0.56)	(0.52)	(2.10)	(1.98)	(1.07)	(1.02)	(6.25)	(2.32)	(3.33)	(1.66)
Tropical Livestock Units (TLU)	5.66	5.23	5.04	6.32	0.96	1.74	2.14	3.86	3.39	4.13
	(11.60)	(10.59)	(11.88)	(14.94)	(3.73)	(12.43)	(4.03)	(6.71)	(9.02)	(12.05)
Asset index (sum)	2.71	2.69	4.56	5.05	3.52	4.81	3.08	3.73	3.49	4.13
	(2.21)	(2.29)	(2.46)	(2.70)	(2.10)	(2.30)	(2.43)	(2.63)	(2.37)	(2.64)
Attrited (0/1)		0.12		0.16		0.08		0.11		0.11
		(0.32)		(0.37)		(0.28)		(0.31)		(0.32)
Observations	704	704	652	652	871	871	437	437	2664	2664

Table 1: Socio-demographic Characteristics

Notes: This table presents means and standard deviations in parentheses for the full sample by country and year surveyed. We summarize under Panel A socio-demographic characteristics of the household head and under Panel B, characteristics of the household. The cropland area in hectares refers to the land size used or owned to grow crops, winsorized at the at the 95th percentile within country and year. The locus of control index is the sum of five binary variables detailed in Table A6. For the dependency ratio, there are 143 households (5.4 percent) in 2019 and 152 households (5.7 percent) in 2023 with zero members of working age. To obtain a non-missing dependency ratio, we treat the number of working members as 1. As a result, the dependency ratio for those households equals the sum of household members that are below 15 and above the age of 65 years.

In the following Sections 3.1 and 3.2, we discuss the construction of our main outcome variables capturing income diversification and our measure for aspirations.

3.1 Livelihoods

Table 2 provides an overview of the economic activities of households across the four countries. We distinguish between economic activities and sources of monetary income, where we define the former irrespective of whether a household generates a positive amount of monetary income from the activity.[3] This is motivated by the fact that following an economic activity may not necessarily translate into a positive cash earning for a household. The most common example is agriculture - even though around 76 percent of households were engaged in agriculture in 2023, which we define by having cultivated some land and grown some crops, only 23 percent reported to have commercialized some of their harvest and thereby generated a positive monetary income. For all economic activities as well as sources of income, we aggregate at the household level; that is, if at least one household member is involved in a given activity, it counts towards the household.

In Panel A of Table 2, we show the shares of households involved in the various economic activities irrespective of whether a positive amount of monetary income was generated. We note that there is some variation across countries and time - for example, while above 95 percent of households in Tanzania and Zambia were involved in agriculture, less than 60 percent in Namibia cultivated land and grew crops. Panel B shows relevant sources of income. We observe a sharp decline in employment between 2019 and 2023 across all countries, which we attribute to two potential reasons: i) a change in the survey instrument and how employment information was elicited and ii) a lasting impact of COVID.[4]

[3]Subsistence agriculture with in-kind income is not counted under this definition, only if households are able to commercialize and generate some monetary income.

[4]For the former, the questionnaire in 2019 missed a logical skip pattern which led to everyone being asked about their off-farm employment, irrespective of whether they were actually engaged in it. We compute indicators that capture potential data inconsistencies. We find it difficult to come to a final conclusion, but it seems that shares for 2019 may even be only lower bounds. See Section A.2 in the Appendix for more details.

	Kenya		Nambia		Tanzania		Zambia		Full Sample	
	2019	2023	2019	2023	2019	2023	2019	2023	2019	2023
Panel A: Economic activities										
Farming	0.70	0.57	0.46	0.57	0.98	0.95	0.95	0.95	0.78	0.76
	(0.46)	(0.50)	(0.50)	(0.50)	(0.13)	(0.21)	(0.22)	(0.23)	(0.41)	(0.43)
Employed	0.58	0.23	0.64	0.15	0.40	0.10	0.47	0.10	0.51	0.15
	(0.49)	(0.42)	(0.48)	(0.36)	(0.49)	(0.30)	(0.50)	(0.30)	(0.50)	(0.35)
Self-employed	0.14	0.23	0.18	0.14	0.29	0.31	0.16	0.16	0.21	0.23
	(0.35)	(0.42)	(0.39)	(0.35)	(0.46)	(0.46)	(0.37)	(0.37)	(0.40)	(0.42)
Livestock	0.90	0.75	0.69	0.49	0.71	0.60	0.77	0.71	0.77	0.63
	(0.31)	(0.43)	(0.46)	(0.50)	(0.45)	(0.49)	(0.42)	(0.46)	(0.42)	(0.48)
Environmental products	0.64	0.99	0.56	0.98	0.63	0.82	0.69	0.99	0.63	0.93
	(0.48)	(0.11)	(0.50)	(0.13)	(0.48)	(0.39)	(0.46)	(0.10)	(0.48)	(0.26)
Number of economic activities	2.95	2.76	2.53	2.33	3.03	2.78	3.03	2.90	2.89	2.69
	(0.99)	(0.88)	(1.18)	(0.90)	(1.05)	(0.90)	(1.03)	(0.76)	(1.08)	(0.90)
Simpson's Diversity Index (SDI)	0.26	0.20	0.17	0.12	0.18	0.19	0.14	0.12	0.19	0.16
	(0.24)	(0.23)	(0.21)	(0.19)	(0.22)	(0.22)	(0.20)	(0.18)	(0.22)	(0.21)
Panel B: Sources of monetary income										
Farming	0.18	0.11	0.03	0.06	0.34	0.51	0.05	0.11	0.18	0.23
	(0.38)	(0.32)	(0.18)	(0.24)	(0.48)	(0.50)	(0.22)	(0.31)	(0.38)	(0.42)
Employed	0.56	0.23	0.46	0.15	0.35	0.10	0.44	0.10	0.45	0.15
	(0.50)	(0.42)	(0.50)	(0.36)	(0.48)	(0.30)	(0.50)	(0.30)	(0.50)	(0.35)
Self-employed	0.14	0.23	0.16	0.13	0.28	0.29	0.15	0.16	0.19	0.21
	(0.35)	(0.42)	(0.37)	(0.34)	(0.45)	(0.45)	(0.36)	(0.36)	(0.40)	(0.41)
Livestock	0.76	0.65	0.24	0.24	0.41	0.31	0.36	0.41	0.45	0.40
	(0.43)	(0.48)	(0.43)	(0.43)	(0.49)	(0.46)	(0.48)	(0.49)	(0.50)	(0.49)
Environmental products	0.15	0.27	0.14	0.04	0.05	0.05	0.13	0.04	0.11	0.10
	(0.36)	(0.45)	(0.34)	(0.20)	(0.21)	(0.21)	(0.33)	(0.19)	(0.31)	(0.31)
Others	0.28	0.30	0.70	0.68	0.26	0.48	0.18	0.58	0.35	0.49
	(0.45)	(0.46)	(0.46)	(0.47)	(0.44)	(0.50)	(0.38)	(0.49)	(0.48)	(0.50)
Remittances	0.07	0.18	0.11	0.33	0.14	0.39	0.09	0.29	0.10	0.30
	(0.25)	(0.39)	(0.31)	(0.47)	(0.34)	(0.49)	(0.29)	(0.46)	(0.31)	(0.46)
Transfers (NGOs, Govt)	0.19	0.10	0.59	0.40	0.07	0.05	0.06	0.18	0.22	0.16
	(0.39)	(0.30)	(0.49)	(0.49)	(0.25)	(0.21)	(0.24)	(0.38)	(0.42)	(0.37)
Rent	0.02	0.01	0.01	0.01	0.05	0.07	0.01	0.01	0.03	0.03
	(0.15)	(0.11)	(0.07)	(0.12)	(0.22)	(0.25)	(0.07)	(0.09)	(0.16)	(0.17)
PES/nature conservation	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	(0.08)	(0.06)	(0.04)	(0.10)	(0.00)	(0.00)	(0.00)	(0.00)	(0.05)	(0.05)
Others	0.03	0.03	0.07	0.03	0.03	0.08	0.02	0.18	0.04	0.07
	(0.16)	(0.18)	(0.25)	(0.18)	(0.17)	(0.28)	(0.15)	(0.39)	(0.19)	(0.26)
Number of income sources	2.06	1.79	1.72	1.31	1.69	1.73	1.31	1.39	1.73	1.59
	(1.07)	(1.14)	(1.00)	(0.86)	(1.13)	(1.14)	(1.06)	(1.00)	(1.10)	(1.08)
More than one source of monetary income	0.68	0.53	0.54	0.36	0.54	0.56	0.40	0.40	0.55	0.48
	(0.47)	(0.50)	(0.50)	(0.48)	(0.50)	(0.50)	(0.49)	(0.49)	(0.50)	(0.50)
Panel C: Share of different monetary income sources from total monetary income										
Farming	0.07	0.05	0.01	0.02	0.19	0.26	0.02	0.03	0.09	0.11
	(0.20)	(0.19)	(0.08)	(0.13)	(0.34)	(0.37)	(0.12)	(0.15)	(0.25)	(0.27)
Employed	0.31	0.15	0.21	0.11	0.17	0.06	0.29	0.06	0.24	0.10
	(0.37)	(0.30)	(0.34)	(0.28)	(0.31)	(0.22)	(0.40)	(0.22)	(0.35)	(0.26)
Self-employed	0.08	0.15	0.09	0.09	0.17	0.21	0.08	0.10	0.11	0.15
	(0.23)	(0.31)	(0.24)	(0.26)	(0.32)	(0.36)	(0.23)	(0.28)	(0.27)	(0.32)
Livestock	0.33	0.33	0.08	0.09	0.16	0.08	0.19	0.19	0.19	0.17
	(0.38)	(0.41)	(0.21)	(0.24)	(0.31)	(0.21)	(0.36)	(0.35)	(0.33)	(0.32)
Environmental products	0.03	0.09	0.02	0.02	0.02	0.01	0.05	0.02	0.03	0.03
	(0.14)	(0.23)	(0.13)	(0.11)	(0.11)	(0.10)	(0.19)	(0.11)	(0.14)	(0.15)
Others	0.12	0.14	0.51	0.52	0.15	0.22	0.11	0.42	0.22	0.30
	(0.26)	(0.29)	(0.43)	(0.45)	(0.32)	(0.35)	(0.28)	(0.45)	(0.37)	(0.41)
Observations	623	623	548	548	798	798	391	391	2360	2360

Table 2: Livelihood Composition

Notes: This table presents means and standard deviations in parentheses for the main economic activities and the sources of income for the balanced sample of households. We distinguish between economic activities, irrespective of their income-generating nature (Panel A), sources of income (Panel B), and show for each source of income the average contribution to the total income in the past 12 months (Panel C). Our main outcomes of interest are whether the household has more than one income source (extensive margin) and the number of income sources (intensive margin), which are both summarized in Panel B. In addition, we use a measure for the degree of income diversification, namely Simpson's Diversity Index (SDI), which is defined as SDI $= 1 - \sum_{s=1}^{6} (\frac{y_s}{Y})^2$, where y_s is the share of income derived from source s and Y the total income. The resulting SDI takes on values between 0 and 1. The income shares and the SDI are summarized under Panel C.

Evidence on the impact of COVID on employment is mixed. Krafft *et al.* (2022) find for five MENA countries that after initial shocks, employment rates recovered rather quickly across their countries under study and were back to pre-COVID levels towards late 2021. In contrast, Contreras-Gonzalez *et al.* (2022) find large impacts of COVID-19 on labor market outcomes for households in Ethiopia, Malawi, Nigeria, and Uganda. The authors still report large job losses in March 2021, one year after the first lockdowns were imposed in March 2020.[5] The rates range from 27 percent in Uganda to 34 percent in Ethiopia, 37 percent in Malawi and 71 percent in Nigeria (Contreras-Gonzalez *et al.* (2022), Figure 6). Using the same data, Kugler *et al.* (2023) find that for some countries' employment rates partially recovered between April and August 2020 already after the initial shock at the start of the pandemic and accompanied restrictions. In addition, Schotte *et al.* (2023) find that lockdowns had large and significant immediate negative impacts on employment in Ghana. The authors detect persistent nationwide declines in both earnings and employment, particularly for small business owners mainly operating in the informal economy.

In summary, we find it difficult to pin this sharp decline in employment to either the change in the instrument or to COVID. As a result, we include in our main estimation results employment as a source of income, but also estimate our main specification excluding wage employment for robustness. (See Section 6 for more details).

On average households follow around 2.3 to 3 different economic activities both in 2019 and 2023, when looking at actual sources of income, this number drops to around 1.6 to 2.1 different sources on average. In all three countries, the most common combination of livelihoods is comprised of crop- and livestock farming as well as the collection of environmental products, such as wood to produce and sell charcoal. Employed activities in agriculture and services and self-employment in retailing and other small non-agricultural

[5]The authors look at early and long term impacts. The longer term, one year impacts seem more suitable as a benchmark since there was quite some heterogeneity with respect to the implementation and relaxation of restrictions. For instance, the Nigerian government gradually lifted restrictions between June and October 2020, while most restrictions in Ethiopia and Uganda remained until October 2021. By mid-November 2020, restrictions in all countries were at their lowest level since April 2020 and did not increase much again until June 2021. Nevertheless, job losses continued after July 2020 in Malawi, Nigeria, and Uganda.

businesses are also commonly observed.

The total possible number of sources of income is extended by other sources, such as remittances from the extended family or friends and transfers from the government or NGOs. We further find that next to other sources of income, which were received by 49 percent of the sample in 2023, livestock with around 40 percent, and then farming with 23 percent to be the main three sources of monetary income for households in our sample. This is followed by self-employment (21 percent) and wage-employment (15 percent). In terms of the monetary income structure of households, other sources (30 percent), livestock (17 percent), and self-employment (15 percent) are the most important components on average (Panel C of Table 2). Yet the composition of incomes vary between countries. In Kenya, livestock is the largest source of income, in Namibia and Zambia transfers are particularly important, while in Tanzania crop farming is a relevant source of cash income.

Our three main outcome variables for diversification are a binary variable that captures whether a household has more than one income source (extensive margin), the number of income sources (intensive margin), and a measure for diversification in form of Simpson's Diversity Index (SDI). The binary indicator for the extensive margin is equal to one if a household has more than one source of income and zero otherwise (bottom of Panel B in Table 2). For the intensive margin of income diversification, we construct a count index for the number of income sources (bottom of Panel B in Table 2). We also show distributions by country for the number of income sources in Figure A1, which follow the shape of a Poisson distribution and influences the choice of our parametric model detailed in Section 4. Our third main outcome is Simpson's Diversity Index (SDI), which is a weighted average, $\text{SDI} = 1 - \sum_{s=1}^{6} (\frac{y_s}{Y})^2$, where y_s is the share of income derived from source s and Y the total income. Both over a reference period of the past 12 months (bottom of Panel C in Table 2).

Before we discuss our measure of aspirations in Section 3.2, we would like to highlight that we do not attempt to promote diversification as a means to assist the rural poor, but empirical evidence still suggests that it is superior to on-farm and off-farm specialization, particularly in contexts where credit markets that would allow for productivity enhancing on-farm investments are either poorly developed or non-existent, and lucrative wage employment opportunities are limited. With our data we demonstrate a clear positive relationship between our three indicators of income diversification and monthly income levels in Tables A2, A3, and A4. Table A3 even shows suggestive evidence for a linear relationship between the number of income sources and total monthly income.

3.2 Aspirations

To test the predictions of the model by Genicot and Ray (2017) as outlined in Section 2, we construct an income aspiration gap. We opted for a one-dimensional metric that is closely related to decisions on income diversification and disregard the dimensions of social status, assets, and education. As Bernard and Taffesse (2014) noted, economists recently relied on ad-hoc measures to elicit aspirations with a lack of consistency across studies. In our paper, we rely on the more novel measure of aspirations by Bernard and Taffesse (2014) that asks directly about aspired income, which is a good measure of financial aspirations to relate to income diversification. The income aspiration question is "Assessing all household income sources, how much income do you aspire to earn per month in 5 years from now?". With the time horizon spanning to 5 years, those aspirations can be considered as rather long-term, reflecting the decisions over income diversification with longer-term investment implications. To construct the income aspiration gap g_i for household i, we benchmark the aspiration against the total income of the household as follows:

$$g_i = \frac{(a_i - c_i)}{a_i} \tag{1}$$

where c_i represents the current and a_i the aspired monetary income for household i in their respective local currency. To compute the current monthly income, we sum the incomes from all sources under Panel B, Table 2.[6] We elicited in 2023 aspirations for up

[6] Incomes were asked with respect to the past 12 months so in order to benchmark against the aspired monthly aspiration, we divide the total sum across all sources by 12.

to two members per household individually. To construct the income aspiration gap we use the average aspired income within each household. The aspiration gap is equal to 1 if a household i has zero income today. The aspiration gap represents a quantifiable measure of the difference between the current income of household i and its aspired income in the future. To account for outliers in aspired and earned incomes, we winsorize the current- and aspired income at the 5th and 95th percentile of the distribution for our main specification.

In Table 3, we provide descriptive statistics of the current- and aspired monthly incomes per capita, as well as the aspiration gaps across countries and years at household level. It shows that there is quite some variation in terms of average current and aspired levels of income across the four countries. Mean income levels are highest in Namibia with around 57 USD PPP per capita in 2023, followed by Kenya with around 45 USD PPP followed by Tanzania with around 37 USD PPP. Mean per capita incomes are lowest in Zambia with around 15 USD PPP. Average income aspirations are highest in Namibia with around 237 USD PPP per capita in, followed by Kenya with around 220 USD PPP. Average aspirations are considerably lower in Tanzania and Zambia with around 136 USD PPP and 88 USD PPP per capita, respectively.Relative to current incomes, average aspirations in Tanzania are the lowest.

	Kenya		**Nambia**		**Tanzania**		**Zambia**	
	2019	2023	2019	2023	2019	2023	2019	2023
Current nominal monthly income per capita (in local currency)	1726.1	1967.5	406.6	412.3	20854.3	23983.7	72.60	101.9
	(3130.1)	(3907.5)	(766.4)	(572.9)	(63267.4)	(38717.4)	(170.3)	(200.6)
Aspired nominal monthly income per capita (in local currency)	6415.6	9634.2	1669.3	1709.6	86633.9	89458.5	623.9	592.4
	(9107.7)	(11573.8)	(2191.7)	(2087.0)	(126158.6)	(96372.9)	(774.3)	(796.1)
Current monthly income per capita (in 2023 USD PPP)	51.25	44.98	66.95	57.03	36.92	36.52	18.49	15.06
	(92.94)	(89.33)	(126.2)	(79.23)	(112.0)	(58.95)	(43.38)	(29.64)
Aspired monthly income per capita (in 2023 USD PPP)	190.5	220.3	274.9	236.5	153.4	136.2	158.9	87.50
	(270.4)	(264.6)	(360.9)	(288.7)	(223.4)	(146.7)	(197.2)	(117.6)
Aspiration gap	0.694	0.678	0.711	0.709	0.774	0.737	0.872	0.793
	(0.259)	(0.288)	(0.250)	(0.256)	(0.260)	(0.266)	(0.212)	(0.248)
Observations	521	475	438	492	618	658	343	330

Table 3: Average Current-, Aspired Incomes, and Aspirations Gaps

Notes: This table shows average current and aspired incomes per capita and the average aspiration gap per household. We show current and aspired incomes both in local currencies as well as in PPP-adjusted USD at 2023 prices. 2019 values are inflated using the national non-food consumer price indices (CPI). To construct the aspiration gap, we rely on the nominal values in local currencies. For details on the conversion to USD PPP see Section A.3 in the Appendix. Both the current and the aspired incomes are winsorized at the top and bottom 5th percent to account for outliers. We also illustrate the distributions of the current and aspired incomes as well as the aspiration gaps across all four countries in Figures A2 and A3.

Figure A2 shows for each country and year the distribution of the current monthly income in 2023 USD PPP. Figure A3 illustrates the aspiration gaps for both 2019 and 2023 for each country. Both Figures aim to illustrate cross-country differences. For example, there are relatively more households with low incomes in Tanzania and Zambia compared to the other countries, which translates to more mass in the aspiration gaps distribution close to and around 1.

4 Econometric strategy

Aspirations are socially determined and shaped by a host of both observable and unobservable external factors (Parlasca *et al.*). We build our estimation approach on previous studies testing for an inverted U-shape relationship between the income aspiration gap and a future-orientated choice of investment. Specifically, we estimate two types of regression specifications. First, we impose a quadratic functional form on the aspiration gap. Second, we allow the aspiration gap to enter non-parametrically. For the first method, we estimate the following Equation using household fixed effects, which allows us to adjust for unobserved time-invariant heterogeneity.

$$y_{i,t} = \alpha_i + \beta_1 g_{i,t} + \beta_2 g_{i,t}^2 + \beta_3 X_{i,t} + \epsilon_{i,t} \tag{2}$$

where $y_{i,t}$ is an indicator for income diversification for household i in time t (see definitions above), α_i is the time-invariant fixed effect for household i, $g_{i,t}$ and $g_{i,t}^2$ represent the income aspiration gap and the income aspiration gap squared of household i, respectively. Vector $X_{i,t}$ contains time-variant control variables, and $\epsilon_{i,t}$ is an error term clustered at the enumeration area, the first level of sampling. We control for household-level characteristics in the respective time period $t = \{2019, 2023\}$, and country fixed effects when running the analysis for the full sample (we also estimate separate regressions for each country). We therefore identify the effect of the income aspiration gap conditional on observable household-level and enumeration area characteristics. However, the income aspiration gap could be endogenous and linked to wealth and behavioral characteristics of the household. We address this by controlling for any observed and unobserved time-invariant heterogeneity and by controlling for a rich set of household-level characteristics such as age, gender, marital status, and highest level of education of the household head, as well as the size and dependency ratio of the household, the cropland and other assets owned, and a measure for the locus of control. Controls are displayed in Table A5.

We use a linear fixed effects model for our measure of the extensive margin of income diversification and Simpson's Diversity Index (SDI), while we choose a Poisson fixed effects model for the count index of the number of income sources based on its distribution (See Figure A1). We refrain from using probit or logit specifications for our binary outcomes, since maximum likelihood estimators in nonlinear panel models with fixed effects are biased and inconsistent when the length of the panel is small and fixed (Neyman and Scott, 1948; Lancaster, 2000).

Given that the aspiration gap is constrained between zero and one, an inverse U-shape requires that the slope of the curve is positive when the function is equal to zero ($\beta_1 > 0$) and negative when the function is equal to 1 (i.e. $\beta_1 + 2\beta_2 < 0$), with one turning point implied by the quadratic functional form.

According to Lind and Mehlum (2010) it is not enough to verify that $\beta_1 > 0$ and $\beta_2 < 0$ to test for an inverse U-shape. Instead, the authors argue for an additional test to confirm an increasing relationship on the left side of the interval and a decreasing relationship at the right side of the interval. This test leverages a framework proposed by Sasabuchi (1980). In addition, Lind and Mehlum (2010) building on Hirschberg and Lye (2005), who show that confidence intervals commonly computed for ratios using the delta method are severely biased for finite samples in those cases, propose to use Fieller confidence intervals instead.

We therefore also present Sasabuchi (1980) p-values that provide a formal test with the null hypothesis that the sign of the minimum and maximum of the argument are identical. If we can reject the null hypothesis, we have further support for an inverted U-shape relationship. In addition, as mentioned before we follow Lind and Mehlum (2010) building on the work by Fieller (1954) and compute Fieller confidence intervals around the turning point of an inverse U-shaped function. If the turning point lies within the Fieller confidence interval, we have further evidence of an inverted U-shaped relationship. Unlike regular confidence intervals computed using the delta method, Fieller confidence intervals are not forced to be symmetric. In cases of ratio with a ratio with two random variables as here in our case, Fieller confidence intervals have been shown to provide superior coverage Hirschberg and Lye (2010).

We further follow Hardle and Mammen (1993) (HM) and test whether a quadratic is the appropriate parametric fit for the relationship. The null hypothesis of their test is that the fit of the non-parametric and the specified polynomial, (of degree 2 in our case) are not different. If not rejected, this would provide additional support for an (inverted) U-shaped relationship.

In a second step, we estimate Equation 3, which is similar to Equation 2, but here the aspiration gap, $g_{i,t}$ of household i in time t enters the Equation non-parametrically. This allows for a more flexible relationship between the income aspirations gap and our measures of income diversification compared to Equation 2, where we impose a quadratic functional form.

$$y_{i,t} = \alpha_i + f(g_{i,t}) + \beta_1 m_{i,t} + \beta_3 X_{i,t} + \epsilon_{i,t} \tag{3}$$

We estimate this semi-parametric regression using the double residual semi-parametric approach developed by Robinson (1988). The Robinson residuals partial out the parametric fit from the regression and allow to test for a fit of a non-parametric term estimating a full model comprising of a parametric and a non-parametric component. Details on the derivation of the Robinson (1988)'s double residual method can be found in Section A.6 in the Appendix.

In addition, we are able to follow Hardle and Mammen (1993) to formally test whether a quadratic fit is superior to a polynomial of a higher degree.

Although we include household fixed effects and a rich set of controls in our estimations of Equations 2 and 3, the estimates may still be biased from unobserved, time varying heterogeneity. To test how sensitive the results are to this issue, we conduct Oster (2019)'s test for unobservable selection and coefficient stability.

Given our observational data, we do not make causal claims with regards to the evidence for aspirations that are ahead, but not too far ahead that cause effort into income diversification. Instead, we aim to test for an inverted U-shaped relationship between aspirations and our measures of income diversification to demonstrate the aspirations matter for this investment decision.

5 Empirical results

In this Section, we present the empirical results from our two estimation approaches. First, we present results from estimating Equation 2 imposing a quadratic fit between the income aspiration gap and income diversification. Second, we discuss our results from the semi-parametric estimation of Equation 3, where we allow the income aspiration gap to enter without any assumption on functional form. We carry out additional robustness tests in Section 6.

Table 4 presents results from estimating Equation 2, where $y_{i,t}$ takes on the value one if a household i in time t has more than one source of income and zero otherwise – our measure for the extensive margin of income diversification. We observe a positive point estimate for the aspiration gap and a negative point estimate for the aspiration gap squared across all specifications, which provides first supporting evidence for an inverse U-shaped relationship. In addition, the estimated extreme value lies within the range of the data and falls within the 95%-Fieller confidence intervals. We are further able to reject the null hypotheses that the relationship between the aspiration gap and income diversification is monotonous over the interval (see Sasabuchi p-values), which provides additional support for an inverted U-shape relationship.

Dependent variable:	Household has more than one source of monetary income (0/1)									
	Kenya		Namibia		Tanzania		Zambia		Full Sample	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Aspiration gap	1.129***	1.180***	1.330***	1.424***	2.063***	2.085***	2.140***	2.216***	1.687***	1.676***
	(0.328)	(0.302)	(0.413)	(0.427)	(0.329)	(0.327)	(0.563)	(0.591)	(0.199)	(0.195)
Aspiration gap squared	-1.347***	-1.323***	-1.467***	-1.541***	-2.293***	-2.299***	-2.212***	-2.286***	-1.878***	-1.849***
	(0.262)	(0.242)	(0.333)	(0.350)	(0.255)	(0.253)	(0.420)	(0.436)	(0.156)	(0.154)
Constant	0.522	0.475	0.310	0.285	0.388	0.375	0.234	0.230	0.375	0.367
R^2	0.672	0.690	0.642	0.655	0.658	0.666	0.673	0.685	0.656	0.664
Observations	782	782	788	788	1044	1044	574	574	3188	3188
Number of clusters	47	47	44	44	60	60	30	30	181	181
Household FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Country FE	No	No	No	No	No	No	No	No	Yes	Yes
U-test results:										
Turning point	0.419	0.446	0.454	0.462	0.450	0.453	0.484	0.485	0.449	0.453
Fieller 95% CI	[0.275;0.495]	[0.328;0.515]	[0.299;0.528]	[0.321;0.533]	[0.387;0.494]	[0.392;0.497]	[0.354;0.552]	[0.350;0.555]	[0.406;0.482]	[0.412;0.486]
Sasabuchi p-value	0.001	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000
Slope at min	1.122	1.173	1.282	1.373	2.058	2.080	2.016	2.087	1.682	1.671
Slope at max	-1.565	-1.466	-1.603	-1.657	-2.523	-2.513	-2.284	-2.356	-2.069	-2.022

Table 4: Aspiration Gap and Income Diversification - Extensive Margin

Notes: This table presents results from linear fixed effects models, where we regress a binary income diversification indicator (value of one if the household has more than one source of income and zero otherwise) on the aspiration gap and the aspiration gap squared. Income sources considered here are farming, employment, self-employment, livestock production, the collection of environmental products as well as other sources of income. All sources of incomes are summarized in Panel B of Table 2. Controls included are age, gender, marital status, highest level of education of the household head, the size and socio-dependency of the household, the cropland, other assets owned and the locus of control. Controls are displayed in Table A5. We also add Fieller confidence intervals around the turning point (Fieller, 1954), which are preferred and less biased than delta method confidence intervals for finite samples as summarized by Lind and Mehlum (2010) in their work on testing for a U-shaped relationship. Standard errors are clustered at the enumeration area level. ***, **, and * indicate significance at the 1, 5, and 10 percent level. Linear predictions from the estimation of the model for the full sample, including controls are illustrated in Figure 1a.

As for the extensive margin, we also find strong evidence at the intensive margin for an inverse U-shaped relationship between the number of income sources and the income aspiration gap (Table 5). We observe a positive point estimate for the aspiration gap and a negative point estimate for the aspiration gap squared across all model specifications. The turning point (estimated extreme value) lies within the range of the data and falls within the 95%-Fieller confidence intervals across all specifications. We are further able to reject the null hypotheses that the relationship between the aspiration gap and the number of income sources is monotonous over the interval (see Sasabuchi p-values), providing additional support for an inverted U-shape relationship. The point estimates as such are not interpretable, but are indicative of the shape and curvature of the inverted U-shape. The point estimates on β_1 are indicative of the slope and curvature of the left side until the turning point, while the estimates on β_2 for the slope and curvature on the right side of the inverted U-shape.

Dependent variable:	**Number of monetary income sources (0-6)**									
	Kenya		**Namibia**		**Tanzania**		**Zambia**		**Full Sample**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Aspiration gap	1.535***	1.565***	3.397***	3.689***	4.186***	4.274***	4.359***	4.232***	3.133***	3.141***
	(0.405)	(0.381)	(0.584)	(0.637)	(0.508)	(0.499)	(0.856)	(0.821)	(0.295)	(0.286)
Aspiration gap squared	-2.024***	-1.980***	-3.699***	-3.955***	-4.598***	-4.677***	-4.577***	-4.416***	-3.571***	-3.552***
	(0.355)	(0.334)	(0.499)	(0.547)	(0.450)	(0.439)	(0.726)	(0.670)	(0.256)	(0.249)
Constant	0.690	0.625	0.094	0.016	0.172	0.153	0.042	0.004	0.325	0.294
Pseudo-R^2	0.148	0.151	0.151	0.155	0.168	0.170	0.173	0.177	0.161	0.163
Observations	776	776	774	774	984	984	530	530	3064	3064
Household FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Country FE	No	No	No	No	No	No	No	No	Yes	Yes
U-test results:										
Turning point	0.379	0.395	0.459	0.466	0.455	0.457	0.476	0.479	0.439	0.442
Fieller 95% CI	[0.270;0.442]	[0.299;0.453]	[0.403;0.499]	[0.413;0.503]	[0.421;0.482]	[0.423;0.484]	[0.415;0.516]	[0.410;0.526]	[0.411;0.461]	[0.416;0.464]
Sasabuchi p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Slope at min	1.525	1.555	3.273	3.558	4.176	4.263	4.101	3.984	3.125	3.133
Slope at max	-2.512	-2.394	-4.001	-4.220	-5.010	-5.080	-4.795	-4.600	-4.008	-3.964

Table 5: Aspiration Gap and Income Diversification - Intensive Margin

Notes: This table presents regression results from Poisson fixed effects models, where we regress the number of income sources on the aspiration gap and the aspiration gap squared. Income sources considered here are farming, employment, self-employment, livestock production, the collection of environmental products as well as other sources of income. All sources of income are summarized in Panel B of Table 2. Controls included are age, gender, marital status, highest level of education of the household head, the size and socio-dependency of the household, the cropland, other assets owned and the locus of control. Controls are displayed in Table A5. We also add Fieller confidence intervals around the turning point (Fieller, 1954), which are preferred and less biased than delta method confidence intervals for finite samples as summarized by Lind and Mehlum (2010) in their work on testing for a U-shaped relationship. Standard errors are clustered at the enumeration area level. ***, **, and * indicate significance at the 1, 5, and 10 percent level. Linear predictions from the estimation of the model for the full sample, including controls are illustrated in Figure 1b.

To illustrate the estimated relationship between the income aspiration gap and the two indicators of income diversification (extensive and intensive margin), we plot the marginal linear predictions including the confidence intervals for both of them in Figures 1a and 1b. In addition, we indicate with the thick dashed-dotted line the estimated turning point and thinner dashed lines the 95% Fieller confidence intervals. The marginal linear predictions intuitively illustrate the theory by Genicot and Ray (2017). Looking at Figure 1a and 1b with the aspiration gap on the horizontal axis and the marginal linear prediction on the vertical axis. It suggests that when levels of diversification are comparably low, they may not be sufficiently far ahead to strongly induce investments into additional income-generating activities. But investments increase with the aspiration gap up to the turning point, after which the gap may be too large, leading to decreasing levels of diversification.

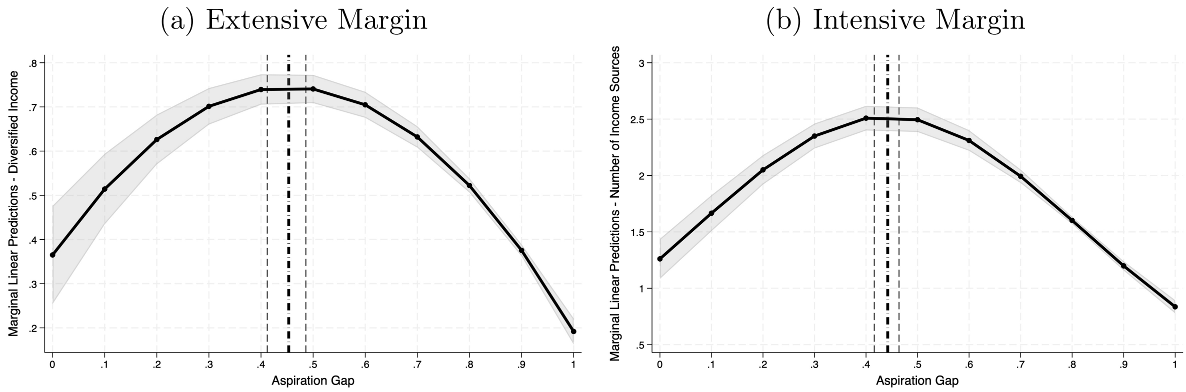


Figure 1: Linear predictions: Income Diversification

Notes: Figures 1a and 1b plot the linear prediction between our measures of the extensive and intensive margin of income diversification and the aspiration gap. The gray shaded area depicts the 95% confidence interval band. In each Figure panel, the thick dashed-dotted line indicates the estimated turning point, while the dashed lines highlight the Fieller confidence intervals (Fieller, 1954), which are preferred and less biased than the delta method confidence intervals for finite samples as summarized by Lind and Mehlum (2010) in their work on testing for a U-shaped relationship. The linear predictions for the extensive margin come from a linear fixed effects model for the full sample of households across the four countries (Table 4). The linear predictions for the intensive margin come from a fixed effects Poisson model for the full sample of households (Table 5). Controls included are age, gender, marital status, highest level of education of the household head, the size and socio-dependency of the household, the cropland, other assets owned and the locus of control.

Table A7 shows additional supporting evidence of an inverted U-shaped relationship between the income aspiration gap and our third outcome indicator - the SDI. We also plot the marginal linear predictions for the SDI in Figure A4 in the Appendix. The findings are in line with our results for the extensive and intensive margin indicators with one exception: for Kenya, we do not find strong evidence for an inverted U-shape relationship using the SDI. Figure A5 illustrates the marginal linear prediction for each country. For Kenya, Figure A5a shows a concave, downward sloped relationship between the aspiration gap and the SDI. Average levels of diversification in Kenya are higher than in the other countries, captured in the constants in Table A7 and the descriptive statistics in Table 2.

Based on the parametric results, we can summarize that we find robust evidence for the existence of an inverted U-shaped relationship between the income aspiration gap and our three measures of income diversification.

To illustrate our results from the non-parametric estimation of Equation 3, we compute expected values for each of the three outcome variables for the mean of the income aspiration gap across 50 quantiles to illustrate the degree of income diversification over the aspiration gap. The methods and results are discussed in detail in Appendix A.6.

For all models, we graphically observe inverted U-shaped relationships between the aspiration gap and income diversification in Figures A6), A7) and A8). However, based on the Hardle and Mammen (1993) test results we have to reject the null hypothesis that the non-parametric and quadratic parametric fits do not differ. When looking at the Figures, it seems that we have more households with an aspiration gap on the right side of the distribution, reflected by the bunching of bin means on the right side in each Figure panels. These results do not challenge the inverted U-shape but suggest that the relationship is not necessarily fully symmetric, as implied by the quadratic functional form.

6 Robustness

In this Section, we test for the robustness of our results by conducting several additional tests. First, to address the potential issue of measurement error of employment during the 2019 data collection, we construct our income diversification indicators without considering any monetary income from employment. The corresponding results for the extensive (Table A8) and intensive margin (Table A9) provide strong evidence that our full results - with monetary income from employment included - are not driven by this potential bias caused by our difference in measurement of employment nor by a potential decrease in employment due to COVID.

Second, we test whether our approach of winsorizing the income and aspiration measures may possibly influence the results in unexpected ways. Incomes and income aspirations are often considered noisy measures. In the main analyses, we account for this by winsorizing both measures at the top and bottom 5 percent levels. To demonstrate robustness to the choice of different cutoffs we show in Tables A10 and A11 alternative results estimated by using different winsorization cutoffs. All results support the main estimates and strengthen our finding of aspiration failure in the case of income diversification.

Third, we test for possible omitted variable bias. We rely on household fixed effects models, thus controlling for for time-invariant unobserved factors, but time-variant unobserved factors might still potentially bias the estimates. To address this concern, we use the method by Oster (2019). In Appendix A.7 we outline the test and report bias-adjusted estimates for β_1 and β_2 across all parametric regressions (Eq. 2). Based on these estimates, we conclude that our results are not very susceptible to omitted variable bias; the bias-adjusted estimates stay within the confidence intervals and the computed δs range outside the critical range between zero and one except for one case.

Fourth, we test for the role of sample attrition over the two survey waves. As Table 1 summarizes, we face a rate of attrition of around 11 percent for the full sample with some variation across countries. Attrition is lowest in Tanzania with around 8 percent, followed by Zambia and Kenya with 11 and 12 percent, respectively, and highest in Namibia with around 16 percent.

To test for differential attrition, we first test for systematic differences between attritors and non-attritors in terms of socio-demographic characteristics and main variables of analysis in round 1 (2019). Results are shown in Table A12. In addition, we test whether differences in aspirations are correlated with attrition. Although some of the differences are statistically significant based on t-tests, the actual differences are small. The normalized differences, which are generally preferred over t-tests because they provide a scale-free comparison (Imbens and Rubin, 2015; Imbens and Wooldridge, 2009; Abadie and Imbens, 2011), are all below 0.25. The result of the joint F-test raises some concerns suggesting that differences between observable variables predict attrition. To address this issue, we carry out two additional tests of robustness against attrition.

Firstly, we estimate all our main models with inverse probability weights (IPW). We obtain these weights from a probit model using 2019 data, regressing attrition status on the set of household characteristics, the main outcome variables, and the aspiration gap. Results from these probit regressions are shown in Table A14. Then we use the weights to re-estimate the relationship between the aspiration gap and our three main outcomes for income diversification (Tables A15), A16), and A17). The IPW results are very similar to the main model estimates, suggesting that attrition is not a threat to the conclusion drawn from our analysis.

Secondly, we test whether the estimates may be driven by the 2023 data by regressing an attrition indicator on the aspiration gap and the aspiration gap squared. Results are summarized in Table A13 and do not suggest that the aspiration gap predicts whether respondents were interviewed in round 2 (2023) of the survey. Taken together, even though it seems that attrition households differ from non-attrition households in some dimensions, we conclude that the threat of potential attrition bias is likely minor.

7 Conclusion

The topic of livelihood diversification has received much attention in the recent literature. Determinants of diversification have been categorized as push and pull factors, where the former are typically characterized by ex-ante risk-management or ex-post risk-coping strategies and the latter catalyzed via commercialization or proximity to urban areas. We provide additional evidence from behavioral economic theory that, next to these external factors, internal, behavioral factors seem to matter in the process of income diversification. Using panel data from four countries in sub-Saharan Africa, we show that income diversification is associated with higher levels of overall income in local rural settings, where secure and lucrative wage employment opportunities are rare. Further, we test and find strong supportive evidence that aspirations matter for the process of income diversification.

More specifically, we find strong evidence for an inverted U-shaped relationship between the aspiration gap and income diversification across all four countries, supporting the theory by Genicot and Ray (2017) that aspirations that are ahead, but not too far ahead are associated with future-oriented behavior. Our results are robust across various indicators of income diversification, namely a simple binary indicator that captures the extensive margin for whether a household pursues more than one income-generating activity, an indicator for the intensive margin counting the number of income-generating activities, and Simpson's Diversity Index (SDI).

With our findings, we do not attempt to promote diversification as a general strategy for poverty reduction and development in rural Africa. Supporting smallholder farming and creating lucrative non-agricultural employment through investments in the rural economy are important policy priorities. However, in the light of widespread market failures, uncertainties, and limited wage employment opportunities that many rural households in Africa face, our results suggest that income diversification is beneficial and that aspirations matter. We believe that our findings are externally valid to other countries with similar conditions, particularly with a scarcity of secure wage employment opportunities.

Finally, our findings underpin the importance to consider behavioral factors when designing development policies. Understanding the role of aspirations and how to influence them in specific contexts may be useful to induce future-oriented investment behavior. If existing levels of aspirations are too low, interventions to raise them should be considered as part of broader development programs. On the other hand, our results also suggest that in situations where aspirations are too high, interventions to lower them to more realistic levels can be useful to promote more future-oriented behavior.

References

Abadie, A. and Imbens, G. W. (2011). Bias-corrected matching estimators for average treatment effects. *Journal of Business & Economic Statistics*, **29** (1), 1–11.

Alobo Loison, S. (2015). Rural livelihood diversification in sub-Saharan Africa: a literature review. *The Journal of Development Studies*, **51** (9), 1125–1138.

Altonji, J. G., Elder, T. E. and Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy*, **113** (1), 151–184.

Barrett, C. B., Reardon, T. and Webb, P. (2001). Nonfarm income diversification and household livelihood strategies in rural Africa: Concepts, dynamics, and policy implications. *Food Policy*, **26** (4), 315–331.

Bernard, T. and Taffesse, A. S. (2014). Aspirations: An approach to measurement with validation using Ethiopian data. *Journal of African Economies*, **23** (2), 189–224.

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

Bloem, J. R. (2021). Aspirations and investments in rural Myanmar. *Journal of Economic Inequality*, **19** (4), 727–752.

Cobb-Clark, D. A. (2015). Locus of control and the labor market. *IZA Journal of Labor Economics*, **4**, 1–19.

Contreras-Gonzalez, I., Oseni, G., Palacios-Lopez, A., Pieters, J. and Weber, M. (2022). Inequalities in job loss and income loss in Sub-Saharan Africa during the covid-19 crisis.

Dalton, P. S., Ghosal, S. and Mani, A. (2016). Poverty and aspirations failure. *The Economic Journal*, **126** (590), 165–188.

De Janvry, A. and Sadoulet, E. (2001). Income strategies among rural households in Mexico: The role of off-farm activities. *World Development*, **29** (3), 467–480.

Dercon, S. (2002). Income risk, coping strategies, and safety nets. *World Bank Research Observer*, **17** (2), 141–166.

Ellis, F. (2000). The determinants of rural livelihood diversification in developing countries. *Journal of Agricultural Economics*, **51** (2), 289–302.

FAO (2021). *Sub-Saharan Africa: Strengthening resilience to safeguard agricultural livelihoods*. Tech. rep., Food and Agriculture Organization of the United Nations.

Fieller, E. C. (1954). Some problems in interval estimation. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **16** (2), 175–185.

Genicot, G. and Ray, D. (2017). Aspirations and inequality. *Econometrica*, **85** (2), 489–519.

Hardle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Annals of Statistics*, pp. 1926–1947.

Hirschberg, J. and Lye, J. (2010). A geometric comparison of the delta and fieller confidence intervals. *The American Statistician*, **64** (3), 234–241.

— and Lye, J. N. (2005). Inferences for the extremum of quadratic regression models. *Available at SSRN 918626*.

Imbens, G. W. and Rubin, D. B. (2015). *Causal inference for statistics, social, and biomedical sciences: An introduction*. Cambridge University Press.

— and Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, **47** (1), 5–86.

Janzen, S. A., Magnan, N., Sharma, S. and Thompson, W. M. (2017). Aspirations failure and formation in rural Nepal. *Journal of Economic Behavior & Organization*, **139**, 1–25.

Kahneman, D. and Tversky, A. (2013). Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I*, World Scientific, pp. 99–127.

Khan, R. and Morrissey, O. (2023). Income diversification and household welfare in Uganda 1992–2012. *Food Policy*, **116**, 102421.

KNBS (2019). *2019 Kenya Population and Housing Census*. Tech. rep., Kenya National Bureau of Statistics.

— (2023). *The Kenya Poverty Report 2021*. Tech. rep., Kenya National Bureau of Statistics.

Krafft, C., Assaad, R., Marouani, M. A., Cheung, R. and LaPlante, A. (2022). Are labor markets in the Middle East and North Africa recovering from the COVID-19 pandemic? *IZA Journal of Development and Migration*, **14** (1).

Kremer, M., Rao, G. and Schilbach, F. (2019). Behavioral development economics. In *Handbook of behavioral economics: applications and foundations 1*, vol. 2, Elsevier, pp. 345–458.

Kugler, M., Viollaz, M., Duque, D., Gaddis, I., Newhouse, D., Palacios-Lopez, A. and Weber, M. (2023). How did the COVID-19 crisis affect different types of workers in the developing world? *World Development*, **170**, 106331.

Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of econometrics*, **95** (2), 391–413.

Lind, J. T. and Mehlum, H. (2010). With or without u? the appropriate test for a u-shaped relationship. *Oxford Bulletin of Economics and Statistics*, **72** (1), 109–118.

Mehraban, N., Kubitza, C., Alamsyah, Z. and Qaim, M. (2021). Oil palm cultivation, household welfare, and exposure to economic risk in the Indonesian small farm sector. *Journal of Agricultural Economics*, **72** (3), 901–915.

MoFP-PED and NBS (2024). *The 2022 Population and Housing Census: Basic Demographic and Socio-Economic Profile; Tanzania Mainland*. Tech. rep., The United Republic of Tanzania (URT), Ministry of Finance, Tanzania National Bureau of Statistics and President's Office - Finance and Planning, Office of the Chief Government Statistician.

Musumba, M., Palm, C. A., Komarek, A. M., Mutuo, P. K. and Kaya, B. (2022). Household livelihood diversification in rural Africa. *Agricultural Economics*, **53** (2), 246–256.

Musungu, A. L., Kubik, Z. and Qaim, M. (2024). Drought shocks and labour reallocation in rural africa: evidence from ethiopia. *European Review of Agricultural Economics*, p. jbae020.

Nagler, P. and Naudé, W. (2017). Non-farm entrepreneurship in rural sub-Saharan Africa: New empirical evidence. *Food Policy*, **67**, 175–191.

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica: journal of the Econometric Society*, pp. 1–32.

NPC (2015). *Namibia Poverty Mapping*. Tech. rep., National Planning Commission.

NSA (2024). *2023 Population & Housing Census: Preliminary Report*. Tech. rep., Namibia Statistics Agency.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, **37** (2), 187–204.

Parlasca, M. C., Martini, C. A., Köster, M. and Ibañez, M. (). Aspirations and weather shocks: Evidence from rural zambia. *Agricultural Economics*.

Ray, D. (2006). Aspirations, poverty, and economic change. *Understanding Poverty*, **1**, 409–421.

Reardon, T. (1997). Using evidence of household income diversification to inform study of the rural nonfarm labor market in Africa. *World Development*, **25** (5), 735–747.

—, Berdegué, J., Barrett, C. B. and Stamoulis, K. (2007). Household income diversification into rural nonfarm activities. *Transforming the rural nonfarm economy: opportunities and threats in the developing world*, pp. 115–140.

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society*, pp. 931–954.

Ross, P. H. (2019). Occupation aspirations, education investment, and cognitive outcomes: Evidence from Indian adolescents. *World Development*, **123**, 104613.

Rotter, J. (1966). Generalized expectancies of internal versus external control of reinforcements. *Psychological monographs*, **80**, 609.

Sasabuchi, S. (1980). A test of a multivariate normal mean with composite hypotheses determined by linear inequalities. *Biometrika*, **67** (2), 429–439.

Schotte, S., Danquah, M., Osei, R. D. and Sen, K. (2023). The labour market impact of COVID-19 lockdowns: Evidence from Ghana. *Journal of African Economies*, **32** (Supplement_2), ii10–ii33.

Tabe-Ojong, M. P., Heckelei, T. and Rasch, S. (2023). Aspirations and investments in livestock: Evidence of aspiration failure in Kenya. *Agricultural Economics*.

Tabe-Ojong, M. P. J., Kedinga, M. E. and Gebrekidan, B. H. (2024). Behavioural factors matter for the adoption of climate-smart agriculture. *Scientific Reports*, **14** (1), 798.

URT (2019). *Tanzania mainland household budget survey 2017-18, key indicators report; Tanzania Mainland*. Tech. rep., Ministry of Finance and Planning - Poverty Eradication Division (MoFP- PED) and National Bureau of Statistics (NBS).

Villacis, A. H., Bloem, J. R. and Mishra, A. K. (2023). Aspirations, risk preferences, and investments in agricultural technologies. *Food Policy*, **120**, 102477.

ZamStats (2024). *Zambia - 2022 Census of Population and Housing*. Tech. rep., "Central Statistical Office Zambia (ZamStats)".

A Appendix

A.1 Sampling and Data collection

In this Section, we detail the sampling procedures, the field activities, and the content of the survey instruments used. For each country, we used a two-stage sampling process to firstly randomly select a fixed number of enumeration areas and secondly within each enumeration area a fixed number of households. For the data collection, a group of well-trained enumerators interviewed the household head or their spouse in their local language using a harmonized household survey across all countries, covering a range of socio-demographic characteristics, agricultural and livestock production, shocks and coping strategies, aspirations, road connectivity, conservancies and environmental products, and ecological shocks. Compared to 2019, the module on labor participation was refined, and new modules on food security, gender roles, and time allocation were introduced. Surveys in both rounds involved computer-assisted personal interviews (CAPI) using tablets and surveyCTO in 2019 and survey solutions in 2023.

In each round of the data collection, data were first collected in Namibia between May - June 2019 and 2023, respectively. Followed by Zambia and Tanzania, where data were collected between June - July 2019 and 2023. Lastly in Kenya data were collected between July - August in 2019 and 2023, respectively.

A.1.1 Kenya

The selection followed a two-stage stratified random sampling process. Within this approach, subgroups (strata) of a population are identified. For Baringo county, these are conservation, intensification and others. In Kenya's National Sample Survey and Evaluation Programme (NASSEP V) the country was subdivided into enumeration areas (EAs). These EAs are used for the household survey and classified according to the above-mentioned strata. For each stratum, 15 EAs are randomly drawn, which is the first stage of the sampling process. This resulted in 47 selected EAs (including two additional contingency EAs) from a total of 1,754 rural EAs in the Baringo county. As the aim of the study is to analyze rural populations, urban EAs were excluded. For the second stage of sampling, 15 Households inside the selected EAs (from stage one), are selected. The total number of possible households for selection into the study was 94,327 out of which 705 were selected randomly. This gives a representative sample of the whole county. The data collection in Kenya took place in between July 16 - August 9, 2019 and July 14 - August 14, 2023 for the second round.

A.1.2 Namibia

In Namibia, we followed a two-stage stratified random sampling process using Namibia's 2011 National Population and Housing Census which subdivides each region into enumeration areas (EAs) (Namibia Statistics Agency 2012). These EAs are then classified according to an intensification, conservation and other strata. For each stratum, 15 EAs are randomly drawn, which is the first stage of the sampling process. For the second stage of the sampling, 15 households inside the selected EAs (from stage one), are selected. This results in 45 selected EAs from a total of 288 rural EAs in the Zambezi region. The total number of possible households for selection into the study was 21,600 out of which 675 were selected randomly. This gives a representative sample of the whole region. Data was collected from May 20 - June 14, 2019, while the second round was collected between May 15, 2023 - June 17, 2023. An additional follow-up survey was conducted between September 20, 2019 and October 1, 2019.

A.1.3 Tanzania

In Tanzania, we followed a two-stage stratified random sampling process. Within this approach, subgroups (strata) of a population are identified. For the Iringa and Morogoro region, these are agriculture schemes, infrastructure projects (SAGCOT, TAZARA etc.) and neither agricultural nor infrastructure. To randomly select households and villages, enumeration areas (EAs) were drawn fromthe 2012 Tanzania Population and Housing Census. These EAs are then classified according to the above-mentioned strata. Our project spans across two regions and four districts. The four districts are composed of

90 wards and 379 villages with a total of 1,145,240 people and 299,762 households. In each region, 30 EAs were selected using probability proportional to size (PPS) sampling, where the measure of size is the number of households in the 2012 census frame. Then, for each EA 15 households are drawn, resulting in a total sample size of 870 households. The resulting sample is thereby representative of the study area. The data collection of the first round took place between June 14 - July 19, 2019. Similar to Namibia, there was an additional follow-up in Tanzania, which was rolled out between July 20 - August 19, 2019. The second round was collected between June 24 - August 3, 2023.

A.1.4 Zambia

In Zambia, the selection followed a two-stage stratified random sampling process. Within this approach, subgroups (strata) of a population are identified. For the Western province, these are conservation, intensification and other. In 2010 a Population and Housing census was performed throughout Zambia. Here, the districts were subdivided into enumeration areas (EAs). For the survey, all rural EAs were selected and classified according to the above-mentioned strata. For each stratum, 10 EAs are randomly drawn, which is the first stage of the sampling process. For the second stage of the sampling, 15 households inside the selected EAs (from stage one), are selected. This results in 30 selected EAs from a total of 25,632 EAs in Zambia. The total number of possible households for selection into the study was 9,500 out of which 437 were selected randomly. This gives a representative sample of the whole province. The data collection in Zambia was rolled out between June 27 and July 14, 2019 and June 5 - June 25, 2023.

A.2 Measurement of Employment

The measurement of employment was changed between the 2019 and 2023 survey waves. In 2019, individual household members above the age of 11 were first asked whether they were employed or involved in any income-generating activity, including farming, followed by a specification of the sector. In a later section, everyone unconditional of the response to the first questions was asked about their off-farm activities, in detail i) “In the last year, how many months did ... work at this off-farm activity?”, ii) “In a typical week, how many days did ... work?”, and iii) “What was ... estimated monthly income?”. As a result, based on this lack of logical skip pattern in the questionnaire, we do not know for certain whether an individual really followed some off-farm activity or not. Table A1 provides an overview of the three employment questions as they were asked in 2019. We also define two indicators that capture potential issues: 1.) the respondents provided no information on the number of days and months they worked off their own farm (time), but stated to have received a positive amount of income. 2.) Respondents did provide information on number of months and days worked, but not on income. The shares for potential issues on 1.) are around 4 percent in Namibia and close to zero in the other countries. For 2.) shares are larger across all countries, largest in Namibia with around 18 percent. For this analysis, we consider that someone was employed off their own farm if they stated to have received a positive amount of monthly income. We consider this definition rather strict and therefore as a lower bound of the true share of off-farm employment.

The employment module was refined in 2023, respondents were first asked again whether they were employed or involved in any income-generating activities, including farming?, followed by more precise questions of a) whether they were employed in any job away from their own farm, including their own business as well as ii) whether they had undertaken any work on the farm in the past 12 months. These additional questions allow us to distinguish with more certainty the economic activities of a given individual. At the end, we do apply the same definition in terms of a positive income generated from employment to be counted towards our indicators of income diversification.

	Kenya 2019	Nambia 2019	Tanzania 2019	Zambia 2019	Full Sample 2019
Questions as in 2019 survey					
In the last year, how many months did (s)he work at this off-farm activity?	2.05 (3.69)	2.13 (3.80)	0.91 (2.70)	2.20 (4.07)	1.74 (3.55)
In a typical week, how many days did (s)he work?	1.38 (2.20)	1.85 (2.69)	1.60 (2.66)	1.40 (2.37)	1.57 (2.52)
What was her/his estimated monthly income?	2696.93 (7929.96)	455.06 (2065.69)	26353.92 (88462.97)	215.04 (784.55)	9159.05 (51157.25)
Potential data issues					
No information on months and days, but on income	0.00 (0.02)	0.04 (0.19)	0.00 (0.06)	0.01 (0.10)	0.01 (0.11)
Information on months and days, but not on income	0.01 (0.12)	0.18 (0.38)	0.07 (0.25)	0.05 (0.21)	0.08 (0.27)
Observations	1770	1742	2115	1092	6719

Table A1: Off-farm employment information

Notes: This table provides an overview of the employment questions asked in 2019 and the share of potential issues across countries.

A.3 Conversions to USD PPP

The surveys were collected between May - August in 2019 and 2023. We decided to convert all monetary values displayed in 2023 USD PPP. To do that, we average the national monthly non-food national consumer price index (CPI) series over the time period in which the data collection in a given country took place. We obtain for Kenya a CPI of 103.3 for 2019 and 134.09 for 2023, for Namibia a CPI of 137.25 for 2019 and 162.85 for 2023, for Tanzania a CPI of 97.18 (116.52 rebased) in 2019 and 112.58 in 2023, and finally for Zambia a CPI of 229.05 in 2019 and 394.54 in 2023. Then we rebase the 2019 CPI factors that 2023 is equal to 1 to obtain national CPI inflation adjusted conversion rates, which we use to multiply the 2019 values with to obtain 2019 monetary values in 2023 terms.

Lastly, to convert to 2023 USD PPP we divide the newly obtained 2019 monetary values in 2023 terms as well as the 2023 values with World Bank's USD PPP conversion factors for private consumption (Link, last accessed August 2023.)

A.4 Tables

Dependent variable:	**Income in 2023 USD PPP**									
	Kenya		**Namibia**		**Tanzania**		**Zambia**		**Full Sample**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
More than one monetary income source	239.470***	211.881***	223.920***	215.589***	264.145***	237.208***	141.193***	140.801***	231.374***	215.471***
	(30.133)	(28.403)	(20.280)	(20.543)	(35.689)	(33.889)	(21.313)	(20.824)	(16.014)	(15.478)
Constant	168.699	185.394	183.856	187.619	128.282	143.118	59.908	60.064	141.201	149.418
R^2	0.711	0.727	0.703	0.720	0.698	0.712	0.704	0.715	0.709	0.718
Observations	1246	1246	1096	1096	1596	1596	782	782	4720	4720
Number of clusters	47	47	44	44	60	60	30	30	181	181
Household FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Country FE	No	No	No	No	No	No	No	No	Yes	Yes

Table A2: Incomes and Income Diversification, Extensive Margin

Notes: This table shows results from linear fixed effects models, where the income in 2023 USD PPP is regressed on a binary income diversification indicator that is equal to one if the household has more than one sources of income and zero otherwise. The income sources considered here are farming, employment, self-employment, livestock production, the collection of environmental products as well as other sources of income. All income sources are summarized in Panel B of Table 2. Controls included are age, gender, marital status, highest level of education of the household head, the size and socio-dependency of the household, the cropland, other assets owned and the locus of control. All controls are displayed in Table A5. Standard errors are clustered at the enumeration area level. ***, **, and * indicate significance at the 1, 5, and 10 percent level.

Dependent variable:	**Income in 2023 USD PPP**									
	Kenya		**Namibia**		**Tanzania**		**Zambia**		**Full Sample**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1 income source	195.575***	196.052***	118.349***	114.243***	60.060***	57.872**	46.072***	46.374***	86.312***	82.514***
	(37.635)	(38.164)	(22.278)	(25.187)	(21.172)	(21.778)	(12.291)	(14.673)	(12.161)	(12.685)
2 income sources	321.897***	305.361***	273.698***	263.397***	211.725***	187.170***	136.450***	140.568***	223.491***	209.493***
	(47.216)	(45.973)	(27.552)	(27.901)	(33.169)	(32.866)	(23.818)	(23.656)	(17.518)	(17.287)
3 income sources	513.914***	488.847***	411.482***	401.599***	429.749***	399.090***	267.382***	259.181***	409.649***	390.135***
	(54.169)	(48.311)	(39.099)	(41.609)	(47.377)	(45.972)	(29.123)	(35.019)	(25.312)	(23.962)
4 or more income sources	672.365***	638.527***	554.111***	552.428***	564.043***	531.620***	256.544***	261.541***	536.210***	511.317***
	(66.187)	(65.668)	(95.462)	(93.555)	(66.549)	(65.740)	(65.015)	(67.442)	(37.508)	(37.067)
Constant	-13.496	-0.439	85.072	91.212	78.587	94.333	26.356	25.916	68.054	78.103
R^2	0.750	0.762	0.722	0.739	0.719	0.732	0.720	0.728	0.732	0.739
Observations	1246	1246	1096	1096	1596	1596	782	782	4720	4720
p(1 source = 2 sources)	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.000	0.000
p(2 sources = 3 sources)	0.000	0.000	0.002	0.002	0.000	0.000	0.000	0.003	0.000	0.000
p(3 sources = 4 sources)	0.004	0.004	0.177	0.130	0.023	0.024	0.881	0.976	0.000	0.000
Number of clusters	47	47	44	44	60	60	30	30	181	181
Household FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Country FE	No	No	No	No	No	No	No	No	Yes	Yes

Table A3: Incomes and Income Diversification, Intensive Margin

Notes: This table shows results from a linear regression model, where the income in 2023 USD PPP is regressed on an income diversification indicator which captures the number of income sources of the household. The income sources considered here are farming, employment, self-employment, livestock production, the collection of environmental products as well as other sources of income. All income sources are summarized under Panel B in Table 2. Controls included are age, gender, marital status, highest level of education of the household head, the size and socio-dependency of the household, the cropland, other assets owned and the locus of control. All controls are displayed in Table A5. Standard errors are clustered at the level of sampling, the enumeration area level. ***, **, and * indicate significance at the 1, 5, and 10 percent level.

Dependent variable:	**Income in 2023 USD PPP**									
	Kenya		**Namibia**		**Tanzania**		**Zambia**		**Full Sample**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Simpson diversity index	403.481***	343.690***	462.626***	434.621***	269.724***	219.005***	145.255**	145.485**	342.639***	305.772***
	(60.879)	(58.833)	(59.490)	(60.576)	(75.997)	(72.098)	(62.088)	(58.431)	(36.788)	(35.398)
Constant	222.235	235.776	217.768	221.837	223.291	232.781	97.751	97.722	199.880	206.431
R^2	0.698	0.716	0.688	0.705	0.669	0.688	0.671	0.683	0.686	0.698
Observations	1246	1246	1096	1096	1596	1596	782	782	4720	4720
Number of clusters	47	47	44	44	60	60	30	30	181	181
Household FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Country FE	No	No	No	No	No	No	No	No	Yes	Yes

Table A4: Incomes and Income Diversification, Simpson's Diversity Index

Notes: This table shows results from a linear fixed effects model, where the income in 2023 USD PPP is regressed on an income diversification indicator, namely Simpson's Diversity Index (SDI). The index is calculated by SDI = 1 - $\sum_{s=1}^{6} (\frac{y_s}{Y})^2$, where y_s is the share of income derived from source s and Y the total income from the past 12 months. The income sources considered here are farming, employment, self-employment, livestock production, the collection of environmental products as well as other sources of income. All income sources are summarized under Panel B in Table 2. Controls included are age, gender, marital status, highest level of education of the household head, the size and socio-dependency of the household, the cropland, other assets owned and the locus of control. All controls are displayed in Table A5. Standard errors are clustered at the level of sampling, the enumeration area level. ***, **, and * indicate significance at the 1, 5, and 10 percent level.

	Kenya		Nambia		Tanzania		Zambia		Full Sample	
	2019	2023	2019	2023	2019	2023	2019	2023	2019	2023
Panel A: Characteristics of household head										
Age (in years)	46.13	49.21	51.31	54.19	48.91	52.09	48.28	52.83	48.63	51.94
	(15.49)	(15.77)	(17.55)	(16.01)	(14.73)	(14.94)	(16.56)	(17.55)	(16.02)	(15.96)
Male (0/1)	0.74	0.71	0.52	0.54	0.75	0.72	0.66	0.71	0.68	0.67
	(0.44)	(0.45)	(0.50)	(0.50)	(0.43)	(0.45)	(0.48)	(0.46)	(0.47)	(0.47)
Married (0/1)	0.73	0.68	0.59	0.55	0.73	0.69	0.68	0.68	0.69	0.65
	(0.45)	(0.47)	(0.49)	(0.50)	(0.44)	(0.46)	(0.47)	(0.47)	(0.46)	(0.48)
Years of schooling	4.69	4.73	7.37	7.67	5.97	6.06	5.94	5.83	5.96	6.04
	(5.03)	(5.12)	(4.46)	(4.32)	(2.88)	(3.21)	(3.90)	(3.91)	(4.18)	(4.28)
Highest level of education of household head										
No formal education (0/1)	0.45	0.46	0.18	0.15	0.14	0.15	0.22	0.20	0.24	0.24
	(0.50)	(0.50)	(0.39)	(0.36)	(0.34)	(0.36)	(0.41)	(0.40)	(0.43)	(0.43)
Some primary school education (0/1)	0.34	0.33	0.21	0.24	0.79	0.77	0.48	0.50	0.49	0.48
	(0.48)	(0.47)	(0.41)	(0.43)	(0.41)	(0.42)	(0.50)	(0.50)	(0.50)	(0.50)
Some secondary school education (0/1)	0.14	0.13	0.58	0.56	0.07	0.07	0.28	0.28	0.24	0.23
	(0.34)	(0.34)	(0.49)	(0.50)	(0.25)	(0.25)	(0.45)	(0.45)	(0.43)	(0.42)
Some post-secondary education (0/1)	0.07	0.08	0.03	0.05	0.01	0.02	0.02	0.03	0.03	0.04
	(0.25)	(0.27)	(0.16)	(0.22)	(0.10)	(0.12)	(0.15)	(0.17)	(0.17)	(0.20)
Panel B: Household characteristics										
Locus of control index (0-5)	3.77	4.27	3.68	4.11	3.84	4.18	3.22	4.18	3.68	4.19
	(1.09)	(1.01)	(1.18)	(0.98)	(1.09)	(1.07)	(1.48)	(1.04)	(1.20)	(1.03)
Household size	6.38	5.88	5.14	5.11	4.63	4.51	5.77	5.78	5.40	5.22
	(2.93)	(2.68)	(2.45)	(2.41)	(2.12)	(2.13)	(2.54)	(2.70)	(2.59)	(2.51)
Socio-dependancy ratio	1.46	1.35	1.18	1.12	0.95	0.99	1.58	1.18	1.25	1.15
	(1.17)	(1.07)	(1.23)	(1.06)	(0.87)	(0.86)	(1.34)	(0.97)	(1.15)	(0.99)
Cropland area (in ha)	0.44	0.50	2.84	2.65	1.39	1.37	4.97	2.49	2.07	1.62
	(0.45)	(0.57)	(3.05)	(2.96)	(0.98)	(1.02)	(4.50)	(2.00)	(2.88)	(1.96)
Area cultivated (in ha)	0.52	0.43	1.33	1.48	1.56	1.36	6.06	2.68	1.98	1.36
	(0.57)	(0.52)	(2.08)	(1.98)	(1.07)	(1.02)	(6.35)	(2.30)	(3.41)	(1.65)
Tropical Livestock Units (TLU)	5.43	5.09	4.96	6.50	0.96	1.78	2.13	4.01	3.26	4.12
	(11.16)	(10.57)	(10.90)	(15.27)	(3.67)	(12.92)	(4.06)	(6.92)	(8.46)	(12.30)
Asset index (sum)	2.82	2.76	4.59	5.01	3.54	4.82	2.91	3.69	3.49	4.13
	(2.20)	(2.26)	(2.50)	(2.69)	(2.10)	(2.30)	(2.23)	(2.52)	(2.34)	(2.60)
Observations	623	623	548	548	798	798	391	391	2360	2360

Table A5: Socio-demographic Characteristics of Balanced Panel

Notes: This table presents means and standard deviations in parentheses of the balanced panel of households that were interviewed in 2019 and 2023. This sample represents the main sample of interest for our analysis presented in Tables 4 and 5.

	Kenya		Nambia		Tanzania		Zambia		Full Sample	
	2019	2023	2019	2023	2019	2023	2019	2023	2019	2023
Q1. My life's course depends on me	0.90	0.98	0.89	0.91	0.91	0.90	0.81	0.90	0.88	0.92
	(0.30)	(0.15)	(0.32)	(0.28)	(0.29)	(0.30)	(0.39)	(0.30)	(0.32)	(0.26)
Q2. Success is a matter of hard work	0.97	0.97	0.95	0.98	0.96	0.97	0.89	0.97	0.95	0.97
	(0.17)	(0.16)	(0.22)	(0.13)	(0.20)	(0.17)	(0.32)	(0.16)	(0.22)	(0.16)
Q3. Abilities are more important than effort	0.55	0.79	0.54	0.86	0.69	0.82	0.53	0.77	0.59	0.81
	(0.50)	(0.41)	(0.50)	(0.35)	(0.46)	(0.38)	(0.50)	(0.42)	(0.49)	(0.39)
Q4. When I make plans, I'm most certain they will work	0.68	0.77	0.55	0.65	0.46	0.63	0.39	0.72	0.53	0.69
	(0.47)	(0.42)	(0.50)	(0.48)	(0.50)	(0.48)	(0.49)	(0.45)	(0.50)	(0.46)
Q5. I'm usually able to protect me and my community's interests	0.67	0.77	0.76	0.70	0.82	0.85	0.60	0.81	0.73	0.79
	(0.47)	(0.42)	(0.43)	(0.46)	(0.39)	(0.36)	(0.49)	(0.39)	(0.44)	(0.41)
Locus of control index (0-5)	3.77	4.27	3.68	4.11	3.84	4.18	3.22	4.18	3.68	4.19
	(1.09)	(1.01)	(1.18)	(0.98)	(1.09)	(1.07)	(1.48)	(1.04)	(1.20)	(1.03)
Observations	623	623	548	548	798	798	391	391	2360	2360

Table A6: Summary Statistics: Locus of Control

Notes: This table presents means and standard deviations in parentheses of the all locus of control questions. Each of the five questions were elicited on a five point scale: i) strongly disagree, ii) disagree, iii) indifferent, iv) agree, and v) strongly agree. We generate binary indicators that are equal to one for a perceived locus of control, so if households agreed or strongly agreed. The locus of control index is the sum over all binary indicators and captures the total perceived locus of control across all dimensions.

Dependent variable:	**Simpson index of income diversification**									
	Kenya		**Namibia**		**Tanzania**		**Zambia**		**Full Sample**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Aspiration gap	0.035	0.022	0.394**	0.413**	0.710***	0.695***	0.532*	0.468*	0.398***	0.390***
	(0.167)	(0.160)	(0.185)	(0.185)	(0.168)	(0.166)	(0.266)	(0.250)	(0.099)	(0.097)
Aspiration gap squared	-0.191	-0.158	-0.428***	-0.448***	-0.761***	-0.746***	-0.553***	-0.494**	-0.483***	-0.469***
	(0.134)	(0.129)	(0.139)	(0.139)	(0.133)	(0.130)	(0.197)	(0.185)	(0.077)	(0.075)
Constant	0.295	0.286	0.097	0.095	0.120	0.122	0.082	0.093	0.165	0.163
R^2	0.661	0.677	0.625	0.639	0.642	0.649	0.635	0.647	0.644	0.651
Observations	782	782	788	788	1044	1044	574	574	3188	3188
Number of clusters	47	47	44	44	60	60	30	30	181	181
Household FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Country FE	No	No	No	No	No	No	No	No	Yes	Yes
U-test results:										
Turning point	0.093	0.071	0.460	0.461	0.467	0.466	0.481	0.474	0.412	0.416
Fieller 95% CI	[.;.]	[.;.]	[0.067;0.574]	[0.112;0.576]	[0.367;0.525]	[0.364;0.524]	[-0.040;0.596]	[-0.182;0.600]	[0.298;0.477]	[0.304;0.481]
Sasabuchi p-value	0.419	0.446	0.021	0.016	0.000	0.000	0.030	0.039	0.000	0.000
Slope at min	0.034	0.022	0.379	0.398	0.709	0.693	0.501	0.440	0.397	0.389
Slope at max	-0.346	-0.293	-0.462	-0.483	-0.811	-0.797	-0.573	-0.519	-0.568	-0.548

Table A7: Aspirations and Income Diversification - Simpson's Diversity Index

Notes: This table shows results from a linear fixed effects, where we regress Simpson's Diversity Index (SDI) for income diversification on the aspiration gap and the aspiration gap squared. The Simpson's Diversity Index is calculated by SDI = 1 - $\sum_{s=1}^{6} (\frac{y_s}{Y})^2$, where y_s is the share of income derived from source s and Y the total income from the past 12 months. The income sources considered here are farming, employment, self-employment, livestock production, the collection of environmental products as well as other sources of income. All sources of incomes are summarized under Panel B in Table 2. Controls included are age, gender, marital status, highest level of education of the household head, the size and socio-dependency of the household, the cropland, other assets owned and the locus of control. Controls are displayed in Table A5. Standard errors are clustered at the level of sampling, the enumeration area level. ***, **, and * indicate significance at the 1, 5, and 10 percent level. Linear predictions from the estimation of the model for the full sample including controls (column 10) are illustrated in Figure A4. Figure A5 shows figures by country for columns (2), (4), (6), and (8).

Dependent variable:	Household has more than one source of monetary income (0/1)									
	Kenya		Namibia		Tanzania		Zambia		Full Sample	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Aspiration gap	0.756*	0.896**	0.904**	0.901**	1.710***	1.764***	1.758***	1.721***	1.250***	1.248***
	(0.389)	(0.378)	(0.400)	(0.401)	(0.406)	(0.396)	(0.536)	(0.589)	(0.222)	(0.219)
Aspiration gap squared	-0.988***	-1.057***	-1.079***	-1.087***	-1.898***	-1.955***	-1.689***	-1.671***	-1.428***	-1.417***
	(0.310)	(0.303)	(0.323)	(0.328)	(0.313)	(0.305)	(0.418)	(0.461)	(0.173)	(0.172)
Constant	0.442	0.384	0.248	0.254	0.322	0.317	0.070	0.090	0.307	0.304
R^2	0.604	0.619	0.617	0.623	0.654	0.665	0.637	0.645	0.630	0.634
Observations	782	782	788	788	1044	1044	574	574	3188	3188
Number of clusters	47	47	44	44	60	60	30	30	181	181
Household FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Country FE	No	No	No	No	No	No	No	No	Yes	Yes
U-test results:										
Turning point	0.382	0.424	0.419	0.414	0.450	0.451	0.520	0.515	0.438	0.440
Fieller 95% CI	[-0.036;0.501]	[0.142;0.527]	[0.108;0.519]	[0.101;0.518]	[0.346;0.509]	[0.355;0.507]	[0.376;0.591]	[0.339;0.587]	[0.368;0.485]	[0.372;0.487]
Sasabuchi p-value	0.030	0.011	0.016	0.016	0.000	0.000	0.001	0.004	0.000	0.000
Slope at min	0.751	0.891	0.868	0.864	1.706	1.759	1.663	1.627	1.247	1.245
Slope at max	-1.221	-1.217	-1.255	-1.273	-2.087	-2.146	-1.620	-1.620	-1.605	-1.586

Table A8: Aspirations and Income Diversification - Extensive Margin (Robustness)

Notes: This table presents results from linear fixed effects model, where we regress an indicator that is one if the household has more than one source of income and zero otherwise on the aspiration gap and the aspiration gap squared. Income sources considered here are farming, self-employment, livestock production, the collection of environmental products as well as other sources of income. We exclude here employment for robustness as outlined in Section 3.1. The main result which include employment as a source of income are displayed in Table 4. Controls included are age, gender, marital status, highest level of education of the household head, the size and socio-dependency of the household, the cropland, other assets owned and the locus of control. Controls are displayed in Table A5. Standard errors are clustered at the level of sampling, the enumeration area level. . ***, **, and * indicate significance at the 1, 5, and 10 percent level.

Dependent variable:	Number of monetary income sources (0-5)									
	Kenya		Namibia		Tanzania		Zambia		Full Sample	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Aspiration gap	1.465***	1.540***	3.625***	3.880***	4.233***	4.332***	4.482***	4.328***	3.193***	3.183***
	(0.410)	(0.404)	(0.637)	(0.691)	(0.570)	(0.571)	(0.928)	(0.978)	(0.311)	(0.307)
Aspiration gap squared	-1.894***	-1.882***	-3.830***	-4.053***	-4.624***	-4.725***	-4.453***	-4.320***	-3.560***	-3.532***
	(0.355)	(0.353)	(0.534)	(0.576)	(0.503)	(0.499)	(0.759)	(0.773)	(0.270)	(0.267)
Constant	0.477	0.397	-0.183	-0.249	0.048	0.039	-0.278	-0.282	0.123	0.104
Pseudo-R^2	0.129	0.132	0.131	0.134	0.166	0.168	0.164	0.168	0.148	0.150
Observations	754	754	744	744	958	958	508	508	2964	2964
Household FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Country FE	No	No	No	No	No	No	No	No	Yes	Yes
U-test results:										
Turning point	0.387	0.409	0.473	0.479	0.458	0.458	0.503	0.501	0.448	0.451
Fieller 95% CI	[0.263;0.460]	[0.299;0.477]	[0.412;0.517]	[0.419;0.520]	[0.420;0.487]	[0.421;0.487]	[0.433;0.550]	[0.415;0.554]	[0.419;0.473]	[0.422;0.475]
Sasabuchi p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Slope at min	1.456	1.531	3.497	3.744	4.223	4.321	4.232	4.085	3.185	3.175
Slope at max	-2.324	-2.225	-4.036	-4.226	-5.015	-5.118	-4.424	-4.313	-3.928	-3.882

Table A9: Aspirations and Income Diversification - Intensive Margin (Robustness)

Notes: This table presents regression results from a Poisson fixed effects model, where we regress the number of income sources on the aspiration gap and the aspiration gap squared. The income sources considered are summarized in Table 2 under Panel B, namely farming, self-employment, livestock production, and collection of environmental products. We exclude here employment as a source of income for robustness of our results displayed in Table 5 and as outlined in Section 3.1. Controls included are age, gender, marital status, highest level of education of the household head, the size and socio-dependency of the household, the cropland, other assets owned and the locus of control. Controls are displayed in Table A5. Standard errors are clustered at the level of sampling, the enumeration area level. . ***, **, and * indicate significance at the 1, 5, and 10 percent level.

Dependent variable:	**Household has more than one monetary source of income (0/1)**								
Winsorization cutoff:	**1 pct**	**2 pct**	**3 pct**	**4 pct**	**5 pct**	**6 pct**	**7 pct**	**8 pct**	**9 pct**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Aspiration gap	1.697***	1.687***	1.686***	1.702***	1.676***	1.694***	1.430***	1.603***	1.539***
	(0.201)	(0.196)	(0.201)	(0.197)	(0.195)	(0.187)	(0.184)	(0.185)	(0.177)
Aspiration gap squared	-1.864***	-1.859***	-1.856***	-1.866***	-1.849***	-1.859***	-1.662***	-1.795***	-1.744***
	(0.160)	(0.157)	(0.159)	(0.156)	(0.154)	(0.149)	(0.145)	(0.147)	(0.142)
Constant	0.360	0.364	0.364	0.357	0.367	0.362	0.438	0.389	0.405
R^2	0.663	0.665	0.665	0.667	0.664	0.666	0.664	0.664	0.662
Observations	3114	3136	3156	3166	3188	3212	3270	3288	3342
Number of clusters	181	181	181	181	181	181	181	181	181
Household FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
U-test results:									
Turning point	0.455	0.454	0.454	0.456	0.453	0.456	0.430	0.447	0.441
Fieller 95% CI	[0.413;0.488]	[0.413;0.486]	[0.411;0.487]	[0.415;0.489]	[0.412;0.486]	[0.417;0.487]	[0.382;0.468]	[0.405;0.480]	[0.400;0.475]
Sasabuchi p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Slope at min	1.692	1.682	1.682	1.698	1.671	1.690	1.427	1.599	1.535
Slope at max	-2.032	-2.031	-2.027	-2.030	-2.022	-2.024	-1.893	-1.986	-1.949

Table A10: Income Diversification, Extensive Margin - Sensitivity Analysis

Notes: This table presents results from linear fixed effects model, where we regress an indicator that is one if the household has more than one source of income and zero otherwise on the aspiration gap and the aspiration gap squared. The results serve as a sensitivity analysis for main results displayed in Table 4. Here, we show results for different winsorization cutoffs of monthly incomes and income aspirations - from the bottom and top 1 percent to the bottom and top 9 percent level, i.e. Column (1) shows result where the aspiration gap was constructed using a winsorized version of the actual monthly - and the aspired monthly income. Column (5) shows the results that correspond to the specification used in our main results, Table 4, Column (10). All regressions contain controls and are estimated using household, year and country fixed effects. Controls included are age, gender, marital status, highest level of education of the household head, the size and socio-dependency of the household, the cropland, other assets owned and the locus of control. Controls are displayed in Table A5. Standard errors are clustered at the enumeration area level. ***, **, and * indicate significance at the 1, 5, and 10 percent level.

Dependent variable:	Number of monetary income sources (0-6)								
Winsorization cutoff:	1 pct	2 pct	3 pct	4 pct	5 pct	6 pct	7 pct	8 pct	9 pct
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Aspiration gap	3.267***	3.151***	3.193***	3.174***	3.141***	3.174***	2.667***	3.071***	2.976***
	(0.300)	(0.290)	(0.292)	(0.301)	(0.286)	(0.280)	(0.265)	(0.279)	(0.259)
Aspiration gap squared	-3.639***	-3.554***	-3.589***	-3.569***	-3.552***	-3.575***	-3.177***	-3.494***	-3.418***
	(0.261)	(0.254)	(0.254)	(0.260)	(0.249)	(0.246)	(0.231)	(0.243)	(0.228)
Constant	0.257	0.291	0.279	0.280	0.294	0.285	0.423	0.311	0.333
Pseudo-R^2	0.163	0.164	0.164	0.163	0.163	0.163	0.164	0.164	0.164
Observations	2990	3012	3032	3042	3064	3088	3146	3164	3218
Number of clusters	181	181	181	181	181	181	181	181	181
Household FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
U-test results:									
Turning point	0.449	0.443	0.445	0.445	0.442	0.444	0.420	0.439	0.435
Fieller 95% CI	[0.423;0.471]	[0.417;0.465]	[0.418;0.467]	[0.417;0.468]	[0.416;0.464]	[0.419;0.465]	[0.388;0.446]	[0.412;0.462]	[0.410;0.457]
Sasabuchi p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Slope at min	15917.021	15544.024	15695.967	15608.594	15537.777	15635.172	13896.028	15281.607	14947.732
Slope at max	-4.011	-3.956	-3.985	-3.963	-3.964	-3.976	-3.687	-3.917	-3.859

Table A11: Income Diversification - Intensive Margin - Sensitivity Analysis

Notes: This table presents regression results from a Poisson fixed effects model, where we regress the number of income sources on the aspiration gap and the aspiration gap squared. The results serve as a sensitivity analysis for main results displayed in Table 5. Here, we show results for different winsorization cutoffs of monthly incomes and income aspirations - from the bottom and top 1 percent to the bottom and top 9 percent level, i.e. Column (1) shows result where the aspiration gap was constructed using a winsorized version of the actual monthly - and the aspired monthly income. Column (5) shows the results that correspond to the specification used in our main results, Table 5, Column (10). All regressions contain controls and are estimated using household, year and country fixed effects. Controls included are age, gender, marital status, highest level of education of the household head, the size and socio-dependency of the household, the cropland, other assets owned and the locus of control. Controls are displayed in Table A5. Standard errors are clustered at the enumeration area level. ***, **, and * indicate significance at the 1, 5, and 10 percent level.

Variable	(1) Non-Attritors N/[Clusters]	Mean/SE	(2) Attritors N/[Clusters]	Mean/SE	(3) Total N/[Clusters]	Mean/SE	T-test Difference (1)-(2)	Normalized difference (1)-(2)
Age of household head	2360	48.628	304	46.928	2664	48.434	1.700*	0.105
	[181]	(0.407)	[128]	(1.110)	[181]	(0.408)		
Household head is male	2360	0.679	304	0.681	2664	0.679	-0.002	-0.004
	[181]	(0.013)	[128]	(0.031)	[181]	(0.012)		
Household head is married	2360	0.689	304	0.618	2664	0.681	0.070*	0.150
	[181]	(0.012)	[128]	(0.032)	[181]	(0.012)		
No formal education	2360	0.243	304	0.296	2664	0.249	-0.053	-0.122
	[181]	(0.015)	[128]	(0.050)	[181]	(0.016)		
Some primary school education	2360	0.485	304	0.312	2664	0.465	0.173***	0.346
	[181]	(0.021)	[128]	(0.035)	[181]	(0.020)		
Some secondary school education	2360	0.272	304	0.391	2664	0.285	-0.120**	-0.265
	[181]	(0.018)	[128]	(0.042)	[181]	(0.018)		
Household size	2360	5.397	304	4.773	2664	5.326	0.624***	0.240
	[181]	(0.086)	[128]	(0.207)	[181]	(0.086)		
Socio-dependancy ratio	2360	1.245	304	1.207	2664	1.241	0.038	0.033
	[181]	(0.033)	[128]	(0.085)	[181]	(0.032)		
Cropland area (in ha)	2360	2.070	304	1.817	2664	2.041	0.254**	0.088
	[181]	(0.135)	[128]	(0.217)	[181]	(0.131)		
Asset index (sum)	2360	3.492	304	3.480	2664	3.491	0.012	0.005
	[181]	(0.089)	[128]	(0.245)	[181]	(0.092)		
Locus of control index (0-5)	2360	3.681	304	3.628	2664	3.675	0.052	0.043
	[181]	(0.037)	[128]	(0.073)	[181]	(0.035)		
Aspiration gap	1920	0.755	255	0.740	2175	0.754	0.015	0.060
	[181]	(0.008)	[120]	(0.017)	[181]	(0.008)		
Number of income sources	2360	1.732	304	1.520	2664	1.708	0.212***	0.194
	[181]	(0.037)	[128]	(0.070)	[181]	(0.036)		
More than one source of income	2360	0.554	304	0.441	2664	0.541	0.113***	0.227
	[181]	(0.015)	[128]	(0.036)	[181]	(0.015)		
Simpson Diversity Index	2360	0.193	304	0.145	2664	0.188	0.048***	0.218
	[181]	(0.007)	[128]	(0.016)	[181]	(0.007)		
F-test of joint significance (p-value)							0.000***	

Table A12: Balancing Test: Attrition

Notes: This table shows a balancing test between households that attrited in 2023. We show simple ttest difference as well as normalized differences as suggested by Imbens and Wooldridge (2009). Standard errors are clustered at the enumeration area level. ***, **, and * indicate significance at the 1, 5, and 10 percent level.

Dependent variable:	**Observed in 2023 (round 2)**									
	Kenya		**Namibia**		**Tanzania**		**Zambia**		**Full Sample**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Aspiration gap	0.138	0.230	0.665*	0.584	-0.103	-0.034	-1.074***	-0.871**	-0.008	0.025
	(0.195)	(0.197)	(0.352)	(0.373)	(0.219)	(0.220)	(0.310)	(0.340)	(0.141)	(0.141)
Aspiration gap squared	-0.199	-0.209	-0.449	-0.396	0.057	-0.006	0.944***	0.715**	0.013	-0.011
	(0.169)	(0.164)	(0.279)	(0.297)	(0.178)	(0.178)	(0.285)	(0.288)	(0.116)	(0.114)
Mean Dep. Var	0.896	0.837	0.619	0.646	0.956	0.945	1.076	1.082	0.880	0.871
Observations	591	591	525	525	676	676	383	383	2175	2175
Adjusted R^2	0.005	0.051	0.008	0.025	-0.002	0.013	0.038	0.115	0.006	0.027
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Country FE	No	No	No	No	No	No	No	No	Yes	Yes

Table A13: Aspirations and Attrition

Notes: This table presents results from a linear regression model, where the dependent variable is binary and takes on value one if the household was observed in 2023 and zero otherwise. The dependent variable is regressed on the income aspiration gap and income aspiration gap squared. Controls included are age, gender, marital status, highest level of education of the household head, the size and socio-dependency of the household, the cropland, other assets owned and the locus of control. Controls are displayed in Table A5. Standard errors are clustered at the enumeration area level. ***, **, and * indicate significance at the 1, 5, and 10 percent level.

Dependent variable:	**Re-interviewed in 2023**				
	Kenya (1)	**Namibia** (2)	**Tanzania** (3)	**Zambia** (4)	**Full Sample** (5)
Age of household head	0.240***	-0.019	0.021	0.159	0.098**
	(0.072)	(0.077)	(0.077)	(0.110)	(0.042)
Household head is male	-0.236	-0.167	-0.357**	-0.012	-0.209***
	(0.144)	(0.109)	(0.143)	(0.121)	(0.064)
Household size	0.086	0.121*	0.013	0.540***	0.105*
	(0.107)	(0.069)	(0.130)	(0.135)	(0.058)
Household head is married	0.051	0.271***	0.331**	-0.176	0.169***
	(0.144)	(0.101)	(0.129)	(0.153)	(0.063)
No formal education	-0.142	0.097	0.102	-0.040	-0.017
	(0.147)	(0.089)	(0.094)	(0.126)	(0.060)
Some primary school education	0.108	0.003	0.276***	0.109	0.157***
	(0.114)	(0.069)	(0.084)	(0.127)	(0.052)
Cropland (in ha)	0.260*	-0.003	-0.065	0.193*	0.043
	(0.145)	(0.079)	(0.079)	(0.101)	(0.043)
Asset index	0.097	-0.048	0.010	-0.331***	-0.053
	(0.129)	(0.082)	(0.095)	(0.109)	(0.052)
Socio-dependancy ratio	-0.074	0.017	-0.010	0.096	-0.005
	(0.081)	(0.064)	(0.079)	(0.098)	(0.037)
Locus of control index	0.019	0.056	0.053	-0.059	0.027
	(0.079)	(0.058)	(0.061)	(0.099)	(0.033)
Number of income sources	-0.007	0.162	0.111	-0.324*	-0.002
	(0.160)	(0.130)	(0.135)	(0.178)	(0.075)
More than one source of income	0.031	0.072	-0.007	0.336	0.113
	(0.183)	(0.124)	(0.158)	(0.231)	(0.085)
Simpson Diversity Index	0.208	-0.167	0.027	0.205	0.049
	(0.158)	(0.120)	(0.160)	(0.189)	(0.074)
Aspiration gap	0.119	0.461	-0.166	-1.466***	-0.053
	(0.347)	(0.327)	(0.471)	(0.506)	(0.201)
Aspiration gap squared	-0.072	-0.340	0.161	1.568***	0.133
	(0.337)	(0.349)	(0.486)	(0.553)	(0.208)
Kenya					-0.002
					(0.076)
Namibia					-0.072
					(0.069)
Tanzania					0.042
					(0.068)
Observations	591	525	676	383	2175
Chi2 Statistic	42.744	45.746	20.946	126.522	86.314
Prob > chi2	0.000	0.000	0.139	0.000	0.000

Table A14: Aspirations and Attrition

Notes: This table presents results from a linear regression model, where the dependent variable is binary and takes on value one if the household was observed in 2023 and zero otherwise on our main set of variables used in the analysis, the main outcomes and indicators of income diversification, the characteristics of the household head and the household. ***, **, and * indicate significance at the 1, 5, and 10 percent level.

Dependent variable:	Household has more than one source of monetary income (0/1)				
	Kenya (1)	**Namibia** (2)	**Tanzania** (3)	**Zambia** (4)	**Full Sample** (5)
Aspiration gap	1.144***	1.305***	2.088***	1.932***	1.656***
	(0.338)	(0.398)	(0.328)	(0.533)	(0.197)
Aspiration gap squared	-1.351***	-1.436***	-2.312***	-2.061***	-1.845***
	(0.271)	(0.323)	(0.254)	(0.402)	(0.155)
Constant	0.490	0.305	0.376	0.260	0.361
R^2	0.673	0.642	0.658	0.668	0.655
Observations	782	788	1044	574	3188
Number of clusters	47	44	60	30	181
Household FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Controls	No	No	No	No	No
Country FE	No	No	No	No	Yes
Inverse Probability Weights	Yes	Yes	Yes	Yes	Yes
U-test results:					
Turning point	0.424	0.454	0.451	0.469	0.449
Fieller 95% CI	[0.277;0.500]	[0.306;0.528]	[0.389;0.495]	[0.328;0.543]	[0.406;0.482]
Sasabuchi p-value	0.001	0.001	0.000	0.001	0.000
Slope at min	1.138	1.257	2.082	1.816	1.652
Slope at max	-1.558	-1.567	-2.537	-2.189	-2.035

Table A15: Aspirations and Income Diversification - Extensive Margin (IPW)

Notes: This table shows results from the same specification as displayed in Table 4, including Inverse Probability Weights (IPW). Weights are obtained from a probit model summarized in Table A14. Standard errors are clustered at the level of sampling, the enumeration area level. ***, **, and * indicate significance at the 1, 5, and 10 percent level.

Dependent variable:	Number of monetary income sources (0-6)				
	Kenya (1)	**Namibia** (2)	**Tanzania** (3)	**Zambia** (4)	**Full Sample** (5)
Aspiration gap	1.639***	3.455***	4.251***	4.245***	3.201***
	(0.412)	(0.576)	(0.508)	(0.854)	(0.292)
Aspiration gap squared	-2.120***	-3.763***	-4.661***	-4.494***	-3.633***
	(0.360)	(0.500)	(0.452)	(0.721)	(0.254)
Constant	0.650	0.078	0.152	0.028	0.292
Pseudo-R^2	0.151	0.155	0.169	0.166	0.161
Observations	776	774	984	530	3064
Household FE	Yes	Yes	Yes	Yes	Yes
Year FE	No	No	No	No	No
Controls	No	No	No	No	No
Country FE	No	No	No	No	Yes
Inverse Probability Weights	Yes	Yes	Yes	Yes	Yes
U-test results:					
Turning point	0.386	0.459	0.456	0.472	0.441
Fieller 95% CI	[0.284;0.447]	[0.406;0.497]	[0.423;0.482]	[0.407;0.513]	[0.414;0.462]
Sasabuchi p-value	0.000	0.000	0.000	0.000	0.000
Slope at min	1.628	3.329	4.241	3.992	3.193
Slope at max	-2.602	-4.071	-5.070	-4.744	-4.065

Table A16: Aspirations and Income Diversification - Intensive Margin (IPW)

Notes: This table shows results from the same specification as displayed in Table 5, including Inverse Probability Weights (IPW). Weights are obtained from a probit model summarized in Table A14. ***, **, and * indicate significance at the 1, 5, and 10 percent level.

Dependent variable:	Simpson index of income diversification				
	Kenya (1)	Namibia (2)	Tanzania (3)	Zambia (4)	Full Sample (5)
Aspiration gap	0.050	0.387**	0.721***	0.488*	0.398***
	(0.166)	(0.181)	(0.166)	(0.254)	(0.097)
Aspiration gap squared	-0.197	-0.421***	-0.768***	-0.524***	-0.480***
	(0.133)	(0.136)	(0.131)	(0.189)	(0.075)
Constant	0.277	0.099	0.114	0.086	0.157
R^2	0.658	0.629	0.642	0.625	0.641
Observations	782	788	1044	574	3188
Number of clusters	47	44	60	30	181
Household FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Controls	No	No	No	No	No
Country FE	No	No	No	No	Yes
Inverse Probability Weights	Yes	Yes	Yes	Yes	Yes
U-test results:					
Turning point	0.127	0.459	0.469	0.465	0.415
Fieller 95% CI	[.;.]	[0.072;0.575]	[0.374;0.526]	[-0.108;0.586]	[0.305;0.479]
Sasabuchi p-value	0.384	0.020	0.000	0.035	0.000
Slope at min	0.049	0.373	0.719	0.458	0.397
Slope at max	-0.344	-0.455	-0.816	-0.560	-0.562

Table A17: Aspirations and Income Diversification - Simpson's Diversity Index (IPW)

Notes: This table shows results from the same specification as displayed in Table A7, including Inverse Probability Weights (IPW). Weights are obtained from a probit model summarized in Table A14. ***, **, and * indicate significance at the 1, 5, and 10 percent level.

A.5 Figures

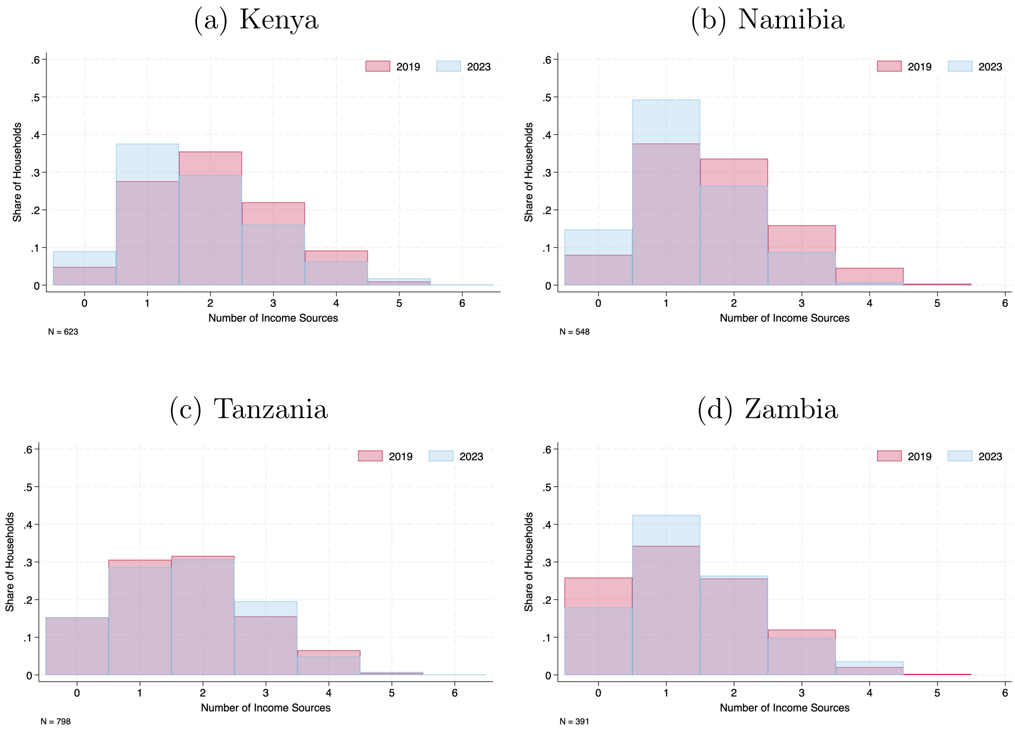


Figure A1: Number of Income Sources, by Country

Notes: These figures illustrate the number of income sources by country and year as also summarized in Table 2. The distributions are similar to a Poisson distribution, which motivates the choice for a Poisson fixed effects model when estimating the relationship between the aspiration gap and the number of income sources, our intensive margin measure of income diversification.

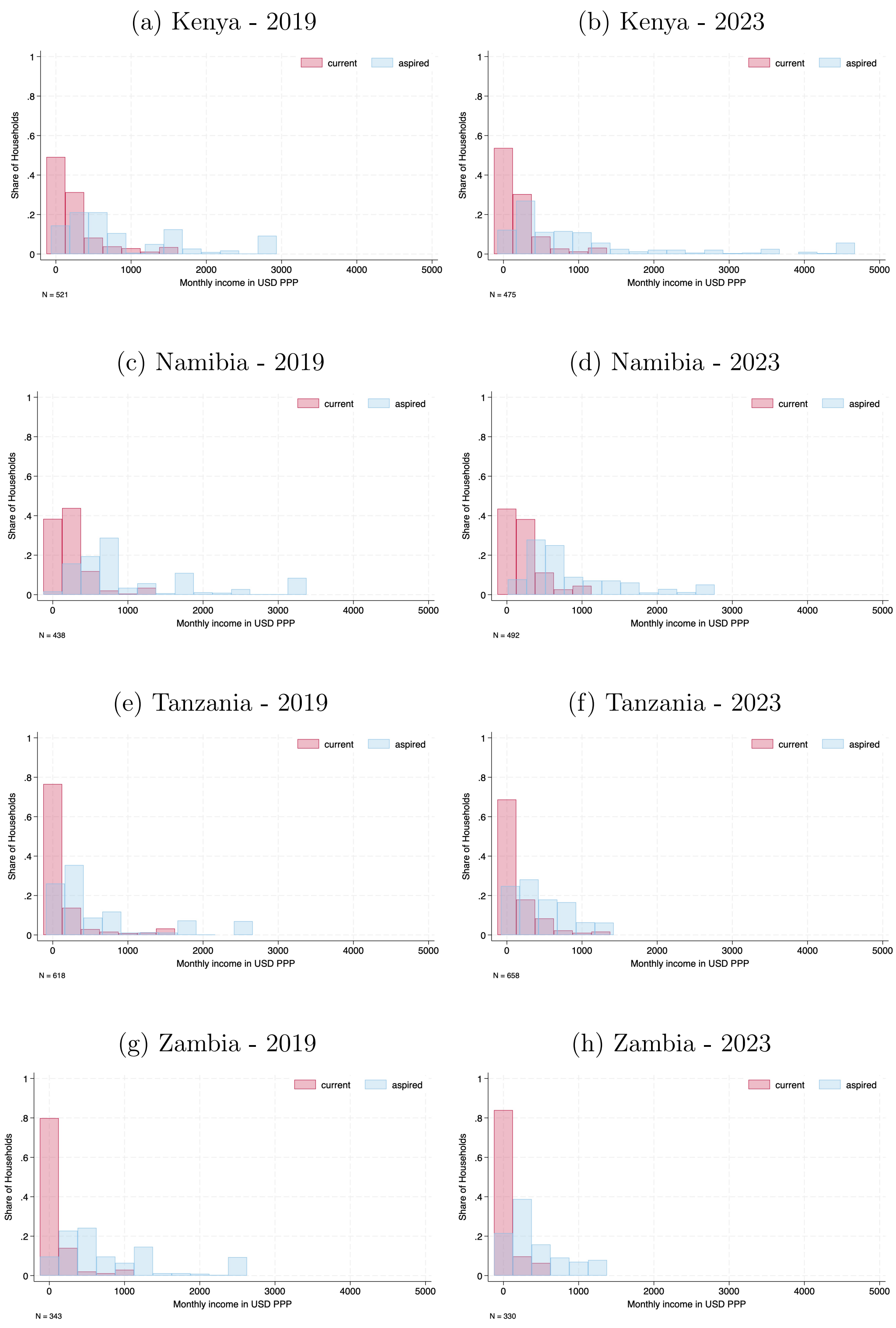
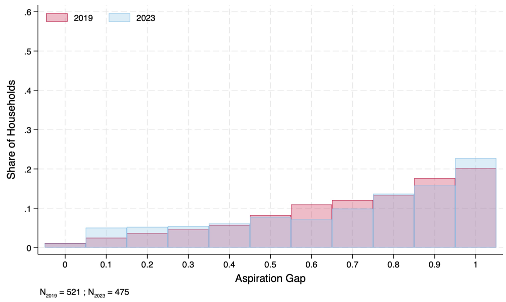


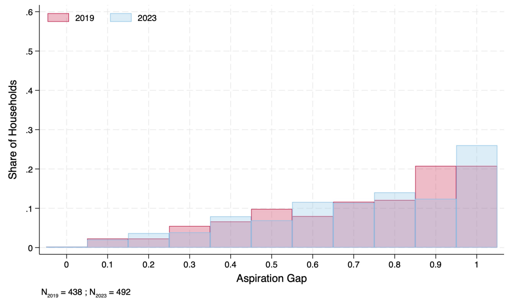
Figure A2: Comparison: Current vs. Aspired Income, by Country and Year

Notes: These figures illustrate for each country and year the current- and the aspired monthly income in 2023 USD PPP. 2019 monetary values are inflated using the national non-food Consumer Price Indices (CPI) and similarly as the 2023 monetary values converted to 2023 USD PPP terms. For details on the conversion to 2023 USD PPP see Section A.3 in the Appendix. Both the current- and the aspired incomes are winsorized at the top and bottom 5th percentile to account for outliers.

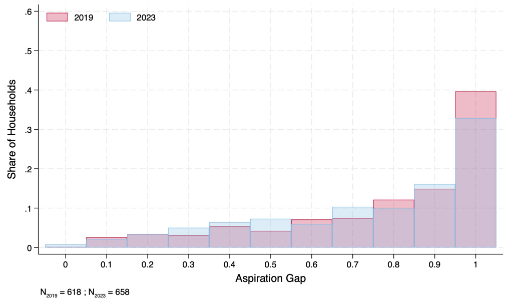
(a) Kenya



(b) Namibia



(c) Tanzania



(d) Zambia

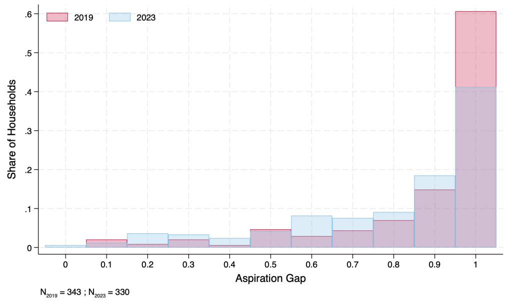


Figure A3: Aspiration Gaps

Notes: These Figures illustrate the aspiration gaps for the final sample of analysis for each country and year as used in Tables 4, 5, and Table A7 to test for an inverted U-shaped relationship between the aspiration gap and our indicators for income diversification. The current- and aspired monthly incomes as well as the aspiration gap are also summarized in Table 3.

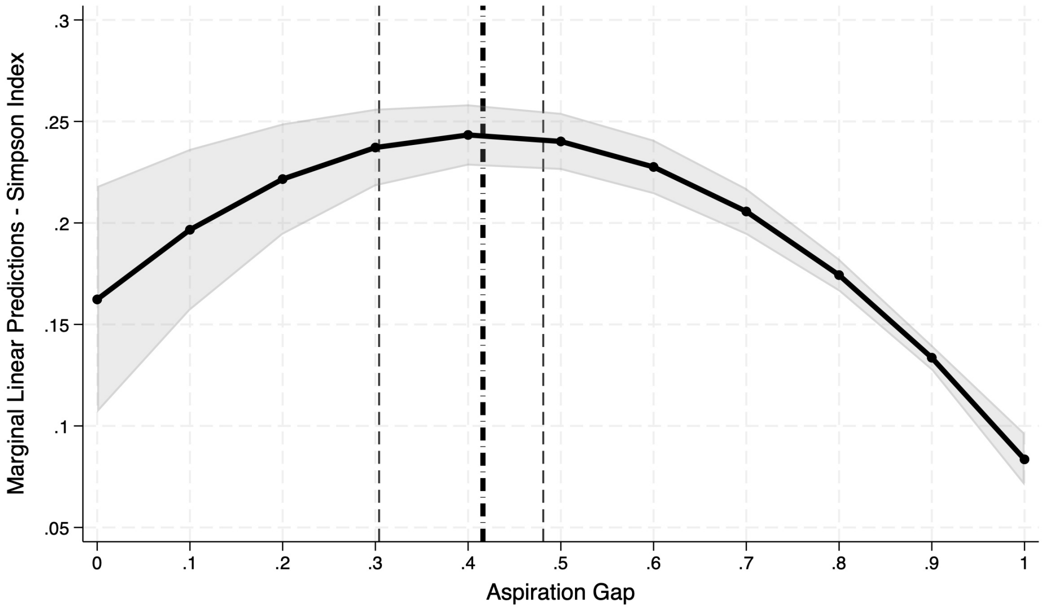


Figure A4: Linear prediction - Simpson's Diversity Index

Notes: This figure illustrates the linear prediction from estimations summarized in Table A7, column (10). The thick dot-dashed line indicates the estimated turning point and the dashed lines the Fieller confidence intervals (Fieller, 1954). The gray shaded area depicts the 95% confidence interval band.

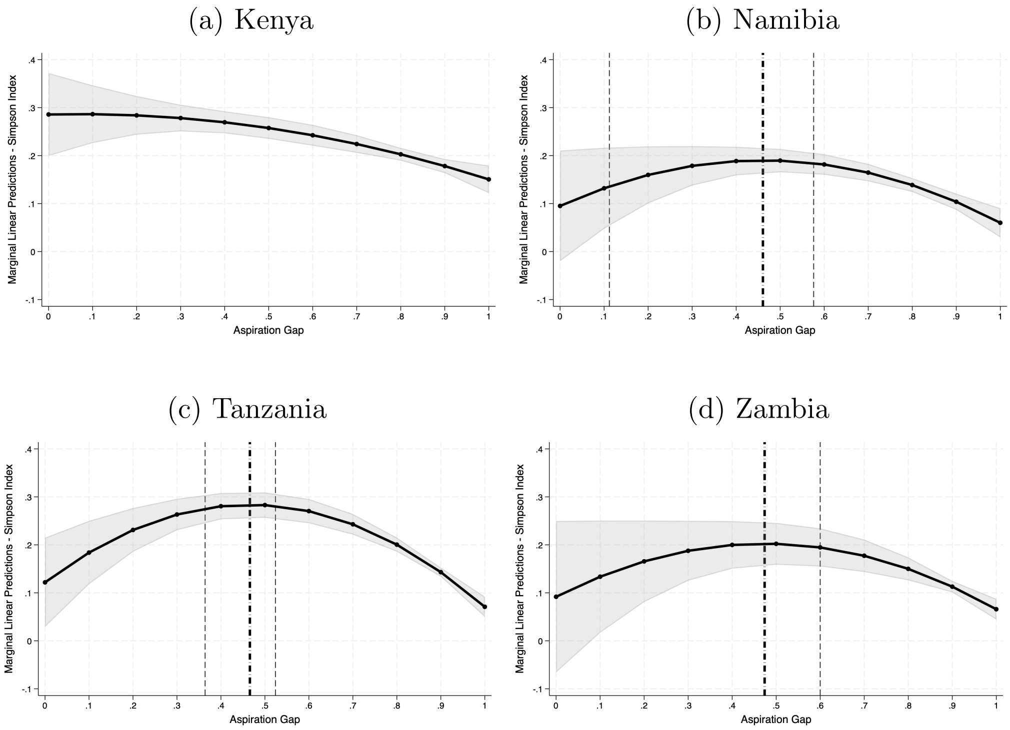


Figure A5: Linear prediction - Simpson's Diversity Index, by Country

Notes: These figures illustrates the linear prediction from estimations summarized in Table A7, columns (2), (4), (6), and (8) including controls. Thick dot-dashed lines indicate the estimated turning point and lighter the dashed lines the Fieller confidence intervals (Fieller, 1954). Gray shaded areas depict the 95% confidence interval band.

A.6 Semi-parametric Estimations

Robinson (1988)'s Double Residual Method

The goal is to estimate a partially linear model, which is a powerful tool to test for a non-linear relationship between regressors and the outcome.

$$E[y_i|x_i, z_i] = \underbrace{x_i\beta}_{linear} + \underbrace{h_0(z_i)}_{non-linear} \tag{R1}$$

To estimate non-linear, non-parametric relationships we rely on Kernel regression, which does not allow for $\beta_0 \neq 0$.

In the following, we detail the Robinson (1988)'s double residual method by starting with a following model that contains a linear and a non-linear component:

$$y_i = x_i\beta_0 + h_0(z_i) + \epsilon_i \tag{R2}$$

Then we take the expectation: $E[\cdot|z_i]$:

$$E[y_i|x_i] = E[y_i|z_i]\beta_0 + h_0(z_i) + E[\epsilon_i|z_i] \tag{R3}$$

where taking $E[\cdot|z_i]$ of $h_0(z_i)$ is no longer stochastic and will yield a number.

Next, we subtract R3 from R2:

$$y_i - E[y_i|x_i] = x_i\beta_0 - E[y_i|z_i]\beta_0 + h_0 - h_0 + (z_i) + \epsilon_i - E[\epsilon_i|z_i] \tag{R4}$$

and end up with:

$$\tilde{y}_i = \tilde{x}_i\beta_0 + \tilde{\epsilon}_i \tag{R5}$$

where $\tilde{a}_i = a_i - E[a_i|z_i]$ is referred to as "residualization", which can be estimated using OLS and the residulization $E[a_i|z_i]$ by kernel regression.

Finally, after estimation - we can move in Equation R3, $x_i\hat{\beta}$ over to get:

$$y_i - x_i\hat{\beta} = h_0(z_i) + \epsilon_i \tag{R6}$$

to obtain $\hat{h}(\cdot)$ from kernel regression of $y_i - x_i\hat{\beta}$ on z_i

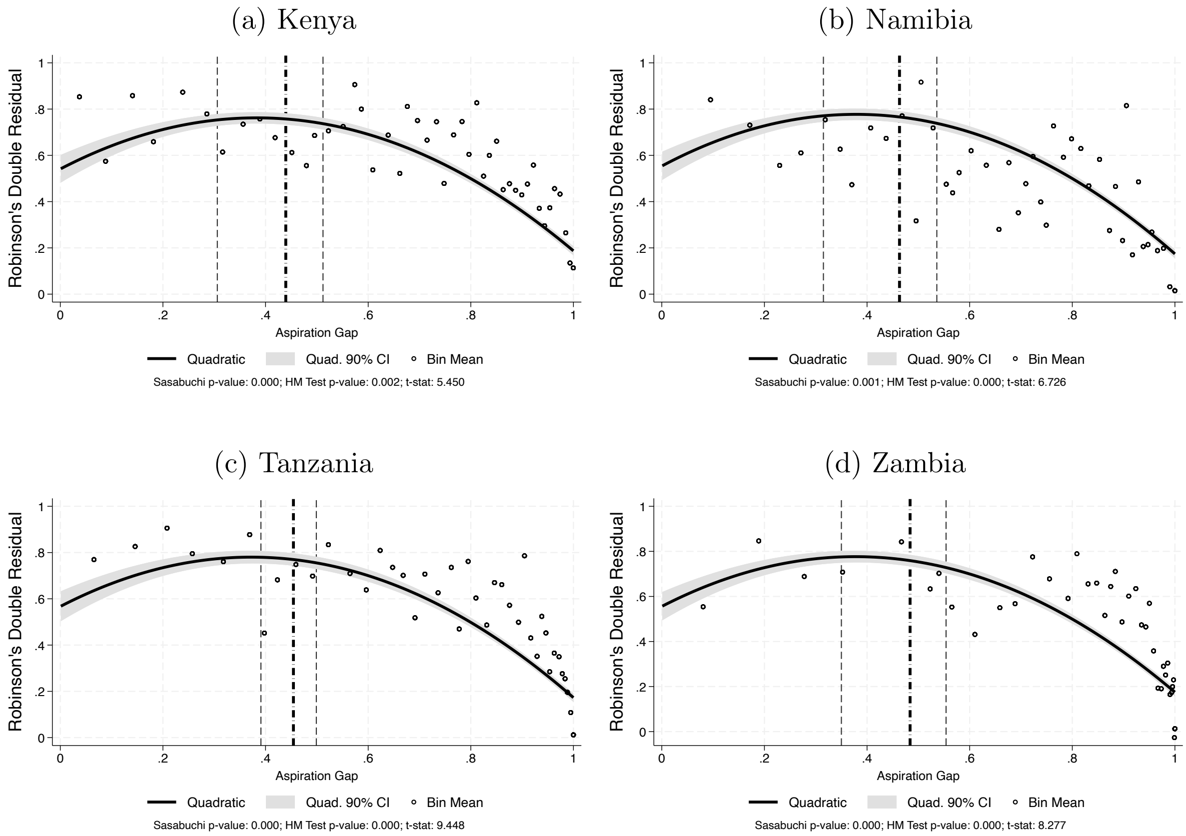


Figure A6: Aspiration Gap and Income Diversification, Extensive Margin

Notes: The Figures illustrate the quadratic and non-parametric regression fit of the income aspiration gap in a semi-parametric model where the dependent variable is equal to one if the household has more than one source of income and zero otherwise. On the vertical axis are the Robinson's double residuals (Robinson, 1988), which are estimated semi-parametrically using Equation 3. Robinson residuals partialled out the parametric fit from the regression and allow to test for a fit of a non-parametric term estimating a full model comprising of a para- and non-parametric component. The thick vertical dashed-dot line indicates the turning point of the quadratic fit. The thin vertical dashed lines indicate the 90% Fieller confidence intervals for turning point. Bin means are the mean value of the outcome variable for the mean of the income aspiration gap for 50 quantiles of the aspiration gap. Sasabuchi p-values are for the null hypothesis that the income aspiration gap is monotone over the interval. The HM p-values are retrieved from a Hardle and Mammen (1993) test with 500 bootstrap replications and a null hypothesis that the non-parametric and quadratic fits are not different, i.e. the same.Controls included are age, gender, marital status, highest level of education of the household head, the size and socio-dependency of the household, the cropland, other assets owned and the locus of control. Controls are displayed in Table A5.

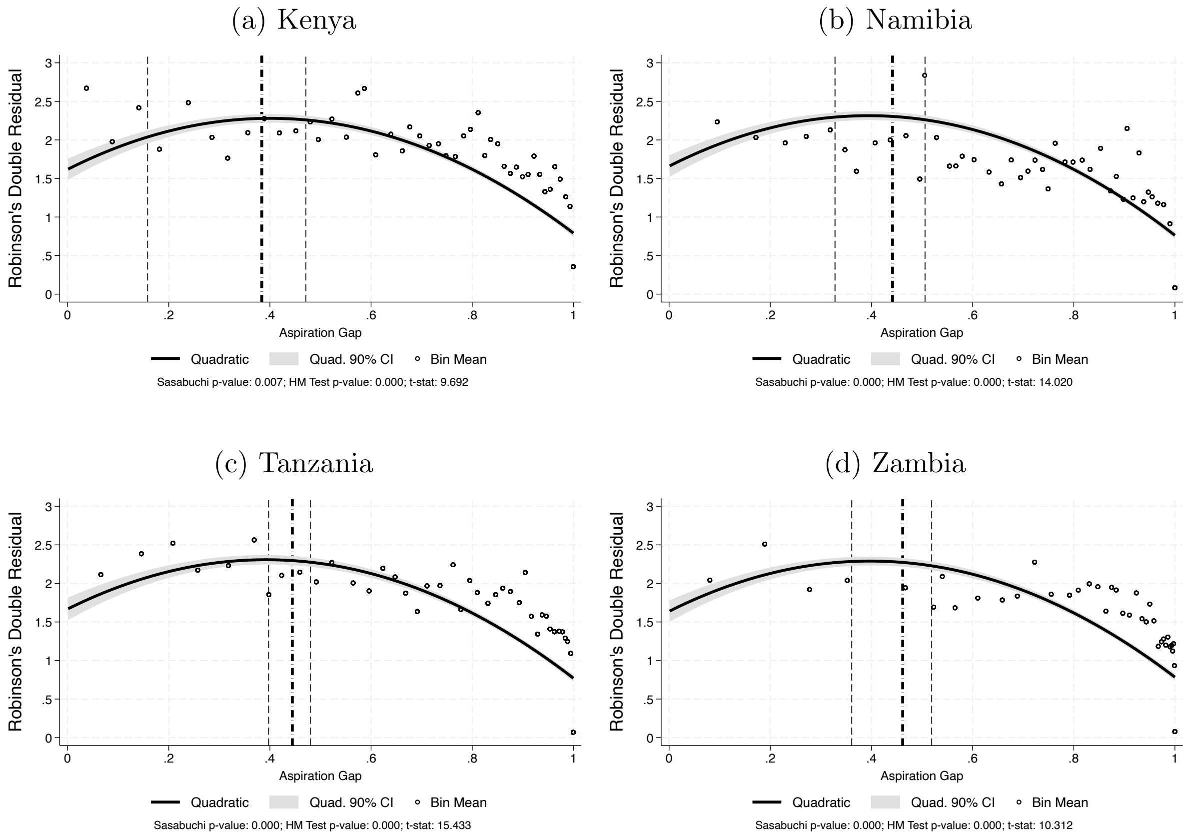


Figure A7: Aspiration Gap and Income Diversification, Intensive Margin

Notes: The Figures illustrate the quadratic and non-parametric regression fit of the income aspiration gap in a semi-parametric model where the dependent variable captures the number of income sources that a household has. On the vertical axis are the Robinson's double residuals (Robinson, 1988), which are estimated semi-parametrically using Equation 3. Robinson residuals partialled out the parametric fit from the regression and allow to test for a fit of a non-parametric term estimating a full model comprising of a para- and non-parametric component. The thick vertical dashed-dot line indicates the turning point of the quadratic fit. The thin vertical dashed lines indicate the 90% Fieller confidence intervals for turning point. Bin means are the mean value of the outcome variable for the mean of the income aspiration gap for 50 quantiles of the aspiration gap. Sasabuchi p-values are for the null hypothesis that the income aspiration gap is monotone over the interval. The HM p-values are retrieved from a Hardle and Mammen (1993) test with 500 bootstrap replications and a null hypothesis that the non-parametric and quadratic fits are not different, i.e. the same. Controls included are age, gender, marital status, highest level of education of the household head, the size and socio-dependency of the household, the cropland, other assets owned and the locus of control. Controls are displayed in Table A5.

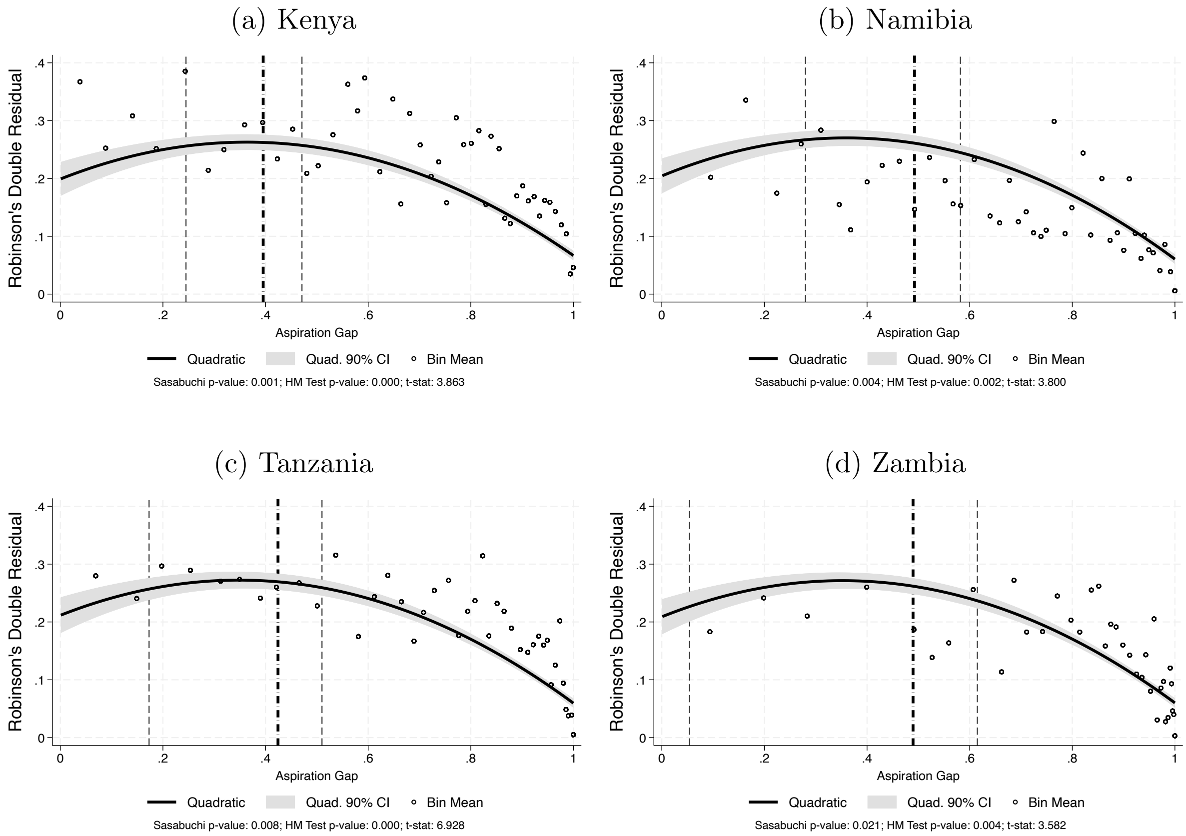


Figure A8: Aspiration Gap and Income Diversification, Simpson's Diversity Index

Notes: The Figures illustrate the quadratic and non-parametric regression fit of the income aspiration gap in a semi-parametric model where the dependent variable is Simpson's Diversity Index (SDI) for income diversification, SID $= 1 - \sum_{s=1}^{6} (\frac{y_s}{Y})^2$, where y_s is the share of income derived from source s and Y the total income from the past 12 months. On the vertical axis are the Robinson's double residuals (Robinson, 1988), which are estimated semi-parametrically using Equation 3. Robinson residuals partialled out the parametric fit from the regression and allow to test for a fit of a non-parametric term estimating a full model comprising of a para- and non-parametric component. The thick vertical dashed-dot line indicates the turning point of the quadratic fit. The thin vertical dashed lines indicate the 90% Fieller confidence intervals for turning point. Bin means are the mean value of the outcome variable for the mean of the income aspiration gap for 50 quantiles of the aspiration gap. Sasabuchi p-values are for the null hypothesis that the income aspiration gap is monotone over the interval. The HM p-values are retrieved from a Hardle and Mammen (1993) test with 500 bootstrap replications and a null hypothesis that the non-parametric and quadratic fits are not different, i.e. the same. Controls included are age, gender, marital status, highest level of education of the household head, the size and socio-dependency of the household, the cropland, other assets owned and the locus of control. Controls are displayed in Table A5.

A.7 Omitted variable bias

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10) (11)
	Uncontrolled			Controlled			Bias-Adjusted		Oster δ	
Outcome	β_1	β_2	R^2	β_1	β_2	R^2	β_1	β_2	δ_1	δ_2
Indicators of Livelihood Diversification										
Extensive margin	1.687	-1.878	0.656	1.676	-1.849	0.664	1.402	-1.134	1.09	1.03
Intensive margin	3.133	-3.571	0.161	3.141	-3.552	0.163	3.408	-2.954	0.93	1.02
Simpson Index	0.398	-0.483	0.644	0.390	-0.469	0.651	0.162	-0.064	1.05	1.01

Table A18: Oster (2019) test for Omitted Variable Bias

Notes: Columns (1)-(3) and (4)-(6) present results from Equation 2 that are also summarized in Tables 4, 5, and Table A7, with and without controls. Columns (4)-(6) are estimated using the full set of controls. In columns (7)-(10), we present indicators proposed by Oster (2019) to test for the presence of omitted variable bias. Columns (7) and (8) contain the bias-adjusted coefficients

Aspirations are not exogenous and therefore the relationship between aspirations and income diversification could also be driven by omitted variable bias.

It is common to include observable controls to test the sensitivity of the estimates. If the estimates are stable to the inclusion of the controls, it is assumed that omitted variable bias is limited. The quality of the control is diagnosed by how much of the variance of the outcome is explained with the inclusion of the controls or in other words, how much the R^2 moves.

We attempt to quantify potential omitted variable bias using Oster (2019)'s test, which is based on the assumption that robustness to omitted variable bias - relationship of outcome to unobservables - can be recovered from the relationship to observables (Altonji *et al.*, 2005). The idea of Oster (2019)'s test is to provide an bias-adjusted estimator that requires to consider the movement in R^2, the variance of the treatment and the outcome.

Key to the framework is a hypothetical R^2, R^2_{max}. So for cases where the outcome can be fully explained by the treatment and the set of control, $R^2_{max} = 1$. Oster (2019) provides two main assumptions under which a consistent bias-adjusted estimator can be calculated: being able to calculate (1) a relative degree of selection on observed and unobserved variables (δ, with $\delta = 1$ being an appropriate upper bound) and (2) a value for R^2_{max}. $R^2_{max} = 1$ may lead to over-adjustment in practice it is therefore preferred to calculate an R^2_{max} based on the data.

The method uses the degree of selection on observables to estimate the degree of selection of unobservables through movement in coefficient and R-squared.

Oster (2019) develops a test based on the theory that for robustness one often checks for movement of coefficients after the inclusion of additional controls. The author connects theory of omitted variable bias to coefficient stability and illustrates that it is necessary to look at coefficient and R-squared movements.