Beyond the formal economy: employment and income perspectives in Tanzania

Jehovaness Aikaeli*and Beatrice Kalinda Mkenda†

Abstract

This study investigates employment and earnings levels in Tanzania's informal and formal sectors, as well as the factors that determine them. Despite growth in the economy, insufficient job creation in the formal sector has rendered the informal sector an inevitable source of employment for youths. We use the Blinder-Oaxaca decomposition to examine the disparity between formal and informal earnings, and Quantile Regression to investigate the factors that determine earning differentials across income groups. Among key findings are: first, there are major differences in average earnings between the formal and informal sectors, with the informal sector having lower average earnings; second, women's earnings are lower in both sectors and across other sectors; and lastly, the formal sector has a higher percentage of employees who work a second job than the informal sector. Employee characteristics, gender, taking on another work, relocating to a rural location, and transferring to private employment and the informal sector are the factors that contribute to earnings disparities. The following policy implications are drawn: investing in education is crucial for raising earnings; women's participation in higher-paying activities is important for empowering them; and an overall improvement in employee earnings is important for reducing the need to get a second job to supplement their incomes.

***Key words**: Formal sector employment; Informal sector employment; Blinder-Oaxaca decomposition; Quantile Regression*

***JEL**: E24; E26; O17*

1. Introduction

Tanzania's Gross Domestic Product (GDP) increased significantly between 2005 and 2019, and yet this expansion did not translate into formal sector jobs. Between 2005 and 2019, Tanzania's GDP grew at an average rate of 6.5% (see Table A1 in Appendix A), before slowing down to 4.7% and 5.2% in 2022 and 2023, respectively. This growth rate is among the highest in sub-Saharan Africa (SSA) over the same period (IMF, 2016). The insufficient growth in formal sector jobs is a source of concern because the burgeoning youth population cannot find quality employment. According to the Population and Housing Census of 2022, Tanzania has a population of 61.7 million people, of

*University of Dar es Salaam School of Economics; jaikaeli@udsm.ac.tz; corresponding author
†University of Dar es Salaam School of Economics; bkmkenda@udsm.ac.tz; bkmkenda@gmail.com

 https://doi.org/10.56279/ter.v14i2.203

which about 50 percent are young people of 0-17 years, and the working population of youth (15 – 35 years) makes 34.5 percent of the country's population (URT[1], 2022). Because of the difficulty to absorb young people into formal sector jobs, the informal sector assumed on the role of employing millions of them, enabling them to earn a living. Looking beyond the formal sector to generate sufficient jobs and provide a means of livelihood is now being explored as a solution to the youth employment crisis. Appendix Table A1 shows that the overall unemployment rate increased from 2005 to 2011, and averaged about 5% between 2005 and 2019. Over the same period, the employment-to-population ratio fell.

According to studies conducted in African countries, the informal sector now employs more people than the formal sector (see, for example, Ellis et al. (2017), Yusuf (2014), Rukundo (2015), and Njaya (2015) for Tanzania, Nigeria, Rwanda, and Zimbabwe, respectively), and it provides a source of income for millions of youths. Such studies underline the importance of not neglecting the sector, but rather ensuring that policies are aimed at improving working conditions so that jobs are decent and give the necessary social protection. In the Sustainable Development Goals (SDGs), the eighth goal of providing decent jobs and economic growth requires sub-Saharan Africa countries to ensure that the informal sector, which employs millions of youths, operates in a decent environment and provides them with sustainable incomes.

In Tanzania, the fact that the informal sector now contributes to both jobs and incomes necessitates a systematic understanding of the following aspects: the extent to which it generates income adequate to sustain decent livelihoods for millions of people who rely on it; the extent to which it has grown; and its characteristics, such as gender distribution, location, and education levels of its participants. This paper examines the size of current informal sector[2] employment and income levels. This article is focused on whether informal sector employment is simply a safety net or if it could contribute significantly to income and welfare. We use labour force survey data and household budget surveys to answer three questions: *first, what characterises Tanzania's informal sector employment and earnings? Second, how do earnings from employment in the two sectors differ, and what factors contribute to the differences? Lastly, what policies could be implemented to ensure that the informal sector is a viable means of a decent livelihood for youths?*

[1] United Republic of Tanzania (URT) Population and Housing Census (2022) Key Findings.

[2] Informal employment refers primarily to employment in enterprises that lack registration and social security coverage for their employees (OECD, 2009). A distinctive feature of this type of employment is lack of social coverage and other related benefits applicable to formal employment, hence, it is highly precarious and vulnerable. The informal labor workforce is mostly unskilled and operate in low productivity jobs, in marginal, small scale and often family-based activities.

The paper is structured as follows: following a brief introduction, Section 2 provides the background by examining Tanzania's economy and job creation rate. It also emphasizes the situation of its youths and the need for job creation, as well as the urban-rural divide that forces them to relocate to urban areas in search for jobs in the urban informal sector. Section three analyses the literature on the informal sector, briefly discussing opposing perspectives on informality, the sector's relevance for generating income and employment, and what characterizes it. Section four outlines the approaches used to answer the research questions on earnings discrepancies between the formal and informal sectors, as well as the factors that explain the disparities. Section 5 concludes and draws some policy implications.

2. Setting the Tanzanian context

The extent of job creation and growth in the Tanzanian economy can be understood by examining its structural transformation over time, the level of youth unemployment, and the disparities in employment and livelihoods between rural and urban areas. This section looks into the situational analysis in view of the economy and jobs creation, youth employment question and the urban-rural setting of the country. The reconfiguration of the economy is observed as the sectoral transformation happens, with some sectors creating more job opportunities while those of the other sectors decreasing.

2.1 The Tanzanian economy and job creation

The Tanzanian economy has undergone structural change over time. Figure A1 in the appendix shows sectoral values as a percentage of GDP averaged over three time periods: 2005-2010, 2011-2015, and 2015-2020. It shows the following trends: a steady increase in the percentage of the value added to GDP in industry; the highest percentage of value added to GDP in services; a low percentage of value added to GDP in manufacturing; and a declining percentage of the value added to GDP in agriculture. The agriculture sector's diminishing share of the value added is a cause for concern, considering that it provides livelihoods for a large portion of Tanzania's population. If the sector were modernizing, this would not be a cause for concern because it would imply structural transformation for the betterment of the majority, which is common when countries develop and individuals leave agriculture to find work in the industrial sector. The services sector's high share of the value added to GDP is encouraging, and it should be regarded a viable sector for spearheading the expansion of jobs.

The Tanzanian economy's structural transformation mirrors the sectoral distribution of employment (Figure A2 in Appendix A), indicating that agriculture remains the largest employer, albeit its average percentage contribution to employment has decreased from 72% between 2005 and 2015 to 66% between 2016 and 2020. The percentage employed in services, the

second largest employer, increased from 23% between 2005 and 2015 to 28% between 2016 and 2020, while industry's share increased by slightly more than a percentage point from 6% between 2005 and 2015 to 7% between 2016 and 2020. These estimates conceal the underlying features of Tanzania's labour market that have led to the informal sector being the dominant source of income employment.

2.2 The youth question and jobs

The situation for youths in Tanzania is particularly concerning, and it is worth investigating how the informal sector may absorb them and become a viable and decent option for them. Table 1 shows that: first, 82.3% of Tanzanian youths are in vulnerable employment, with a larger percentage for those with lower levels of education and females; second, long-term unemployment among youths is a problem, since 42.1% of them have been unemployed for more than a year, with female youths having the greatest rates in Dar es Salaam and other urban areas.

Data on the number of people working in the informal sector are only available for 2006 and 2014, and it may be argued that its growth has coincided with a rise in youth unemployment. According to National Bureau of Statistics (NBS) data, the percentage of the labour force who engaged in the informal sector as their main activity rose by 8% in other urban areas between 2006 and 2014, but in Dar es Salaam, there was a 0.1% decrease. Between 2006 and 2014, the share of the rural labour force participating in informal activities as their main activity declined by about 8%. According to the data on gender distribution, more women are employed in the informal sector in Dar es Salaam and other areas than in rural areas (NBS, 2006 & 2014). By age groups, the statistics of 2006 show that the share of employed youth aged 25-34 years was 51.1%, which indicates high unemployment of youth in the country.

The third aspect is long-term unemployment, describes youths who have been unemployed for a long period and have become discouraged to search for employment. Table 1 indicates that youths are the most likely to be discouraged from seeking employment. The fourth aspect refers to the population that is neither employed nor enrolled in educational or training institutions (NEETs). These youths are inactive; according to Table 1, 16.6% are NEETs, with females and youths in Dar es Salaam having a higher percentage.

.

Table 1: The youth question in Tanzania, 2014

Youths in Vulnerable Employment by Level of Education & Gender				*Long-term Youth Unemployment by Area & Gender*				*Discouraged Job Seekers by Age Group & Gender*				*NEETs* by Area & Gender*			
	Male	**Female**	**Both**		**Male**	**Female**	**Both**		**Male**	**Female**	**Both**		**Male**	**Female**	**Both**
Never Attended	93.3	95.9	94.8	Dar es Salaam	63.0	67.7	66.4	15 - 24	15.6	33.3	48.9	Dar es Salaam	18.7	44.1	32.7
Primary	81.6	89.4	85.4	Other Urban	39.9	52.4	48.5	25 - 35	4.4	25.3	29.7	Other Urban	9.7	21.6	16.1
Secondary	73.0	83.0	77.6	Rural	15.3	23.4	19.5	35 - 64	2.8	16.2	19.0	Rural	10.7	15.3	13.0
Vocational Training	34.5	37.4	35.7	**Total**	**31.2**	**47.1**	**41.2**	65+	2.2	0.2	2.5	**Total**	**11.5**	**21.4**	**16.6**
Tertiary non-university	33.3	16.2	23.9					**Total**	**25.0**	**75.0**	**100.0**				
University	16.2	6.0	12.4												
Total	**78.2**	**86.4**	**82.3**												

***** *Neither in employment nor in education or training institutions.*

Source: *NBS, (2014), LFS Report.*

2.3 The urban-rural setting

The disparities between rural and urban areas in Tanzania are crucial to understanding the burgeoning informal sector, particularly in urban areas. First, rural areas have higher poverty rates than urban areas. The head count ratios derived from the past two household budget surveys indicate that poverty is mostly a rural phenomenon (see Table 2). Table 3 shows that, while both basic needs poverty and food poverty decreased nationally and in rural and urban areas between 2007 and 2011/12, they were larger in rural areas than in urban areas. The higher level of rural poverty implies that households relying on agriculture as their primary source of livelihood are among Tanzania's poorest, and poverty-reduction measures must focus on the agriculture sector, where the majority of them earn a living.

Table 2: Food and basic need head count rates and real per capita expenditure by area – Tanzania Mainland, 2007 and 2011/12

Indicator	Dar es Salaam	Other Urban Areas	Rural Areas	Tanzania Mainland
Basic Needs Poverty Line (Monthly per adult equivalent) (%)	14.1	22.7	39.4	34.4
2007	4.1	21.7	33.3	28.2
2011/12				
Food Poverty Line or Extreme Poverty (Monthly per adult equivalent) (%)				
2007	3.2	8.9	13.5	11.8
2011/12	1.0	8.7	11.3	9.7
Real per Capita Expenditure (Monthly TZS)				
2007				
Mean	74,904.51	52,800.30	93,502.41	59,997.16
Median	59,120.57	42,659.58	71,235.96	46,034.38
2011/12				
Mean	73,876.22	52,763.86	109,030.25	62,395.51
Median	57,559.85	45,089.14	85,959.56	49,346.80

Source: *NBS (2014)*

In terms of income levels, the second aspect is that, according to expenditure data, both the mean and median expenditure levels are lower in rural areas than in urban areas; Table 3 illustrates this difference; in 2011/12, Dar es Salaam's median per capita income was slightly more than twice that of rural areas, an improvement over 2007 when it was 60% higher in Dar es Salaam than in rural areas. It is not implausible to conjecture that this disparity in expenditure levels may be a contributing factor to youths migrating to urban areas in search of a better life, and given the scarce formal sector job opportunities, it means that the youths try their luck in the informal sector.

Third, there is a significant socioeconomic disparity between Tanzania's rural and urban areas. This disparity in socioeconomic conditions is another key factor that may drive youths to migrate and work in the informal sector in urban areas. Social service provision, for example, is lower in rural areas than in urban areas. The quantity and quality of schools in rural areas are lower, which may explain why a higher percentage of household members lack education in rural areas than in urban areas: the percentage of households without education was six times higher and three times higher in rural areas than in Dar es Salaam and other urban areas, respectively (LFS data).

The low quality of life in rural areas has been one of reasons for rural-urban migration as youth chase better job opportunities in the urban informal sector. Water and sanitation facilities are also scarce in rural areas: an

increasing proportion of the rural population practices open defecation, and while the percentage of rural people who have access to basic drinking water and sanitation services has increased over time, it remains significantly lower than in urban areas. Such differences between rural and urban areas encourage youths to pursue opportunities in urban areas, where services are better. There is also a disparity in access to electricity between rural and urban areas. Although the percentage of the rural population with access to electricity has steadily increased since 2010, it still falls short of the national average (by half) and four times that of the urban population. For youths, limited access to electricity deprives them of jobs that can be created in various kinds of non-farm enterprises, causing them to move to cities where electricity is more readily available in order to try their hand in the informal sector.

3. A brief review of literature

Policymakers view the informal sector's position from two perspectives. The first is that the informal sector is an important generator of income and growth and hence requires adequate support. The second associates the informal economy with economic disarray, making it unsuitable for support and necessitating reorganization to formalize it (Aikaeli and Mkenda, 2014). Both perspectives have been observed in Tanzania's policy regimes over the past two decades. The dividing eras include times when the government prioritized formalization of businesses and discouraging informal activity, particularly itinerant trading and other informal activities in urban areas. In recent years, a shift in perspective has occurred, fueled in part by political goals, in which the informal sector is embraced as an important part of the economy, with some due protection for people working in informal jobs and seemingly little emphasis on formalization. Clearly, policy consistency is essential to provide guidance on the role of informal activities in generating employment and income.

While there is recognition of the informal sector's importance for employment and income generation, more research is needed to provide policy guidance on which specific activities deserve more support in terms of potentially high employment and average earnings, what determines employment and earnings in both formal and informal sectors. Such research is missing in Tanzania. This study aims to fill that gap. According to Hussmanns (2001), the informal sector contributes significantly to job creation, income generation, and poverty reduction in many countries, particularly developing and transition countries. Statistics on the informal economy are required as an evidence-based tool for research and policymaking, as well as to enhance knowledge of the numerous workers and their economic contributions. This study goes beyond statistics and investigates what drives employment and earnings in informal activities.

In sub-Saharan Africa (SSA), the informal economy is estimated to account for 42% of GDP and 34% of Tanzania's national economy (Becker, 2004; Economic and Social Research Foundation (ESRF), 2011). According to the International Labour Organization (ILO) (2013), the informal economy accounts for 50% to 75% of all non-agricultural employment in developing countries. The growing importance of the informal economy has been attributed to the fact that informal sector activities constitute a primary source of survival for many people, particularly youths, and for some formal workers, informal sector activities are essential for supplementing their earnings.

Informal employment is characterized by a lack of wage protection, mandatory overtime or extra shifts, layoffs without notice or compensation, unsafe working conditions, and the absence of social benefits such as pensions, sick pay, leave, and health insurance (Aikaeli and Mkenda, 2014; Fields, 2011). Women, migrants, and other vulnerable groups of workers who are excluded from other employment opportunities have no choice than to choose informal, low-quality jobs. Given this, the informal economy has remained a useful concept for activists, policymakers, and researchers, as a major portion of employment and income occurs outside of the regulated formal sector (Chen, 2007).

4. Methodology and analysis

4.1 Analytical approach and variables

We use the Blinder-Oaxaca (Blinder 1973; Oaxaca 1973) decomposition for linear regression models to examine the disparity between formal and informal earnings. The Blinder–Oaxaca decomposition is a statistical method that breaks down the difference in means between two groups into two parts. It's often used to study a variety of differences such as wage gaps by race or sex, income gaps by groups, etc. This model is suitable for this study since we have two groups (formal and informal income earners), and income (Y) is hypothesised to be determined by individual characteristics or predictors (X). The question addressed is: *what is the mean variation (D) attributable to group differences in the predictors?* The methodology used to answer these questions is detailed in Appendix B. Using the appended estimation model, we estimate equation (3b) to understand the contribution of the endowments (individuals' characteristics); coefficients and the interaction (endowment and coefficients) to the differential in the earnings between formal employment and informal employment. Also, for the robustness of our results, we estimate the two components in equation (8b) to separate the explained and unexplained parts of the differential, and then examine their drivers. The estimation of equation (3b) is done as follows; let $\hat{\beta}_f$ and $\widehat{\beta}_{inf}$ be the least-squares estimates for the parameters of formal employment earnings, β_f and β_{inf}, which are obtained independently from the two group-specific samples.

Taking the group means ($\bar{X}_f$ and $\bar{X}_{inf}$) as the expected values of the predictors of the earnings from both groups, equation (4b) is written as;

$$\hat{D} = \bar{X}_f - \bar{X}_{inf} = (\bar{X}_f - \bar{X}_{inf})'\hat{\beta}_{inf} + \bar{X}'_{inf}(\hat{\beta}_f - \hat{\beta}_{inf}) + (\bar{X}_f - \bar{X}_{inf})'(\hat{\beta}_f - \hat{\beta}_{inf}) \quad (1)$$

If low earnings are in informal employment only for example, and assuming that there is no purposeful positive favour to the formal employment to earn more, we use $\hat{\beta}_f$ as an estimate of β^*, and write equation (5) as,

$$\hat{D} = (\bar{X}_f - \bar{X}_{inf})'\hat{\beta}_f + \bar{X}'_{inf}(\hat{\beta}_f - \hat{\beta}_{inf}) \ . \quad (2)$$

The variables employed are described in Table A3 in Appendix A, and they were selected based on our understanding of the factors affecting the incomes of formal and informal workers, as well as findings from the other similar studies (Dasgupta et al., 2015; Baskaya and Hulagu, 2011).

4.2 Data and estimations

We use the Household Budget Survey (HBS) dataset of 2012 to estimate B-O decomposition regressions and cross-sectional determinants of earnings. The reason for using the 2012 HBS dataset is to leverage the existence of the two concurrent sets of Integrated Labour Force Surveys of 2006 and 2014, which have consistently comparable key indicators, with the 2012 HBS lying between these sets that are considered as the baseline and end-line for this case. We did not pool the various household budget surveys due to differences in the methodology employed to collect data for each wave. We use the most recent set, which has improved over previous methodologies and meets the analytical needs of our study. After cleaning the data, we obtained a sample of 2500 workers from the HBS (no missing values), with 69% men and 31% women.

4.2.1 Income and employment by categories and sectors

Average monthly earnings are lower in informal employment, and females earn less than males in both formal and informal employment (Table A4 in Appendix A). Earnings disparities between men and women exist in the formal-unofficial categories, as well as across sectors. It is worth noting that women in informal rural areas earn more on average per month than their urban counterparts. This is not surprising given that women constitute a key source of labour in rural areas, whereas in urban areas, the majority of women work in unpaid jobs as housewives.

Despite a few extreme values, earnings from formal employment are mostly distributed around the mode, whereas earnings in the informal sector generally distributed below the mode. If one deviates from the central tendency for informal earnings, he or she is likely to earn lower earnings (Figure 1).

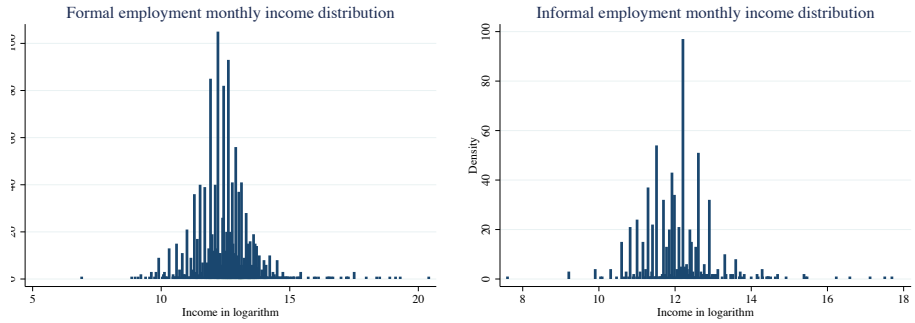


Figure 2: *Monthly average income distribution*

Formal workers in Tanzania perform extra activities to supplement their incomes because they have the resources to do so. According to the data, around 18% and 10% of employees in formal and informal employment held a second job, respectively. For those in formal employment, it is not possible to tell whether their second job is formal since no information is given in the data. Decent jobs, especially in the public sector, are the most rewarding, followed by those in the private sector, particularly in urban areas.

Tables A5 and A6 in Appendix A show the extent of earnings and engagement in the two sectors based on education. Table A5 shows that the higher an employee's education level is, the higher their average earnings. There is a substantial disparity in income between educational levels; workers with a secondary school education earn three times more than those with a primary school education, while those with a university education earn five times more than those with a secondary school education. Engagement in the two sectors varies by education level; Table A6 demonstrates that the proportion of informal work decreases as education level increases. This is because obtaining a higher degree increases one's chances of being recruited in decent formal jobs.

4.2.2 Earnings disparity between formal and informal employment

To analyse the earnings disparity between formal and informal employment, we apply the standard Blinder-Oaxaca three-stage decomposition. Table 3 shows that informal employment has lower mean earnings than formal employment, as indicated by the significant logarithmic coefficients of earnings for formal and informal estimates. The difference in earnings between formal and informal employment is significant at the 1% level, accounting for 0.48 of the log of earnings in these two categories. Individual characteristics contribute substantially to the earnings disparity (which is statistically significant at 5%), accounting for 35.7% of the total earnings gap. The rest of the gap is explained by the interaction of individual characteristics and coefficients, including the intercept, which is

significant at 1% and accounts for 48.7% of the overall earnings gap. The expected change in the mean outcome of informal earnings if informal earners had formal earners' coefficients, is insignificant.

Table 3: Blinder-Oaxaca threefold decomposition of earnings gap between formal and informal employment

Blinder-Oaxaca decomposition Number of obs. = 2,500
Model = linear
Group 1: formal N of obs. 1 = 1714
Group 2: informal N of obs. 2 = 786

log of income	Coef.	Std. Err	z	P>|z|	[95% Conf. Interval]	
Overall						
formal estimation ***	12.5475	0.0267	470.75	0.0000	12.4953	12.5998
informal estimation ***	12.0649	0.0335	359.70	0.0000	11.9991	12.1306
difference ***	0.4827	0.0428	11.27	0.0000	0.3987	0.5666
endowments **	0.1721	0.0798	2.16	0.0310	0.0158	0.3284
coefficients	0.0755	0.0501	1.51	0.1320	-0.0228	0.1737
interaction ***	0.2351	0.0844	2.78	0.0050	0.0696	0.4006

Note: *** denotes significance at 1%; ** significance at 5%; and * significance at 10%.
Source: Authors' estimations

Table 4 provides a detailed decomposition of the factors driving the three terms of the earnings gap: endowment, coefficients, and interaction terms. Education, as a proxy for knowledge and abilities, is important for explaining the earnings gap between formal and informal employment, and none of the other explanatory variables are significant. The more highly educated workers are employed, the wider the earnings gap between formal and informal employment. This is because a higher proportion of highly educated people are more likely to be employed in well-paid formal (and decent) jobs, as opposed to the majority of individuals with only a primary education, who engage in relatively low-paying informal jobs.

If informal income earners had the coefficients of formal income earners instead, demographic factors (*age and gender*) and education would worsen the earnings gap. Table 4 shows that age and gender are both positively and significantly related to the earnings gap: the more elderly persons and males work in the formal sector, the larger the earnings gap. The implication is that enhancing youth and female participation in formal employment would reduce the earnings gap. Furthermore, increasing workers' education levels would raise formal employment and incomes, thereby widening the earnings gap. This indicates that if the gap is to be narrowed, education must be universally improved. The findings also indicate that the interaction

between endowments and coefficients is influenced by age, which deepens the earnings gap. The premise is that as workers get older, they get more experience or endowment and hence earn more. The results also indicate that the interaction between endowments and coefficients is driven by age, which widens the earnings gap. The assumption is that as workers get older, they get more experience, or are more endowed, and hence are likely to earn more.

Table 4: Blinder-Oaxaca threefold detailed decomposition of earnings gap between formal and informal employment

Blinder-Oaxaca decomposition Number of obs. = 2,500
Model = linear
Group 1: formal N of obs. 1 = 1714
Group 2: informal N of obs. 2 = 786

log of income	Coef.	Std. Err	z	P>|z|	[95% Conf.	Interval]
endowments						
age	-0.0050	0.0134	-0.37	0.708	-0.0314	0.0213
sex	-0.0096	0.0122	-0.78	0.433	-0.0336	0.0144
rural-urban location	-0.0077	0.0048	-1.61	0.107	-0.0172	0.0017
education ***	0.1814	0.0431	4.20	0.000	0.0968	0.2660
multiple jobs	-0.0136	0.0090	-1.52	0.128	-0.0312	0.0039
private sector	-0.0159	0.0172	-0.92	0.357	-0.0497	0.0179
household sector	0.0426	0.0826	0.52	0.606	-0.1193	0.2046
coefficients						
age **	0.2685	0.1372	1.96	0.050	-0.0004	0.5374
sex ***	0.5918	0.1229	4.81	0.000	0.3509	0.8327
rural-urban location	0.0351	0.0276	1.27	0.204	-0.0190	0.0893
education **	0.1685	0.0811	2.08	0.038	0.0095	0.3274
multiple jobs	-0.0038	0.0130	-0.29	0.771	-0.0292	0.0217
private sector	-0.0233	0.0950	-0.25	0.806	-0.2095	0.1629
household sector	-0.1404	0.1213	-1.16	0.247	-0.3781	0.0972
Constant **	-0.8209	0.3246	-2.53	0.011	-1.4571	-0.1847
interaction						
age *	0.0325	0.0169	1.92	0.055	-0.0007	0.0657
sex	0.0066	0.0085	0.78	0.437	-0.0100	0.0232
rural-urban location	0.0048	0.0044	1.08	0.279	-0.0039	0.0135
education **	0.0953	0.0460	2.07	0.039	0.0050	0.1855
multiple jobs	-0.0030	0.0102	-0.29	0.771	-0.0229	0.0169
private sector	-0.0042	0.0174	-0.24	0.807	-0.0383	0.0298
household sector	0.1031	0.0891	1.16	0.247	-0.0716	0.2778

Note: *** denotes significance at 1%; ** significance at 5%; and * significance at 10%.

Source: Authors' estimations

Table 5 provides the results for the extent to which variables explain total earnings. The model's specified variables explain 75% of the total earnings gap, leaving 25% unexplained, and both explained and unexplained components are statistically significant at 1%. Individual variables have a significant effect on the explained part of the total earnings gap, with age and education being positive and significant at 5% and 1%, respectively. An increase in the participation of the elderly and highly educated workers

raises earnings, particularly in formal employment, and widens the explained gap. Participation in informal household employment is positive and substantial at 1%, implying that it contributes to the widening of the explained discrepancy. The differences contributing to reducing the explained earnings gap are the rate of participation in formal private sector employment, which is significant at 1%, and working multiple jobs (at least two), which is significant at 5%.

Table 5: Blinder-Oaxaca two-stage overall and detailed decompo sition of earnings gap between formal and informal employment

Blinder-Oaxaca decomposition Number of obs. = 2,500
Model = linear
Group 1: formal N of obs. 1 = 1714
Group 2: informal N of obs. 2 = 786

log of income	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf. Interval]	
Overall						
formal estimation ***	12.5475	0.0266	471.56	0.000	12.4954	12.5997
informal estimation ***	12.0649	0.0334	361.26	0.000	11.9994	12.1303
difference ***	0.4827	0.0427	11.30	0.000	0.3990	0.5664
explained ***	0.3614	0.0298	12.11	0.000	0.3029	0.4198
unexplained ***	0.1213	0.0466	2.60	0.009	0.0300	0.2126
Explained						
age **	0.0196	0.0084	2.34	0.019	0.0032	0.0360
sex	-0.0050	0.0064	-0.78	0.433	-0.0176	0.0075
rural- locationurban	-0.0045	0.0029	-1.57	0.117	-0.0102	0.0011
education ***	0.2709	0.0210	12.93	0.000	0.2298	0.3119
multiple jobs **	-0.0143	0.0057	-2.49	0.013	-0.0255	-0.0030
private sector ***	-0.0237	0.0079	-3.01	0.003	-0.0391	-0.0083
household sector ***	0.1184	0.0281	4.21	0.000	0.0633	0.1735
Unexplained						
age *	0.2764	0.1534	1.80	0.072	-0.0242	0.5770
sex ***	0.5938	0.1020	5.82	0.000	0.3938	0.7938
rural- locationurban	0.0367	0.0288	1.27	0.203	-0.0198	0.0932
education *	0.1742	0.0937	1.86	0.063	-0.0094	0.3579
multiple jobs	-0.0061	0.0191	-0.32	0.749	-0.0435	0.0313
private sector	-0.0197	0.1388	-0.14	0.887	-0.2917	0.2523
household sector	-0.1131	0.1608	-0.70	0.482	-0.4283	0.2022
constant *	-0.8209	0.4114	-2.00	0.046	-1.6273	-0.0146

Note: *** denotes significance at 1%; ** significance at 5%; and * significance at 10%.
Source: Authors' estimations

The unexplained part of the total earnings gap is difficult to interpret since it includes the impacts of sector discrimination and omitted variables. Nonetheless, because the majority of the total disparity is explained, we

assume the omitted variables effect is not substantial, implying the presence of significant discrimination for some predictors. There is significant positive discrimination in earnings toward formal employment, which is attributed to elderly workers' participation, male workers' dominance in formal employment, and the dividend of education in general. This positive discrimination effect on formal income earners widens the unexplained total income gap between formal and informal employment.

4.2.3 Determinants of earnings

Table 6 gives the results of estimating equation (14) using the same explanatory variables used in the previous B-O decomposition.

Table 6: Estimated determinants of earnings by quantiles

	Median/standard regression			*Significance of other quantiles*								
				25th Percentile regression			*75th Percentile regression*			*90th Percentile regression*		
log of income	Coef.	t	P>|t|	Coef.	t	P>|t|	Coef.	t	P>|t|	Coef.	t	P>|t|
age	0.0071	4.50	0.000***	0.0054	3.08	0.002***	0.0132	7.47	0.000***	0.0081	2.85	0.004***
sex	-0.3435	-8.10	0.000***	-0.3756	-7.89	0.000***	-0.3441	-7.22	0.000***	-0.3365	-4.37	0.000***
rural-urban location	-0.1032	-2.69	0.007***	-0.1979	-4.60	0.000***	-0.0665	-1.54	0.123	0.0606	0.87	0.384
education	0.3430	18.63	0.000***	0.3552	17.17	0.000***	0.3937	19.01	0.000***	0.4336	12.95	0.000***
multiple jobs	-0.1678	-3.38	0.001***	-0.2328	-4.17	0.000***	-0.1185	-2.12	0.034	-0.1083	-1.20	0.230
private sector	-0.3410	-7.60	0.000***	-0.3525	-7.00	0.000***	-0.2313	-4.59	0.000***	-0.2017	-2.47	0.013**
household sector	-0.4408	-8.25	0.000***	-0.4935	-8.23	0.000***	-0.1989	-3.31	0.001***	-0.1693	-1.74	0.081*
Constant	12.2442	112.86	0.000***	11.9711	98.23	0.000***	12.2320	100.28	0.000***	12.7118	64.47	0.000***

Note: *** denotes significance at 1%; ** significance at 5%; and * significance at 10%.

Source: Authors' estimations

In the standard quantile regression, all the specified earnings determinants are significant at the 1% level. The age of income earners and education have a positive sign, but the rest of the variables are negative. Earnings increase with age because an older worker is more likely to earn an experience premium that new employees or youths do not. The positive age premium accounts for approximately 0.7% of an increase in employment in the median category. This is consistent with the results for the 25th, 75th, and 90th percentiles, which show premiums of 0.5%, 1.3%, and 0.8%, respectively.

Increasing women's labour force participation by 1% reduces earnings by 29.1%, owing to the negative discrimination against women indicated in the preceding B-O decomposition. The same pattern of results is observed for the rest of quantiles estimated, with negative changes of 31.3%, 29.1%, and 28.6% in income earner groups within the 25th, 75th, and 90th percentiles, respectively. This finding coincides with Adolfo and Cruz (2014) and Dasgupta et al. (2015).

Although switching the location of 1% of income earners from urban to rural areas would reduce earnings by 9.8% for the median category and by 18% for the 25th percentile category. However, shifting the earners' location to rural

areas for the estimated higher quantiles (75th and 90th) would not cause significant change in earnings within these groups.

The education variable indicates a positive change in the respective quantiles as a result of progressing from lower to higher levels of education. The median regression findings reveal that a 1% increase in education raises earnings by around 41% for the median category. A similar pattern is observed in the other quantiles, with rises of 42.6%, 48.3%, and 54.3% for the 25th, 75th, and 90th percentiles, respectively.

Reallocating 1% of single-job workers to multiple jobs would reduce median earnings by 15.4%. This is most likely due to the fact that the majority of people who work at least two jobs earn a low salary, whereas the majority of high-income earners work only one job. This finding is especially evident because it is only significant at the median and lower quantiles.

Switching 1% of public sector employees to private employment would result in a considerable decrease in earnings. A 1% shift of workers to the private sector would result in a 28.9% fall in median earnings. Similar large reductions are observed for the other estimated quantiles: 29.7%, 20.7%, and 18.3% for the 25th, 75th, and 90th percentiles, respectively. The reasons for such an outcome could be several, but the private sector faces many challenges in terms of for example, skills, and business environment. Shifting workers from the public to the private sector, for example, would not result in the same or higher remuneration.

5. Conclusion and policy implications

Tanzania's challenge of inadequacy in the creation of jobs, even with growth in the economy, has made it inevitable for the informal sector to be important for absorbing the burgeoning youth population. Around 82.3% of Tanzanian youths were in vulnerable employment, with a larger proportion for those with lower levels of education and females. The long-term unemployment among youths has been a problem, as 42.1% of them have been unemployed for more than a year, with female youths having the greatest rates. The number of people who were working in the informal sector between 2006 and 2014 show that the economic growth coincided with a rise in youth unemployment. The share of the rural labour force participating in the informal activities as their main activity declined while rising in urban areas, owing to factors such as rural-urban migrations as the economy transformed.

Using the Blinder-Oxaca (B-0) decomposition and Quantile Regression on the HBS dataset, we found the following: a significant difference in average earnings between the formal and informal sectors (with the informal sector having lower average earnings); women get lower earnings in both sectors and across other sectors in the economy; and among women themselves,

earnings differ by location, with a significantly higher average income of women in rural areas compared to their urban counterparts. Other findings related to moonlighting and correlation between education and income, and education and involvement in informal sector activities are: although employees in both sectors have a second job, the percentage of employees in the formal sector with a second job is higher, providing evidence of moonlighting; a positive correlation between education and income earned, with a threefold increase in average income for people with secondary school education compared to those with primary school education, and a five-times increase in average income for those with university education compared to those with secondary school education; and a negative correlation between education and involvement in informal sector activities, meaning that as the level of education rises, the proportion of people engaged in informal employment falls.

The larger proportion of the highly educated people is more likely to be employed in well-paid formal and decent jobs compared to the majority of individuals with only a primary education, who engage in relatively low-paying informal jobs. This explains the paramount importance of the informal sector in the developing economy as a cushion for the lowly educated and unemployed youth. Education, age, and gender are specific factors that contribute to the earnings disparities. These factors account for three-quarters of the variation, leaving one-quarter unexplained. Using quantile regression, we found that age and education were key drivers of earnings, with both increasing earnings. Other significant determinants are gender, taking on another job, locating in a rural, and switching to private employment and the informal sector, all of which reduce earnings.

Some policy implications emerging from the findings are; first, given its positive contribution to earnings, human capital development through education deserves special attention. Educating youths is critical for improving their earning potential and reducing participation in informal sector activities that generate lower incomes than formal activities. Second, women's lower earnings across sectors indicates a need for ensuring that women are employed in activities that generate higher incomes, and are empowered through participating in economic activities, if gender equality is to be achieved. The higher involvement of women in the informal sector where average incomes are low contributes to their lower average incomes, especially in urban areas. Third, employees engaging in moonlighting activities, referred to as "double jobbing," reduces their attention and work input in their principal job. Studies have shown that moonlighting activities have a negative effect on performance and productivity in the principal sector (van der Gaag et al., 2009). This calls for improving the level of earnings for employees in the formal sector so that they focus on their work, rather than finding other ways to supplement their incomes.

References

Aikaeli, J. & B. K. Mkenda, (2014), "Determinants of Informal Employment: A Case of Tanzania's Construction Industry", Botswana Journal of Economics, *Botswana Journal of Economics*, Vol. 12 (2): 51-73.

Angel-Urdinola, D. F. & K. Tanabe, (2012), "Micro-Determinants of Informal Employment in The Middle East and North Africa Region", Social Protection Discussion Papers, The World Bank.

Bank of Tanzania, (2016), Bank of Tanzania Annual Report 2014/15.

Baskaya, Y. & T. Hulagu, (2011), "Informal-formal Worker Wage Gap in Turkey: Evidence from a Semiparametric Approach", *ERSA conference papers* ERSA11p409 (Louvain-la-Neuve, European Regional Science Association).

Baum, C. F. (2013), "Applied Econometrics Notes", Boston college – Spring 2013, *unpublished*.

Becker, K. F., (2004), "Informal Economy: Facts Finding Study", Swedish International Development Cooperation Agency (SIDA), Stockholm.

Blinder, A. S., (1973), "Wage discrimination: Reduced form and structural estimates", *Journal of Human Resources* 8: 436–455.

Chen, M. A., (2007), "Rethinking the Informal Economy: Linkages with the Formal Economy and the Formal Regulatory Environment", DESA Working Paper No. 46 ST/ESA/2007/DWP/46.

Dasgupta, S., R. Bhula-Or & T. Fakthong (2015), "Earnings differentials between formal and informal employment in Thailand", *ILO Asia-Pacific Working Paper Series*, November 2015.

Daymont, T. N., & P. J. Andrisani (1984), "Job preferences, college major, and thegender gap in earnings", *Journal of Human Resources* 19: 408–428.

Economic and Social Research Foundation (ESRF) (2011), Informal Sector Taxation, ESRF Publication, Dar es Salaam.

Fields, G.S., (2011), "Poverty and Low Earnings in the Developing World", Working Papers, 7-2011, Cornell University ILR School.

Ghose, A. (2003). *Jobs and Incomes in a Globalizing World*, International Labour Office, Geneva.

Horrace, W. C., & R. L. Oaxaca, (2001, "Inter-industry wage differentials and the gender wage gap: An identification problem", *Industrial and Labor Relations Review* 54: 611–618.

Hussmanns, R., (2001), "Informal sector and informal employment: elements of a conceptual framework", Paper presented at the Fifth Meeting of the Expert Group on Informal Sector Statistics (Delhi Group), New Delhi, 19-21, September.

ILO (2013), "Informal Economy",
http://www.ilo.org/global/topics/employmentpromotion/informal-economy/lang--en/index.htm; accessed: November, 2013.

IMF (2016), *Regional Economic Outlook: Sub-Saharan Africa – Time for a Policy Reset*, Washington DC, April.

Jones, F. L., & J. Kelley (1984), "Decomposing differences between groups: A cautionary note on measuring discrimination", Sociological Methods and Research 12: 323–343.

Jones, S. & F. Tarp, (2015)," Priorities for Boosting Employment in Sub-Saharan Africa: Evidence for Mozambique", *African Development Review*, Vol. 27, No. S1: p. 56–70.

Koenker, R, & Bilias, Y. (2001), "Quantile Regression for Duration Data: A Reappraisal of the Pennsylvania Reemployment Bonus Experiments." *Empirical Economics*. March, 26:1, pp. 199–220.

Koenker, R. & Bassett, G. Jr. (1978), "Regression quantiles", *Econometrica*, January, 46:1, pp. 33–50.

Lemieux, R., (2003), "The 'Mincer Equation' Theirty Years *after Schooling, Experience, and Earnings*", Center for Labor Economics, University of Berkeley Working Paper No.62.

National Bureau Of Statistics (NBS). 2013a. *Key Findings 2011/2012 HBS: Tanzania Mainland*. Dar es Salaam.

National Bureau of Statistics (NBS). 2013b. Population Distribution by Age and Sex. Dar es Salaam.

National Bureau of Statistics (NBS). 2014. *Tanzania Household Budget Survey Report.* Dar es Salaam.

National Bureau of Statistics (NBS). 2015. *Integrated Labour Force Survey 2014*: *Analytical Report*. Dar es Salaam.

Njaya, T., (2015), "Informal Sector, Panacea to the High Unemployment in Zimbabwe? Case of Informal Sector Enterprises of Harare Metropolitan", *International Journal of Research in Humanities and Social Studies*, Vol. 2, Issue 2, February, PP 97-106.

Oaxaca, R. L., & M. R. Ransom, (1999) "Identification in detailed wage decompositions", *Review of Economics and Statistics*, 81: 154–157.

OECD (2009), *Is informal Normal? Towards More and Better Jobs in Developing Countries*, OECD, Paris. Edited by J. P. Jütting and J. R. de Laiglesia

Oenker, R. & Hallock K. (2001), "Quantile Regression: An Introduction," *Journal of Economic Perspectives*, 15, 143–156.

Rukundo, J. B., (2015), "Understanding Informal Sector Employment in Rwanda", *International Review of Research in Emerging Markets and the Global Economy* (*IRREM*),

UNDP, (2017), http://www.undp.org/: accessed, March 2017.

Van Der Gaag, M. Stelcner, W. Vijverberg, (1989), "Public-Private Sector Wage Comparisons and Moonlighting in Developing Countries: Evidence from Côte d'Ivoire and Peru", LSMS Working Paper Number 52, The World Bank, Washington DC. Vol. 1 Issue 2, pp. 295-309.

Wamuthenya, R.W., (2010), "Determinants of Employment in the Formal and Informal Sectors of the Urban Areas of Kenya", AERC Research Paper 194.

WIDER Working Paper 2017/16 Employment and productivity growth in Tanzania's service sector Mia Ellis, Margaret McMillan, and Jed Silver.

Yusuf, S. A., (2014), "Informal Sector and Employment Generation in Nigeria", University of Lagos, Nigeria, MPRA Paper No. 55538: http://mpra.ub.uni-muenchen.de/55538/; accessed December 2017.

Appendices

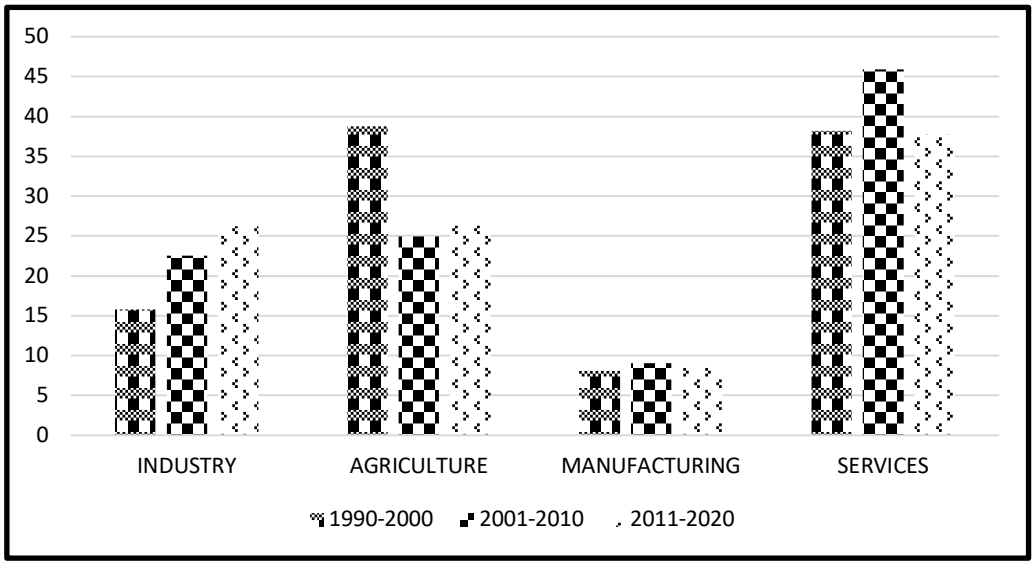
Appendix A: Tables and figures

Table A1: Tanzania's selected indicators on growth and employment, 2005-2022

	Growth Rate of GDP		**Unemployment, youth total (% of total labour force ages 15-24) (Modelled ILO estimate)**		**Employment to population ratio, 15+, total (%) (Modelled ILO estimate)**	
	Annual	Average (2005-2022)	Annual	Average (2005-2022)	Annual	Average (2005-2022)
2005	7.5		5.9		85.2	
2006	6.5		6.1		85.4	
2007	6.8		5.7		85.2	
2008	5.7		5.3		84.9	
2009	5.3		4.9		84.7	
2010	6.3		6.0		83.7	
2011	7.7		6.9		82.8	
2012	4.5		6.4		82.4	
2013	6.8		5.8		82.1	
2014	6.7		3.8		82.2	
2015	6.2		3.6		81.7	
2016	6.9		3.6		81.2	
2017	6.7		3.5		80.7	
2018	5.5		3.5		80.2	
2019	5.8		3.5		79.6	
2020	2.0		3.8		78.2	
2021	4.3		3.8		77.9	
2022	4.6	5.9	3.5	4.8	79.0	82.1

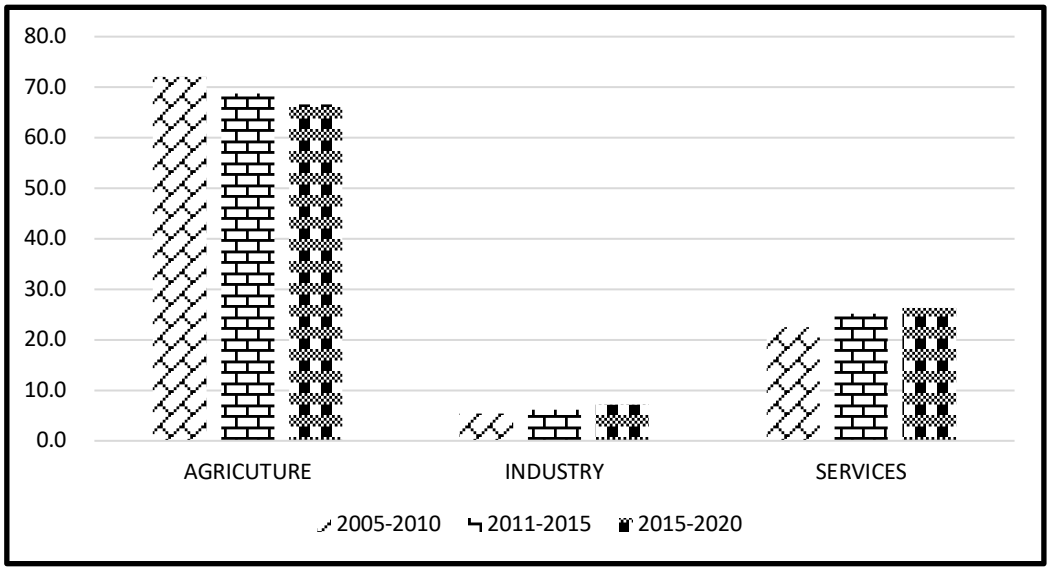
Source: *World Bank (2020), World Development Indicators online; http://data.worldbank.org; accessed, March 2024.*

Figure A1: Sectoral value added (% GDP), 1990-2020



Source: *World Bank (2020), World Development Indicators online; http://data.worldbank.org; accessed, March 2024.*

Figure A2: Sectoral distribution of employment



Source: *World Bank (2020), World Development Indicators online; http://data.worldbank.org; accessed, March 2024.*

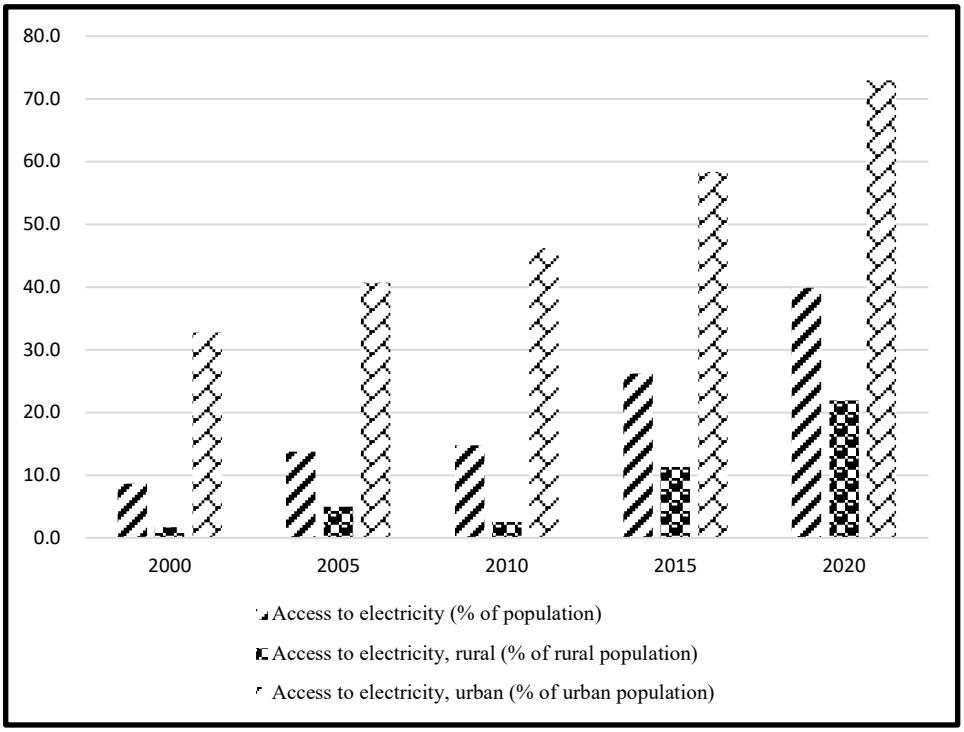
Table A2: Percentage distribution of household members with no education

Level of Education	Dar es Salaam		Other Urban Areas		Rural Areas		Tanzania Mainland	
	2007	2011/12	2007	2011/12	2007	2011/12	2007	2011/12
No education	7.9	4.4	12.1	8.9	28.5	24.2	23.6	18.6
Primary 1-4	5.2	3.4	7.9	6.5	12.3	10.8	10.9	9.0
Primary 5-8	57.0	48.6	58.9	50.6	52.4	51.8	54.0	51.2
Form 1-4	16.6	27.2	13.7	25.3	4.1	10.6	7.0	15.7
Other*	13.3	16.4	7.4	8.7	2.7	2.6	4.5	5.5
Total	100	100	100	100	100	100	100	100

Note: **Other includes: Pre-School; adult education; Form 5-6; Diploma/university; Courses after primary, Form IV and Form VI; and other certificates.*

Source: *NBS (2014)*

Figure A3: Access to electricity



Source: *World Bank (2020), World Development Indicators online; http://data.worldbank.org; accessed, March 2024.*

Table A3: Variable selection

Variable	***Description***
Monthly income	Log of total monthly earnings (total of both monetary wage and estimated pecuniary value of payment in kind)
Formal & informal employment	Informal employment is defined as any job without formal contractual agreement while formal employment is any job with a contractual agreement
Gender	Female = 2, male = 1
Age	Number of years of those who are 15 and above
Location	Rural = 1, urban = 0
Education (categorical)	
Primary school	Yes = 1
Ordinary (O) - level secondary school	Yes = 2
Advanced (A) - level secondary school	Yes = 3
University undergraduate	Yes = 4
University postgraduate	Yes = 5
Non-public employment	
Private sector	Yes = 1, no = 0 (private firms, non-governmental organizations and self-employment (non-farm)
Household sector	Yes = 1, no = 0 (unpaid helpers or household-farm labours as household sector workers)
Multiple jobs	With at least a second job = 1, with single job = 0

Source: Authors' selection

Table A4: Monthly average income by gender and categories/sectors (in 2012)

Category/sector	*Sex*	*Average income (Tsh.)*
Formal	Male & female	1,533,181
	Male	1,787,424
	Female	1,382,044
Informal	Male & female	442,997
	Male	512,205
	Female	126,408
Public sector	Male & female	3,004,576
	Male	3,752,204
	Female	658,571
Private sector	Male & female	686,366
	Male	777,327
	Female	233,965
Household sector	Male & female	490,616
	Male	555,793
	Female	106,944
Urban locations	Male & female	1,264,971
	Male	1,495,296
	Female	298,035
Rural locations	Male & female	1,037,980
	Male	1,153,857
	Female	511,053

Source: Authors' computation based on HBS, 2012

Note: 1. In this study, public sector contains government (local & central) and public corporations; private sector includes private firms, non-governmental organizations and self-employed non-farm organizations; while households sector comprises household farm and household unpaid categories.

3. Annual average exchanges rate in 2012 was Tsh. 1,571.7 per 1 USD. This rate can be used to convert figures from Tanzania shilling to US dollar (Bank of Tanzania, 2016).

Table A5: Average monthly income by education

Education level	*Average monthly income (Tsh.)*
Primary	350,371
Secondary	1,210,368
University	5,879,572

Source: Authors' compilation from HBS, 2012

Table A6: Engagement in formal and informal employment by education

Education level	*Category*	*Number*	*Percent*
Primary	*Informal*	*605*	*46*
	Formal	699	54
Secondary ordinary	*Informal*	*158*	*22*
	Formal	572	78
Secondary advanced	*Informal*	*18*	*7*
	Formal	252	93
University undergraduate	*Informal*	*4*	*3*
	Formal	149	97
University postgraduate	*Informal*	*1*	*2*
	Formal	41	98

Source: Authors' computation based on HBS, 2012

Appendix B: Detailing the methodology

B1: Explained and unexplained decomposition

The Blinder–Oaxaca (B-O) decomposition of the group difference in the regressors is written as;

$$D = E(Y_f) - E(Y_{inf}). \quad (1b)$$

where
D is the average difference in monthly earnings,
$E(Y_f)$ is the expected value of formal sector monthly earnings, and
$E(Y_{inf})$ is the expected value of informal sector monthly earnings.

The expected difference in outcomes of the two sectors is the difference in the linear prediction of the group-specific means of the explanatory variables (predictors), written as;

$$D = E(Y_f) - E(Y_{inf}) = E(\boldsymbol{X}_f)'\beta_f - E(\boldsymbol{X}_{inf})'\beta_{inf}. \quad (2b)$$

where, with a linear relationship of earnings and predictors of labour income,
$Y_l = \boldsymbol{X}_l^i\beta_l + \varepsilon_l,\ \ l \in (f, inf)$
$\boldsymbol{X}$ is a vector of the predictors of earnings and its constant;
β_l denotes slope parameters of the respective predictors and an intercept; and
ε_l is the error term.

By assumption, $E(Y_l) = E(\boldsymbol{X}_l^i\beta_l + \varepsilon_l) = E(\boldsymbol{X}_l^i\beta_l) + E(\varepsilon_l) = E(\boldsymbol{X}_l^i\beta_l)$ and $E(\beta_l) = \beta_l$, while $E(\varepsilon_l) = 0$. To identify the contribution of group differences in predictors to the overall outcome difference, equation (2b) can be rearranged as (see Jones & Kelley (1984), and Daymont & Andrisani (1984));

$$D = \{E(\boldsymbol{X}_f) - E(\boldsymbol{X}_{inf})\}'\beta_{inf} + E(\boldsymbol{X}_{inf})'(\beta_f - \beta_{inf}) + \{E(\boldsymbol{X}_f) - E(\boldsymbol{X}_{inf})\}'(\beta_f - \beta_{inf}) \quad (3b)$$

Equation (3) is referred to as the *threefold* decomposition of group differences, which is divided into three components;

$$D = E + C + I\ . \quad (4b)$$

The first component in equation (4) is part of the income differential that is due to group differences in the predictors (*characteristics effect*) and is given as;

$$\mathrm{E} = \{E(\boldsymbol{X}_f) - E(\boldsymbol{X}_{inf})\}'\beta_{inf}. \quad (5b)$$

The second component measures the contribution of differences in the coefficients (including variation in the intercept) to the differential, and it is given as;

$$C = E(\boldsymbol{X}_{inf})'(\beta_f - \beta_{inf}), \quad (6b)$$

The third component is an interaction term that takes care of the effect of differences in characteristics and coefficients that occur simultaneously between the two sectors, written as;

$$I = \{E(\boldsymbol{X}_f) - E(\boldsymbol{X}_{inf})\}'(\beta_f - \beta_{inf}). \quad (7b)$$

The decomposition in equation (3) is expressed from the viewpoint of informal employment earnings (*inf*). That is, the group variations in the explanatory variables (*predictors*) are weighted by the coefficients of informal employment earnings to determine the endowments. The E component measures the expected change in informal employment earnings' mean outcome if informal group had the formal group's predictor levels.

Similarly, the C component that takes care of differences in coefficients is weighted by informal earnings' predictor levels, which is the component that measures the expected change in informal earnings' mean outcome if informal employment group had formal employment group's coefficients. Intuitively, the differential could as well be explained from the viewpoint of formal earnings; that is, by looking at the opposite expression of this threefold decomposition.

The decomposition can be thought of as a vector of non-discriminatory coefficients that can be applied to determine the contribution of the differences in the predictors. If β^* is defined as the vector of non-discriminatory coefficients, the difference in the outcome can be expressed as,

$$D = \{E(\boldsymbol{X}_f) - E(\boldsymbol{X}_{inf})\}'\beta^* + \{E(\boldsymbol{X}_f)'(\beta_f - \beta^*) + E(\boldsymbol{X}_{inf})'(\beta^* - \beta_{inf})\} \quad (8b)$$

Equation (9b) is a twofold decomposition, with two terms; *EX* and *UEX*, that is, $D = EX + UEX$.

The first component is part of the outcome differential that is explained by group differences in the characteristics (the predictors), given by;

$$EX = \{E(\boldsymbol{X}_f) - E(\boldsymbol{X}_{inf})\}'\beta^* \quad (9b)$$

The second component is the unexplained part of the outcome differential that is attributed to discrimination to address the potential effects of differences in unobserved variables, given by;

$$UEX = E(\boldsymbol{X}_f)'(\beta_f - \beta^*) + E(\boldsymbol{X}_{inf})'(\beta^* - \beta_{inf}) \quad (10b)$$

The unexplained part of equation (9b) can be decomposed further into two components of parameters of respective employment groups; $\beta_f = \beta^* + \tau_f$ and $\beta_{inf} = \beta^* + \tau_{inf}$, where τ_f and τ_{inf} are vectors of group-specific discrimination parameters, which can be positive or negative discrimination as indicated by the sign.

The *UEX* component is now written as, $UEX = E(\boldsymbol{X}_f)'^{\tau_f} - E(\boldsymbol{X}_{inf})'\tau_{inf}$, which is made up of two components:

(i) $UEX_f = E(\boldsymbol{X}_f)'\tau_f$, the part that measures discrimination in favour of formal sector; and

(ii) $UEX_{inf} = -E(\boldsymbol{X}_{inf})'\tau_{inf}$, which measures discrimination against informal sector[3].

B2: Detailed decomposition

For a detailed understanding of the contributions of each explanatory/predictor variable, we use the B-O approach for identifying the contribution of the individual predictors to the explained part of the differential. Since the total differential is the total sum of unexplained and explained parts, we can separate the two, and then sum them up.

From equation (7), the explained part of the twofold-decomposed differential is:

$$\widehat{EX} = (\bar{X}_f - \bar{X}_{inf})'\hat{\beta}_f = (\bar{X}_{1f} - \bar{X}_{1inf})'\hat{\beta}_{1f} + (\bar{X}_{2f} - \bar{X}_{2inf})'\hat{\beta}_{2f} + \cdots \quad (11b)$$

Note:

X_1, X_2, … are the means of the single predictors, and

[3] If *UEX* is positive, it indicates that the formal sector is being discriminated against, whereas if it is negative, it indicates that the informal sector is being favoured.

$\hat{\beta}_1$, $\hat{\beta}_2$, … are their corresponding parameters.

The estimation of the errors of the individual predictors is done from equation (7); the unexplained part of the decomposed differential of the twofold equation is given as;

$$\widehat{UEX} = \left(\bar{X}_{inf}\right)'\left(\hat{\beta}_f - \hat{\beta}_{inf}\right) = \bar{X}'_{1inf}\left(\hat{\beta}_{1f} - \hat{\beta}_{1inf}\right) + \bar{X}'_{2inf}\left(\hat{\beta}_{2f} - \hat{\beta}_{2inf}\right) + \qquad (12b)$$

Earnings Determinants Across Income Groups

After establishment of contributors to the total income differential between formal and informal earnings, a question that remains is what are the independent variables that determine earnings, and more especially for the different income groups. The predictors attributing income differential are the ones that may determine earnings but with different implications for the respective income groups, especially the individual characteristics of employees. Because groups of income earners are heterogeneous, it is important that we run quantile regression (QR) model to find out the contributions of the respective predictors to each quantile of the earnings (Koenker and Bassett, 1978; Koenker and Bilias, 2001). QR coefficient shows a change in the respective quantile attributed to a unit change in independent variable. QR is conditional on median of the distribution. As a matter of definition, a quantile $q \in (0, 1)$ is the *y* that splits the dataset into proportions *y* below and *1-q* above. This means, $F(y_q) = q$ and $y_q = F^{-1}(q)$.

We can use QR to model conditional quantile of joint distribution of dependent variable (*y*) and the respective predictors ($\boldsymbol{x}$). Let the predictor function be denoted $\hat{y}(x)$ and its error as $e(x) = y - \hat{y}(x)$ to write a loss function owing to predictor errors as,

$$L\big(e(x)\big) = L(y - \hat{y}(x)) . \qquad (13)$$

If $L(e) = e^2$, the loss is the same like squared errors and so OLS is an optimal predictor, but if $L(e) = |e|$, the optimal predictor is the conditional median (med$(y|x)$. This means in QR, the optimal predictor is $\hat{\beta}$ that minimises $\sum_i |y_i - x'_i|$. In view of this, the objective function that is minimised by the quantile regression estimator for quantile q is,

$$Q(\beta_q) = \sum_{i:y_i \geq x'_i\beta}^{N} q\left|y_i - x'_i\beta_q\right| + \sum_{i:y_i < x'_i\beta}^{N} (1-q)\left|y_i - x'_i\beta_q\right| . \qquad (14)$$

The coefficient of q^{th} quantile is β_q. Quantile regression, equation (10) is estimated for earnings quantiles to analyse determinants of earnings for the different income categories (by percentiles) in Tanzania.