



# Informal Work in Sub-Saharan Africa: Dead End or Stepping Stone?

**DOI:**  
[10.2478/izajodm-2021-0015](https://doi.org/10.2478/izajodm-2021-0015)

**Document Version**  
Final published version

[Link to publication record in Manchester Research Explorer](#)

**Citation for published version (APA):**  
danquah, M., Schotte, S., & Sen, K. (2021). Informal Work in Sub-Saharan Africa: Dead End or Stepping Stone? *IZA Journal of Development and Migration*, 12(15), [0015]. <https://doi.org/10.2478/izajodm-2021-0015>

**Published in:**  
IZA Journal of Development and Migration

**Citing this paper**  
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

**General rights**  
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

**Takedown policy**  
If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact [uml.scholarlycommunications@manchester.ac.uk](mailto:uml.scholarlycommunications@manchester.ac.uk) providing relevant details, so we can investigate your claim.



Michael Danquah, Simone Schotte\* and Kunal Sen

# Informal work in sub-Saharan Africa: Dead end or stepping-stone?

## Abstract

Despite rapid economic growth in recent decades, informality remains a persistent phenomenon in the labor markets of many low- and middle-income countries. A key issue in this regard concerns the extent to which informality itself is a persistent state. Using panel data from Ghana, South Africa, Tanzania, and Uganda, this paper presents one of the very few analyses providing evidence on this question in the context of sub-Saharan Africa. Our results reveal an important extent of heterogeneity in the transition patterns observed for workers in upper-tier versus lower-tier informality. Given the limited alternative job opportunities available, particularly to those in lower-tier informal self-employment who often remain locked in a situation of inferior pay and conditions, specific policies that seek to enhance the livelihoods of workers in this most disadvantaged segment may be more relevant in the sub-Saharan context than policies that aim to reduce the regulatory barriers to formalization.

**Current version:** May 27, 2021

**Keywords:** informality, segmentation, labor market dynamics, sub-Saharan Africa

**JEL codes:** D31, J46, J62, O12

**Corresponding author:** Simone Schotte  
schotte@wider.unu.edu

---

United Nations University, World Institute for Development Economics Research (UNU-WIDER), Katajanokanlaituri 6 B, FI-00160 Helsinki, Finland

## 1 Introduction

In classical accounts of economic development, economic growth is seen to be accompanied by a decline in informal employment. Yet, in many low- and middle-income countries, informal forms of economic activity remain a persistent phenomenon despite rapid economic growth in recent decades (Kanbur, 2017). Given these trends, the informal economy now comprises >60% of total global employment and >90% of all micro and small enterprises worldwide—with the highest shares being observed in sub-Saharan Africa and South Asia (ILO, 2018). With premature deindustrialization and the growth of the informal service sector across these regions, it seems likely that the trajectory toward informalization may further be intensified in the future (Rodrik, 2016).

Informal employment has primarily been portrayed as providing a source of livelihood for the poor (La Porta and Shleifer, 2014). At the same time, there is an increasing consensus in recent theoretical and empirical studies that recognizes the extent of heterogeneity in informal work (see, Perry et al., 2007; Echevin and Murtin, 2009; Chen, 2012; Grimm et al., 2012; De Vreyer and Roubaud, 2013; Lehmann, 2015; Radchenko, 2017; Basu et al., 2018). For example, within informal wage employment, one may observe workers employed as casual laborers in poorly paid unskilled jobs, at the lower end, along with skilled workers employed in better-paid jobs that are not covered by labor legislation or social protection provisions but nevertheless require some professional training to obtain these jobs. Similarly, within informal self-employment, one may observe subsistence own-account or household entrepreneurs (often referred to as “penniless entrepreneurs”, see Banerjee and Duflo (2007) or the “reluctant self-employed”, see Basu et al. (2018)) along with larger and more productive non-household enterprises employing hired labor (sometimes referred to as “constrained gazelles” and “top performers” see Grimm et al. (2012)). Informal workers thus range from multi-dimensionally deprived individuals in subsistence activities which exhibit low returns, are easily accessible, and are undesirable relative to formal sector employment to workers in activities that are better paid, exhibit barriers to entry, and may even be preferred to formal sector employment (Hayami et al., 2006; Fields, 2019). This internal duality between a “lower-tier” and an “upper-tier” in informality can be observed both in wage employment and in self-employment.

A key issue concerning the persistence of informality in the labor markets of many low- and middle-income countries is whether informality itself—especially in lower-tier work—is a persistent state, such that the most disadvantaged workers are locked in a situation of inferior pay and conditions or whether informality is a transient state that all workers are roughly equally likely to experience at some point throughout their working life. Closely related to this is the question of whether informal employment provides a “stepping-stone” toward formal positions or, on the contrary, presents a “dead end” without better job perspectives, with the result that informal workers either stay in this position or drop out of the labor force (Slonimczyk and Gimpelson, 2015; Lehmann and Pignatti, 2018). In this paper, we examine the likelihood of workers moving from lower-tier to upper-tier informal work and to formal work (and vice versa), as well as the earning implications of such transitions using comparable panel data for four countries in sub-Saharan Africa—Ghana, South Africa, Tanzania, and Uganda.

A growing literature has examined the extent of heterogeneity in informal work and patterns of mobility within and across informal and formal employment in the Latin American

context (Perry et al., 2007) and, more recently, a small set of transition countries (Lehmann et al., 2012; Commander et al., 2013; Lehmann, 2015; Slonimczyk and Gimpelson, 2015; Lehmann and Pignatti, 2018). For instance, Lehmann and Pignatti (2018) examine whether the Ukrainian labor market is segmented along the formal–informal divide. Their analysis of the short-term and medium-term transitions between five employment states, unemployment, and inactivity shows that segmentation is present for dependent employees. For the majority of informal dependent employees, informal employment is used as a waiting phase to enter formal salaried employment and is not voluntarily chosen. The evidence by Lehmann and Pignatti (2018) for self-employment is, however, mixed with respect to segmentation in the Ukrainian labor market.

We contribute to this literature by examining the nature, magnitude, and direction of employment transition patterns in the informal economy in sub-Saharan Africa, where our knowledge of such transitions is limited. A strength of our analysis is the comparative nature of our study, which allows us to assess whether the patterns of transitions that we observe are specific to one country's context or holds true for other countries in our sample. The location of the countries in our study—in Western Africa (Ghana), Eastern Africa (Tanzania and Uganda), and Southern Africa (South Africa)—provides a basis for making generalizable claims on the patterns of mobility within and across the informal economy in sub-Saharan Africa. The four are among the few sub-Saharan African countries for which at least two waves of household panel data are available and where variable definitions can be harmonized across countries.

Specifically, in this paper, we offer a comparative perspective on the composition of employment and document transition patterns across different formality statuses, distinguishing between wage and self-employment. Importantly, we not only distinguish between formal and informal employment, understood as the set of economic activities that are not regulated or protected by the state, but offer a more nuanced view that additionally differentiates between “upper-tier” and “lower-tier” informality. We then examine the individual- and household-level characteristics that can be associated with these transition patterns and analyze the link between employment and income dynamics. The differentiation between formal, upper-tier informal, and lower-tier informal employment proves consequential in terms of both employment dynamics and earnings differentials, with earnings being significantly higher in upper-tier informal activities compared to the lower-tier and highest in formal work. Our results show high persistence in the lower-tier segment of informality, where self-employed workers, in particular, tend to remain locked in a situation of inferior pay and conditions. Informal wage jobs, by contrast, can present a steppingstone into formal employment relationships, especially for those in the more dynamic upper-tier segment. Last, we find a relatively strong segmentation between wage and self-employment in the sub-Saharan African case, with few workers exiting formal or upper-informal wage employment for self-employment.

The paper is organized as follows. Section 2 discusses alternate approaches to the issue of heterogeneity in informal employment and proposes the approach we adopt in our paper. Section 3 introduces the data and discusses the empirical methods used to analyze employment and income dynamics. Section 4 investigates the patterns of employment mobility using transition matrices. Section 5 presents the results from the econometric analysis. Section 6 concludes.

## 2 Heterogeneity in Work Statuses

The early literature on modeling labor markets in developing countries characterized the dualism inherent in these labor markets in terms of two formality statuses— formal employment, which offers relatively attractive wages and other terms and conditions of employment, and informal employment, which offers relatively unattractive pay and conditions of employment (Harris and Todaro, 1970). More recent literature has pointed out the multi-layered nature of labor markets in developing countries, arguing that two-sector models do not seem to be consistent with the empirical realities of labor markets in developing countries (Fields, 2005, 2019). In particular, there are two distinct characteristics of work status that need to be captured in an empirically grounded model of the labor market in developing countries. First, workers can be either in wage employment or in self-employment, which can exist in both formal and informal employment.<sup>1</sup> Second, informal employment is characterized by its own internal duality, where both wage employed and self-employed workers can be in upper-tier or lower-tier informal employment (Fields, 1990). This implies that any particular individual who is employed can be in one of six possible work statuses at a given point of time: (i) formal self-employed, (ii) formal wage employed, (iii) upper-tier informal self-employed, (iv) upper-tier informal wage employed, (v) lower-tier informal self-employed, and (vi) lower-tier informal wage employed. In this section, we discuss alternate approaches to classifying workers in different work statuses and propose our preferred approach operationalizing the six work statuses for the four countries in our study.

We first need to make a distinction between informal and formal employment. Here, the 17th International Conference of Labor Statisticians at the ILO has provided a consistent definition of informal employment which has been widely adopted in the literature, which we follow in our study (see ILO, 2018). According to this definition, informal employment is understood as work that lacks any type of legal recognition or protection, and where workers do not have secure employment contracts, workers' benefits, social protection, or workers' representation. This implies that within self-employment, formal self-employed are those enterprises that are registered with national state authorities (for example, with social security, sales, or income tax authorities) or contributing to social security, while informal self-employed are those enterprises that are unregistered. Within wage employment, formal wage employed are workers who contribute to social security and informal wage workers are those who do not contribute to social security.

On the distinction between upper-tier and lower-tier informal work, three different approaches have been commonly discussed in the literature. The first approach is to take upper-tier informal employment as being “voluntary” in nature, where workers choose to be in jobs that offer more independence and better earnings and working conditions as compared to working in the formal sector (the so-called “exit” view of informal work, see Maloney, 1999, 2004). In contrast, lower-tier informal work is “involuntary” and employment of last resort, when individuals cannot find employment in formal or upper-tier informal work (the “exclusion” view of informality, see Fields, 2005, 2019).

---

<sup>1</sup> The informal self-employed include employers, own-account workers, and contributing family workers (Gindling and Newhouse 2014). Own-account workers are self-employed individuals who do not employ others. Contributing family workers are those who hold self-employment jobs as own-account workers in a market-oriented establishment operated by a related person living in the same household.

To operationalize this approach, it is necessary to ascertain directly from labor force surveys whether a worker is voluntarily or involuntarily engaged in informal work. An example of such an approach is provided by Maloney (1999), where workers who moved to informal jobs from formal jobs in Mexico were asked in the survey whether the reason for the move was due to a desire for greater independence, or for higher pay, or was the move involuntary. Such direct questions to workers on their motives for changing jobs are extremely rare in standard labor force surveys in developing countries, especially in the Sub-Saharan African context. The lack of such information on worker motives for changing jobs in the data we have for the four countries in our study (we describe our data in greater detail in the next section) does not allow us to use an approach that can infer upper-tier and lower-tier work status from workers' self-reported reasons.

A second approach for classifying workers in upper-tier or lower-tier informality is to use outcome-based performance measures—such as earnings, business profits, or enterprise productivity—as the sorting criteria (see, for example, Grimm et al., 2012, who use a size and productivity criterion to classify informal entrepreneurs as “upper-tier” and “lower-tier”). As outcomes in the upper-tier are by definition superior to those in the lower-tier and sometimes comparable to the outcomes of formal units, this has been interpreted as evidence for the voluntariness of upper-tier informal employment.<sup>2</sup> However, this approach has two limitations. First, income or earnings would not capture other characteristics and benefits that may be associated with a job (such as the intrinsic value in terms of autonomy and independence that workers may attach to being self-employed in upper-tier informal work as compared to being a wage employee in a formal firm) (Maloney, 2004). Second, if a key objective of the analysis of worker transition is to understand whether transitions from lower-tier to upper-tier informal work are welfare-enhancing, using an income or earnings measure as the sorting criteria does not allow us to separate out the factors that may explain movements in work status from its consequences in terms of income gains or losses.

A third approach takes upper-tier informal work as “restricted entry” employment and lower-tier informal work as “free entry” (Fields, 1990). In the former case, there are barriers to entry to the job, which could be a certain level of capital if the worker is self-employed or some necessary professional training required for the job if the worker is wage employed. In the latter case, by definition, “free entry” employment does not require a sizeable accumulation of financial capital or are jobs without any need for prior training. The advantage of this approach is that the classification of informal work as upper-tier or lower-tier is undertaken based on the *observable characteristics of the job*, rather than the latent unobserved preferences of workers as in the first approach and outcome measures of job hierarchies as in the second approach. However, a limitation of this approach is that classifying informal workers as upper-tier or lower-tier does not imply a judgment about exit versus exclusion.

Our preferred approach is to follow the third approach to classify upper-tier and lower-tier informal employment. The surveys that we use in the four countries in the study have the necessary information to classify informal jobs as upper-tier or lower-tier, based on the

---

2 Similarly, Günther and Launov (2012) fit a finite mixture model to household survey data from Côte d'Ivoire to test for unobserved earnings heterogeneity in the informal sector. They identify an upper-tier segment that is superior to the lower-tier in terms of significantly higher average earnings as well as higher returns to education and experience. Their results also indicate that those in the upper-tier tend to have a comparative advantage in the informal sector while for those in the lower-tier informality is a strategy of last resort to escape unemployment.

observable characteristics of the job. We next discuss how we operationalize the classification of all informal jobs as upper-tier informal self-employment, lower-tier informal self-employment, upper-tier informal wage employment, and lower-tier informal wage employment for Ghana, South Africa, Tanzania, and Uganda.

We take upper-tier informal self-employment as self-employed workers with unregistered business activities who either employ at least one person (who is not a household member) or are in activities that require some type of professional training (defined as ISCO groups 1–4, covering managers, professionals, technicians, and clerks), while other non-professional own-account workers with unregistered business activities are classified as lower informal (examples of which are street vendors and waste pickers). The rationale for classifying informal employers as upper-tier informal self-employment is that the hiring of non-family workers for a household enterprise involves an implicit “barrier to entry” as these employers typically need to finance the wages of hired workers by borrowing from credit markets or through the profits of the enterprise (Banerjee et al., 2016).<sup>3</sup> All contributing family workers are classified as lower informal, irrespective of the nature of the enterprise. Workers in smallholder agriculture (family farms) would be classified as lower informal but have been excluded from the main analysis presented in this paper (see complementary tables in Appendix B for results including family farms).

Among the wage workers not covered by social protection provisions (who are classified as informal workers, in line with the ILO definition), those in professions that require some type of professional training (ISCO 1–4) are classified as upper informal, as they are “restricted entry”. In addition, we check whether workers report having a written employment agreement and/or are entitled to de facto benefits such as paid sick or maternity leave. The remainder is classified as lower informal.

In defining the activity status of a worker, we consider the status of the worker in the major job—defined as the economic activity on which the worker has spent most of his or her working time in the case of reporting multiple jobs, as well as the income from the main job.<sup>4</sup>

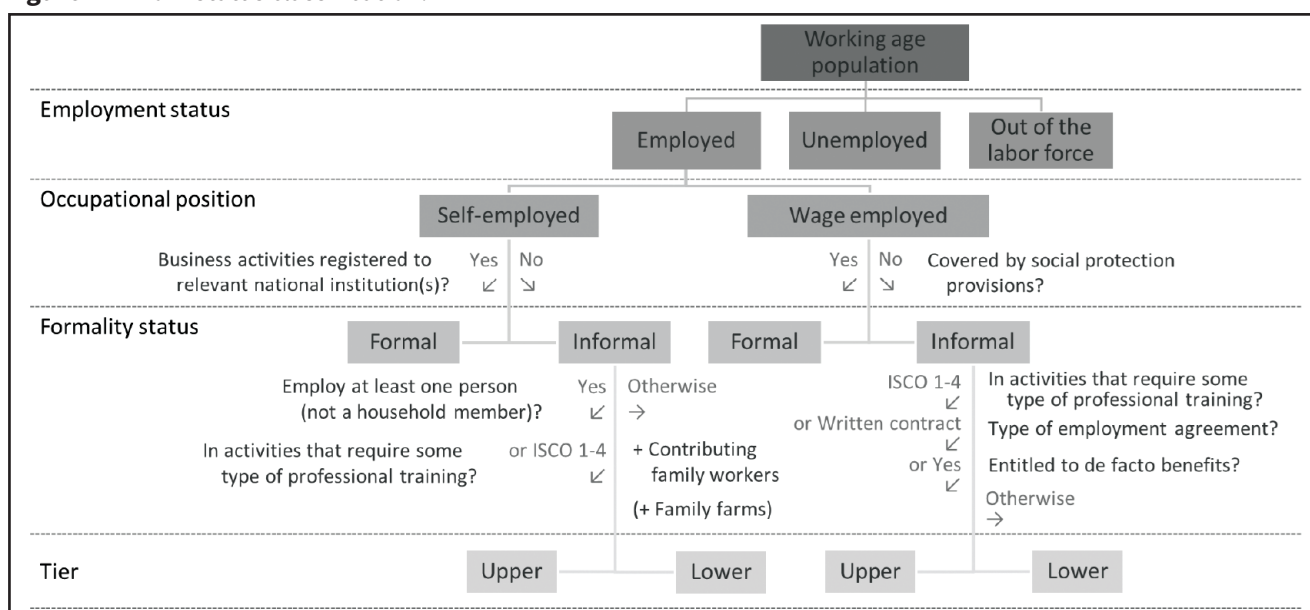
In Figure 1, we provide a diagrammatic overview of our classification schema. We start with the working-age population, where an individual may be employed, unemployed, or out of the labor force. Among those employed, workers may be self-employed or wage employed, depending on their occupational position. The self-employed and wage employed can be in formal or informal work. Within informal employment, the characteristics of the job or activity allow us to classify workers in upper-tier or lower-tier work, as discussed earlier. Thus, any individual in employment will be uniquely assigned to one out of six possible work statuses: (i) formal wage employment, (ii) formal self-employment, (iii) upper-tier informal wage employment, (iv) upper-tier informal self-employment, (v) lower-tier informal wage employment, and (vi) lower-tier informal self-employment.

Workers can switch from one work status to another over time. The schema that we propose does not constrain the direction of movement and allows for a combination of movements across occupational positions, formality states, and tiers, thus allowing for a complex set of

<sup>3</sup> In their study of informal enterprises in West Africa, Grimm et al. (2011) find significant entry barriers in the form of large sunk costs (such as the purchase of raw materials and building up inventories) in some segments of informal self-employment but not in others.

<sup>4</sup> For reasons of cross-country comparability, only information on the current/most recent employment status (past week/7 days) is considered. We do not account for workers holding multiple jobs.

**Figure 1** Work status classification.



Source: Authors’ own construct.

transition possibilities across all six work statuses. In the following, we aim to quantify the magnitudes of these various transitions in the countries under study and to quantify the extent to which these transitions are associated with income gains or losses.

### 3 Data and Methodology

#### 3.1 Data

This study is based on the employment modules of living standard household surveys from four sub-Saharan African countries: Ghana (Ghana Socio-Economic Panel Survey [GSPS]), South Africa (National Income Dynamics Study [NIDS]), Tanzania (Tanzania National Panel Study [TZNPS]), and Uganda (Uganda National Panel Study [UNPS]). In addition to the geographic focus on sub-Saharan Africa, the choice of countries is based on the share of employment outside of smallholder agriculture and the availability of at least two recent waves of nationally representative panel data with individual-level information on demographic characteristics, labor earnings, and employment, including direct or indirect information concerning the individual’s formality status in employment. For reasons of data availability and cross-country comparability, we focus the analysis on the two most recent waves of panel data available in each of the four countries under study. The data were collected between 2010 and 2017 with a two- to four-year time gap between panel waves (see Appendix A for full references to the sources of the data for the four countries). The sample is restricted to workers of prime working age (15 years to 65 years old). We convert the labor earnings reported for different time periods to monthly earnings based on self-reported working times.<sup>5</sup>

<sup>5</sup> All income data is deflated to 2010 prices and, for reasons of cross-country comparability, converted to international dollars using the World Bank’s purchasing power parity conversion factor for private consumption. Income levels above the 99th percentile of the distribution are considered outliers and replaced by the cut-off value. Only individuals working and reporting strictly positive cash income are included, whereas in-kind income is not taken into consideration. Agricultural income generated by family farms is excluded from the analysis as data on agricultural revenues and costs are relatively noisy.



## 3.2 Methodology

### 3.2.1 Employment transitions

We start the analysis of dynamics in work status using basic transition matrices. While these offer an interesting first assessment of the average chances of moving from one work status to another, the derived conditional transition probabilities ignore that the attributes of workers in different work statuses are bound to be different. However, we are interested in questions such as whether upper-tier informal wage workers are on average more likely to transition into formal wage employment than lower-tier informal wage workers, even after controlling for differences in observable worker characteristics. Therefore, we use a dynamic multinomial logistic regression model that allows estimating the extent to which the probability of being in specific work status at a time  $t = 1$  depends on the initial status in employment at the time  $t = 0$  and on observed individual and household attributes, including workers' age, gender, education, and geographic location (rural/urban).<sup>6</sup>

Dynamic multinomial models have been used in the labor economics literature for quite some time to model individual work status choices and the extent of state dependence. For example, the approach is used by Gong et al. (2004) to model the choice between informal and formal sector work in Mexico, Lee et al. (2018) to model work status choices in Korea, Beusch and van Soest (2019) to model dynamics of self-employment in the Netherlands, Buddelmeyer and Wooden (2011) and Oguzoglu (2016) to model labor dynamics in Australia, and Prowse (2012) to model transitions between full-time and part-time employment using data from Britain. The results from these studies show that not only individual and household level characteristics affect work status choices but also previous employment outcomes.

Most of the previous research has used longitudinal data to quantify the extent of “genuine state dependence,” understood as the extent to which the past work status itself affects future states—for example, through signaling or scarring effects to future employers, network effects, or the accumulation or depreciation of human capital—as opposed to heterogeneity in unobserved characteristics and preferences that cause individuals to sort and remain in certain employment states (Heckman, 1981, 1991). This distinction would require a larger number of repeated observations per individual than available in the data used here and is not the focus of this paper. However, the adopted regression framework allows us to understand to what extent observable characteristics (such as age, gender, and education) affect the chances of being in specific work status and to examine the extent to which we continue to see a significant association between past and present employment outcomes, even after controlling for these observable differences.

Following standard practices in the literature (Maloney, 1999; Gong et al., 2004; Liu, 2015), the statistical model can be formulated as follows. Suppose that each individual  $i$  is observed at two points in time  $t = \{0,1\}$ , where  $S_{i0}$  denotes the initial work status and  $S_{i1}$  denotes the final work status. If there are  $K$  possible response states indicating the individual's work status, then  $Pr(S_{it} = k | X_{it})$ , with  $k = \{1, \dots, K\}$ , is the probability that individual  $i$  has a response  $k$  at the time  $t$  given  $X_{it}$ , where  $X_{it}$  is a column vector of covariates for that observation.

<sup>6</sup> We experimented with adding additional controls on the worker's marital status and composition of the household to the analysis. While these are characteristics previously shown to be associated with the probability to work and the sector choice of workers (McKay et al. 2018; Sarkar et al. 2019), adding these to our analysis did not significantly affect our findings on the key variables of interest. Results are available from the authors upon request.

The multinomial model can then be expressed as

$$\pi_{i1k} = \Pr(S_{i1} = k | S_{i0}, X_{i0}) = \left[ 1 + \sum_{j=1}^K e^{S_{i0}\theta_j + X_{i0}\beta_j} \right]^{-1} e^{S_{i0}\theta_k + X_{i0}\beta_k} \quad (1)$$

where  $\pi_{i1k}$  represents the transition probability from work status  $S_{i0}$  at baseline to destination state  $S_{i1} = k$  at the end of the period,  $\theta_k$  is the vector of regression parameters for the work status at baseline, and  $\beta_k$  is the vector of regression parameters for the other covariates on outcome state  $k$ . As in any regular multinomial logit regression, the probability  $\pi_{i1(K+1)}$  is specified as the residual probability (see Liu, 2015).

In our analysis, the response has six states: (i) formal wage employment, (ii) formal self-employment, (iii) upper-tier informal wage employment, (iv) upper-tier informal self-employment, (v) lower-tier informal wage employment, and (vi) lower-tier informal self-employment. For identifiability, lower-tier informal self-employment is set as the reference category.

By functional transformation, the multinomial logit model can be expressed as a generalized linear model, given by

$$\log \left( \frac{\pi_{i1k}}{\pi_{i1(K+1)}} \right) = S_{i0}\theta_k + X_{i0}\beta_k \quad \text{where } k = 1, \dots, K. \quad (2)$$

### 3.2.2 Labor income dynamics

We further analyze the link between employment mobility and changes in earnings. To do so, we regress the change in the logarithm of individual labor earnings between time  $t = 0$  and time  $t = 1$ ,  $\Delta y_{i1}$ , on the individual's initial log earnings,  $y_{i0}$ , initial work status,  $S_{i0}$ , and the set of initial worker characteristics,  $X_{i0}$ . Our main interest is in estimating the relationship between transitions in work status and changes in workers' earnings. For this purpose, we add an interaction term between the initial and the final work status,  $(S_{i0} \times S_{i1})$ . Our dynamic income model is given by

$$\Delta y_{i1} = \alpha + \delta y_{i0} + \nu_0 S_{i0} + \nu_1 (S_{i0} \times S_{i1}) + \varphi X_{i0} + \varepsilon_i \quad (3)$$

where  $\varepsilon_i$  denotes the estimation error. In combination, the coefficient estimates for  $\vartheta_0$  and  $\vartheta_1$  are interpreted as a measure of the conditional earnings premium (or penalty) associated with staying in or moving to particular work status. The advantage of this approach is that it does not assume that the earnings change associated with a transition from work status  $S_{i0} = k$  to  $S_{i1} = l$  is reversely identical to the earnings change associated with a transition from work status  $S_{i0} = l$  to  $S_{i1} = k$ . We compare our results to the earnings premia by work status estimated using a fixed-effects panel regression model.

### 3.2.3 Initial employment and attrition

We focus our analysis on the employment dynamics observed among individuals who are active in the labor market. Accordingly, our study does not cover entry or exit dynamics. We concentrate specifically on the subset of workers who were employed in non-farm activities in

the initial panel wave as opposed to working on a family farm or being unemployed.<sup>7</sup> These are likely a non-random sample of the active population (Heckman, 1981). Furthermore, we may have a sample selection problem if attrition is endogenous for estimating employment transitions, as some of the existing literature suggests (Cappellari and Jenkins, 2004; Sarkar et al., 2019).

In order to correct for the double selection problem due to initial employment and panel attrition, we adopt a similar empirical strategy as Sarkar et al. (2019) drawing on previous studies. This method essentially builds on Heckman's two-stage estimation procedure using the inverse Mills ratio to correct for the selection bias. We specify the initial employment decision and sample retention in the following set of equations:

$$E_{i0}^* = \gamma' Z_{i0} + v_{i0} \quad \text{with} \quad E_{i0} = I(E_{i0}^* > 0) \quad (4)$$

$$R_{i1}^* = \psi' W_{i0} + u_{i1} \quad \text{with} \quad R_{i1} = I(R_{i1}^* > 0) \quad (5)$$

where  $E_{i0}^*$  captures the latent propensity of non-farm employment in  $t = 0$ ,  $R_{i1}^*$  is the latent propensity of panel retention from  $t = 0$  to  $t = 1$ ,  $Z_{i0}$  and  $W_{i0}$  are vectors of baseline characteristics, and  $I(\cdot)$  are binary indicator functions equal to one if the underlying latent propensity exceeds some unobserved value (which can be set to zero without loss of generality) and equal to zero otherwise.

Since initial employment and sample attrition are likely to be correlated, we adopt the framework by Sarkar et al. (2019) following the approach suggested by Vella (1998) and estimate questions (4) and (5) simultaneously using a bivariate probit model. We then calculate the selection correction terms for initial employment and sample retention (for the derivation, see Sarkar et al., 2019) and control for these when estimating Eqs. (2) and (3).<sup>8</sup> For identification,  $Z_{i0}$  and  $W_{i0}$  should include some explanatory variables that affect initial employment and panel retention, respectively, but are excludable from the main employment transitions equation. Relevant test statistics for the selected variables are provided and discussed in section 5.1.

For the propensity of initial employment in non-farm activities, we use a binary variable identifying household heads as an instrument. Our key assumption here is that household decision-making leads to household heads having higher chances to be employed outside of agriculture (while other household members stay behind to work on the family farm) but, after controlling for differences in observable characteristics (such as age, gender, and education), their transition probabilities between work status categories do not differ from workers who are not heads of household.

Finding a valid instrument for the propensity of panel retention that is available across countries proved difficult. Following a similar approach to Schotte et al. (2018), for South Africa, Tanzania, and Uganda—where at least one previous wave of panel data is available (see Appendix A for details)—we use a binary variable indicating whether the respondent was a sample member in the previous survey wave. The identifying assumption here is that original sample members are more stable survey members compared to newly added members and

<sup>7</sup> We focus our analysis on transitions between the six specified employment states. An expansion of this analysis including family farm activities (under lower-tier informal self-employment) and unemployment as additional destination states is provided in Appendix B.

<sup>8</sup> The inclusion of the sample selection terms marginally affects the size of the coefficient estimates (mostly at the two- or three-digit level) but does not change the conclusions presented in the paper.

that the previous round's sample membership status has no direct effect on the employment transition propensity. In Ghana, identification relies on the non-linear form of the inverse Mills ratio.

## 4 Descriptive Analysis

A key strength of our analysis is its comparative nature. The countries included in our study cover different regions—Western Africa (Ghana), Eastern Africa (Tanzania and Uganda), and Southern Africa (South Africa)—as well as the different levels of development—including upper-middle-income (South Africa), lower-middle-income (Ghana), and low-income (Tanzania and Uganda) countries. We begin this section with a brief discussion of the macroeconomic and institutional context in the four countries under study. We then use the survey data to assess differences and commonalities in the composition of employment across the four countries both from a static and from a dynamic perspective and attempt to link these back to the countries' economic and labor market structure.

### 4.1 Macroeconomic and institutional country profiles

Economic development and structural transformation, along with the type and quality of labor market institutions, shape the structure of employment and determine what labor market opportunities are available to workers. Table B1 in Appendix B, therefore, provides a set of summary statistics on selected key indicators, averaged over the 2010–2017 study period, which may be relevant to understanding the employment transition patterns analyzed in this paper.

As an upper-middle-income emerging market economy, South Africa is the most economically developed country in our sample. Yet, economic growth has decelerated in recent years, slowing to an average rate of 0.4% estimated over the 2010–2017 period. In comparison, solid growth in per capita gross domestic product (GDP) was recorded in Ghana (4.4%) and Tanzania (3.4%) while Uganda experienced slower but yet positive economic growth (1.7%). In all four countries, the service sector is the main contributor to GDP. While in Ghana, Tanzania, and Uganda, agriculture and industry contribute about the same share to GDP, agriculture plays a substantially less important role in South Africa (accounting for only 2.2% of GDP).

Importantly, South Africa's comparatively high average income level does not equally benefit all its citizens, considering high levels of inequality in the country. Despite being more economically advanced, South Africa's poverty headcount—on average amounting to 17.5% of the population living in extreme poverty (\$1.90 a day) and 36.5% in moderate poverty (\$3.20 a day)—has remained at a higher level than estimated in Ghana. The poverty rate was substantially higher in Uganda and highest in Tanzania—with an average of 49.5% of the population living in extreme poverty and 78.2% in moderate poverty by international standards.

Mirroring the differences in economic development, South Africa has the highest educational attainment, followed by Ghana. Educational attainment at the primary level is higher in Tanzania than Uganda, but lower- and upper-secondary educational attainment is lower compared to Uganda. However, the quality of vocational training seems to be higher in Tanzania, followed by Ghana, Uganda, and South Africa (World Economic Forum, 2019).

Regarding the labor market structure, employment in agriculture and services is relatively balanced in Ghana, together accounting for 83.4% of total employment (41.6% and 42.2%, respectively). In South Africa, employment in services is highest (71.3%), whilst the share in agriculture is lowest (5%). In Tanzania and Uganda, the dominant share of employment is in agriculture (68.3% and 70.1%, respectively), followed by the service sector (25.5% and 22.6%, respectively) and low shares in the industry (6.1% and 7.2%, respectively). Public sector employment makes up almost half (49.8%) of formal employment in Tanzania, the highest share among the four countries, followed by Ghana (45.4%), Uganda (39.5%), and South Africa (22.7%). Unionization rates are highest in South Africa, where 20.4% of private paid employees have union membership, followed by Ghana (17.5%), Tanzania (9.4%), and Uganda (3%).

The ease of doing business score indicates that South Africa is relatively more business-friendly, followed by Ghana, Uganda, and Tanzania.

The overall governance quality index places Ghana and South Africa ahead of Uganda and Tanzania. The regulatory performance of these countries in terms of tax payments shows that South Africa does better than the other countries. This is followed by Uganda and Ghana. With respect to workers' rights, South Africa performs much better in labor protection, an indication of a highly regulated formal labor market. This is followed by Ghana and Tanzania.

## 4.2 Composition of employment

The summary statistics presented in Table 1 show the aggregated distribution of workers in employment by work status. As widely established in the literature, the composition of the workforce in South Africa differs remarkably from the employment structure observed in poorer sub-Saharan African countries. In South Africa, we observe that 60.6% of those in non-farm employment are formally employed (see Table 1). By contrast, in Ghana and Tanzania, only about 20% and, in Uganda, 16.6% of those employed in non-farm activities are in formal employment, which means that about 80% of the non-farm employment in these three countries is informal. Most of the informal workers in these three countries are in lower-tier informal self-employment, accounting for >40% of all non-farm employment. The latter share would be yet substantially larger if family farms were included in the analysis (forming part of the lower-tier segment of informality), raising the informality rate in Tanzania and Uganda to just >90% (see Table B2 in Appendix B).

The main difference thus consists in the relative absence of lower-tier informal self-employment in South Africa compared to the three other countries. This difference can be attributed to two factors. First, South Africa's economy provides relatively more employment opportunities in the formal economy. Second, in South Africa, a larger share of workers can afford to be openly unemployed (23.3%) compared to the three poorer countries (<2%), where workers revert to survivalist self-employment strategies in the absence of other job opportunities and sufficiently developed social protection systems.

We observe no large changes in these country-level employment structures between survey waves (see Table B3 in Appendix B). In Ghana, South Africa, and Uganda, the share of individuals in self-employment in the balanced panel (workers employed in non-farm activities in both panel waves) moderately increased, while it slightly decreased in Tanzania. Ghana, South Africa, and Tanzania show an increase in the aggregate rate of formal employment among the

**Table 1** Distribution of workers by work status (%)

		Ghana	South Africa	Tanzania	Uganda	
a) Proportion of employment by work status						
Wage employed	Formal	13.2	56.7	11.7	11.7	
	Informal	Upper	5.3	8.9	3.3	10.8
		Lower	18.9	21.4	28.6	26.0
Self-employed	Formal	8.9	4.0	9.1	3.1	
	Informal	Upper	11.9	5.4	3.9	5.8
		Lower	41.8	3.7	43.4	42.6
TOTAL		100	100	100	100	
b) Proportion of formal vs. informal employment						
Formal		22.1	60.6	20.7	14.9	
Informal	Upper	17.2	14.3	7.2	16.6	
	Lower	60.7	25.1	72.0	68.6	
TOTAL		100	100	100	100	
c) Proportion of upper-tier informality in informal employment						
Upper informal in total informal employment		22.1	36.4	9.1	19.4	
Upper informal in informal self-employment		22.2	59.7	8.2	11.9	
Upper informal in informal wage employment		22.0	29.4	10.5	29.4	

*Note:* For each country, summary statistics are compiled for the initial wave of panel study under study. Workers employed on family farms have been excluded from the analysis.

*Source:* Authors' own calculations based on survey data from GSPS 2009/10, NIDS 2014/15, TZNPS 2010/11, and UNPS 2010/11.

GSPS, Ghana Socioeconomic Panel Study Survey; NIDS, National Income Dynamics Study; TZNPS, Tanzania National Panel Study; UNPS, Uganda National Panel Study.

balanced panel by 1.5 to 2.5 percentage points (ppts). In Ghana and Tanzania, this was mainly driven by a rise in the share of formal wage employment, mirrored by a decline in the upper-tier segment of informal wage work. By contrast, in South Africa and Uganda, we see an expansion in formal self-employment, accompanied by a moderate decline in formal wage employment. These relatively small changes in aggregate shares tend to mask substantial mobility of workers across employment categories, which will be discussed in the next section.

Table B4 in Appendix B presents three additional descriptive features of our data. First, we find that informal employment is more common among younger workers. Second, women tend to be under-represented in formal wage and self-employment and to be importantly over-represented in lower informal employment. Finally, workers with secondary or tertiary education are over-represented in formal employment, while workers who have either no education or only completed primary schooling are dominantly found in informal employment.

We also test for statistically significant differences in mean labour earnings across the defined work status groups, using an analysis of variance (ANOVA) test (see Figure C1 in Appendix C). As expected, labor incomes tend to be higher in formal employment than in informal employment. Moreover, considerable heterogeneity exists within the two informal sub-segments, with earnings being significantly higher in upper-tier informal activities compared to the lower-tier. This pattern is consistent across countries.

### 4.3 Patterns of transitions

We now move from the preceding static assessment to a dynamic perspective. The transition matrices in Table 2 report the probabilities with which workers are observed in a certain work status at the end of the period, conditional on their initial state. Accordingly, the elements in the main diagonal give the probabilities of staying in the same work status, while the elements outside the main diagonal give the probabilities of moving to a different work status. The share of stayers, defined as the proportion of workers who remain in their work status, is calculated as the product of the highlighted diagonals and the initial share of workers in the respective category. We observe the highest employment mobility in Ghana, where just about 50% of all initially employed individuals were observed in the same work status at the end of the period. In the other three countries, the same was true for about 60% of all workers.

While we observe some differences in transition patterns across countries, a key commonality observed is that employment stability tends to be highest among the formally wage employed. This may partly be attributed to these jobs being regulated and protected by existing legal standards. In South Africa and Tanzania, around 80% of all workers in formal wage employment remain in this work status from one survey wave to the next. This share is somewhat lower in Ghana, at 65.1%, and lowest in Uganda, at 48.3%. The stability of formal wage employment in South Africa may be related to the country's highly regulated formal labor market with strong trade unions, while in Tanzania, the same may be attributed to the dominant contribution of public sector employment to formal wage employment (see Table B1 in Appendix B).

Labor turnover tends to be higher in formal self-employment, with important differences observed across countries. In South Africa, among the formally self-employed, 50.8% stay in this state, 13% move into formal wage employment, 23% move into upper-tier informality, and only 13.2% move into lower-tier informality (being either self- or wage employed). On the contrary, only around 30% of the formally self-employed in Ghana, Tanzania, and Uganda remain in formal self-employment from one survey wave to the next, while up to 40% move into lower-tier informal self-employment. While these movements may partly be explained by reporting errors, business instability is assumed to also play a major role. This gap in the stability of formal business activities between South Africa and the other three countries can be related to differences in the countries' level of economic development as higher levels of income will provide more stability to the formal business sector.

Furthermore, in Ghana, Tanzania, and Uganda, we observe high stability within lower-tier informal self-employment, with >80% of the respective workers either remaining in this category or moving into lower-tier informal wage-employment. The "stickiness" in this segment reflects the limited alternative job opportunities available to workers in this group. This can be related to the countries' lower level of economic development and economic structure, offering relatively fewer employment opportunities in the upper-tier or formal segments of the labor market. Notably, when including family farming under lower-tier informal self-employment activities in the destination state, we observe an even higher level of persistence in this segment. Hence, a non-negligible share of workers may draw on a combination of self-employment in agriculture and in lower-tier non-agricultural informality (see Table B5 in Appendix B).

**Table 2** Transition matrices across work status groups

			WAVE t = 1						Share of stayers	
			Wage employed			Self-employed				
			Formal	Informal		Formal	Informal			
				Upper	Lower		Upper	Lower		
a) Ghana										
WAVE t = 0	Wage employed	Formal	65.1	6.9	9.6	4.3	3.6	10.6	8.6	
		Informal	Upper	33.0	11.8	23.7	2.5	3.2	25.8	0.6
	Lower		21.7	5.3	32.2	7.5	6.4	26.9	6.1	
	Self-employed	Formal	2.1	3.4	12.1	29.5	15.8	37.1	2.6	
		Informal	Upper	4.0	2.5	15.2	14.7	44.7	18.9	5.3
			Lower	2.1	2.0	15.8	6.8	6.3	67.2	28.1
	TOTAL		21.0	4.5	18.4	9.1	11.2	35.8	51.3	
b) South Africa										
WAVE t = 0	Wage employed	Formal	83.3	7.1	5.8	2.2	1.1	0.5	47.2	
		Informal	Upper	50.1	25.1	14.7	6.0	1.0	3.1	2.2
			Lower	26.4	13.1	47.7	4.7	3.9	4.3	10.2
	Self-employed	Formal	13.0	9.7	5.1	50.8	13.3	8.1	2.0	
		Informal	Upper	12.2	11.1	19.5	16.1	23.5	17.6	1.3
			Lower	14.7	6.2	16.8	6.9	24.6	30.8	1.1
	TOTAL		63.4	9.9	14.9	5.2	3.4	3.1	64.0	
c) Tanzania										
WAVE t = 0	Wage employed	Formal	79.7	4.4	9.7	1.5	1.0	3.6	9.3	
		Informal	Upper	45.8	5.8	33.1	6.5	2.9	6.0	0.2
			Lower	14.8	0.9	62.4	2.9	3.5	15.5	17.9
	Self-employed	Formal	2.0	0.0	5.0	31.7	18.2	43.1	2.9	
		Informal	Upper	4.5	0.0	21.8	15.2	23.1	35.4	0.9
			Lower	5.8	1.5	15.3	8.3	4.1	64.9	28.2
	TOTAL		22.2	1.8	27.3	8.6	6.0	34.2	59.3	
d) Uganda										
WAVE t = 0	Wage employed	Formal	58.0	27.0	8.1	5.2	0.0	1.7	6.8	
		Informal	Upper	20.0	48.3	14.1	5.7	6.1	5.9	5.2
			Lower	4.7	6.2	68.3	2.7	4.3	13.9	17.7
	Self-employed	Formal	2.7	0.0	10.5	20.7	15.0	51.1	0.6	
		Informal	Upper	0.0	4.6	5.0	13.9	39.8	36.8	2.3
			Lower	2.2	1.0	11.9	4.4	7.2	73.2	31.2
	TOTAL		14.8	13.3	24.3	5.7	7.9	34.1	63.9	

Note: Each row indicates work status in the base period, and each column in transition matrices indicates work status in the next period; transition matrix rows sum to 100. The likelihood of staying in the same employment status conditional on the base year employment status is highlighted in gray. The share of stayers (proportion of workers who remain in their work status) is calculated as the product of highlighted diagonals and initial size.

Source: Authors' own calculations based on survey data from GSPS 2009/10-2013/14, NIDS 2014/15-2017, TZNPS 2010/11-2012/13, and UNPS 2010/11-2011/12.

GSPS, Ghana Socioeconomic Panel Study Survey; NIDS, National Income Dynamics Study; TZNPS, Tanzania National Panel Study; UNPS, Uganda National Panel Study.



In all four countries, those in upper-tier informality are more likely to move into formality compared to those in lower-tier informality. This difference tends to be more pronounced among the wage employed. This may imply that workers take on informal employment to gain work experience (either voluntarily or due to the limited supply of formal (or regular) job opportunities) before moving into better-paying activities. It may also reflect information asymmetries, where employers first employ workers informally to test their abilities before providing formal contracts. Interestingly, this type of upward mobility is observed to be highest in South Africa and Tanzania, where about one out of two workers moved from upper-tier informal into formal wage employment between survey waves, followed by Ghana, where the same applied to one out of three upper-tier informal wage employees, and lowest in Uganda, applying to one out of five upper-tier informal wage workers. Higher levels of secondary education, along with more formal job opportunities, better regulatory performance, and stricter enforcement of workers' rights in South Africa can explain this pattern (see Table B1 in Appendix B). The case of Tanzania appears less clear-cut. While the relatively higher quality of vocational training may play a role (see Section 4.1), the small size of upper-tier informal wage employment in Tanzania may also affect the estimate.

Furthermore, in South Africa and Ghana, recognizable mobility out of lower-tier informal wage employment into formal wage employment is observed, suggesting that for about 20% of all workers in this group lower-tier informal wage employment can present a stepping-stone into formal employment relationships. The same applies to about 15% of lower-tier informal wage employees in Tanzania and <5% of the same groups of workers in Uganda.

## 5 Regression Analysis

We begin this section with a short discussion of the estimation results of the two selection equations capturing initial employment and panel retention. Subsequently, we focus on the dynamics in employment status and labor income.

### 5.1 Initial employment and attrition

The subset of workers who were employed in non-farm activities in the initial panel wave, as opposed to working on a family farm or being unemployed, ranges from 78.1% in South Africa, 51.7% in Ghana, 34% in Uganda to 33.2% in Tanzania (see Table B6 in Appendix B). Panel retention rates among the active workforce are highest in South Africa (75.9%) and lowest in Tanzania (69.1%).

As indicated in Section 3.2, for the propensity of initial employment in non-farm activities, we use a binary variable identifying household heads (as opposed to other household members) as an instrument. Across countries, heads of household are significantly more likely to be initially employed in non-farm activities (see Table B7 in Appendix B), while the variable is validly excludable from the main employment transition equation (see Table B8 in Appendix B).

Analogously, meeting our identifying assumptions, the propensity of panel retention is significantly higher among respondents who were members of the previous survey wave (see Table B7 in Appendix B), while the previous round's sample membership status is validly excludable from the main employment transition equation (see Table B8 in Appendix B).

## 5.2 Employment transitions

Table 3 presents the conditional transition probabilities estimated from the multinomial logit regression. The average marginal effects in each column are calculated by destination work status in  $t = 1$ . The reference status that is used as both transitions' starting point and destination is lower-tier informal self-employment. To extract overall patterns, in our main analysis we pool the data for all countries so that the displayed results present cross-country average marginal effects. To investigate between-country differences in transition patterns, we re-estimate the regression including interactions between the initial work status and country indicators (see Figure C2 in Appendix C). We also test how our results change when including unemployment or being out of the labor force as an additional destination state (see Table B9 in Appendix B) and further explore the coefficient estimates of our control variables using a condensed ordered logit specification (see Table B10 in Appendix B).<sup>9</sup>

The average marginal effects on initial work status reported in the upper panel of Table 3 can be read similarly to the conditional transition probabilities of a transition matrix.<sup>10</sup> In this sense, the coefficient estimates on being in employment state  $k = \{1, \dots, 5\}$  at time  $t = 1$ , conditional on being observed in the same state  $k$  at time  $t = 0$  (main diagonal) give an indication of the degree of persistence or state dependence in employment status that is not explained by differences in education, age, gender, and geographic location (Gong et al., 2004; Liu, 2015).

Interestingly, we find a relatively strong segmentation between wage and self-employment. Workers in formal self-employment are most likely to remain in this state or move into informal self-employment. By contrast, transitions from formal self- to formal wage employment are rare and even less likely to occur than a move from lower informal to formal wage employment. Similarly, we find a high degree of persistence in formal wage employment, which is expected given the prevalence of permanent contracts in these groups. In line with the unconditional transition probabilities reported in Section 4.3, the estimation including interactions between the work status of origin and country fixed-effects shows that the conditional marginal probability of remaining in formal wage employment is highest in South Africa and Tanzania, and lowest in Uganda (see Figure C2a in Appendix C). In comparison, formal self-employment is a much more dynamic state, with higher dropouts into informality being observed in Ghana, Tanzania, and Uganda than in South Africa (see Figure C2d in Appendix C).

Different from the findings in the literature on Latin America (Maloney, 1999; Bosch and Maloney, 2010), we find workers originating from formal wage jobs do not display an elevated likelihood of moving into formal or upper-tier informal self-employment in any of the countries. Our evidence thus does not lend support to the hypothesis that workers use the human capital acquired in formal wage jobs to set up their own businesses and benefit from greater flexibility and independence. While workers in lower-tier informal wage jobs are more likely than other wage workers to move into self-employment, we still observe an important extent of segregation between wage and self-employment even in the lower-tier of informality.

<sup>9</sup> For the ordered logit specification, compared to the multinomial logit model specified in Eq. (2), we group together workers in wage and self-employment and define the individual's formality status,  $F_i$ , on an ordinal scale, taking on three possible values: lower-tier informal employment,  $F_i = 1$ ; upper-tier informal employment,  $F_i = 2$ ; and formal employment,  $F_i = 3$ . We split the sample by the formality status at time  $t = 0$  and express the formality status at time  $t = 1$  as a function of the same set of initial worker attributes  $X_{i0}$ .

<sup>10</sup> The likelihood ratio test rejects the null hypothesis that the dynamic factors play no role, i.e., the lagged labor states are jointly significant ( $p$ -value of 0.0000) and in most cases also individually significant.

**Table 3** Employment transitions

Multinomial logistic regression		Number of obs = 7,816			
Average marginal effects on work status in $t = 1$		Log-likelihood = -7,851.1291			
Base outcome: Lower-tier informal wage employed		Pseudo $R^2 = 0.3353$			
	(1)	(2)	(3)	(4)	(5)
Variables	Formal wage employed	Upper-tier informal wage employed	Lower-tier informal wage employed	Formal self-employed	Upper-tier informal self-employed
Work status in $t = 0$ (base: lower-tier informal self-employed)					
(1) Formal wage employed	0.628*** (0.028)	0.037** (0.016)	-0.066** (0.033)	-0.128*** (0.014)	-0.096*** (0.029)
(2) Upper-tier informal wage employed	0.293*** (0.026)	0.192*** (0.015)	0.044 (0.033)	-0.099*** (0.014)	-0.078*** (0.028)
(3) Lower-tier informal wage employed	0.153*** (0.008)	0.062** (0.024)	0.264*** (0.042)	-0.104*** (0.016)	-0.062** (0.025)
(4) Formal self-Employed	-0.045* (0.026)	-0.007 (0.023)	-0.079*** (0.020)	0.227*** (0.033)	0.043 (0.044)
(5) Upper-tier informal self-employed	-0.012 (0.024)	0.028 (0.022)	-0.023 (0.030)	0.016 (0.014)	0.201*** (0.056)
Level of education (base: no schooling)					
Primary	0.053*** (0.018)	0.001 (0.009)	-0.089*** (0.025)	0.023*** (0.006)	0.001 (0.004)
Post-primary	0.121*** (0.023)	0.011 (0.026)	-0.160*** (0.033)	0.050*** (0.005)	0.004 (0.005)
Secondary	0.196*** (0.025)	0.008 (0.022)	-0.223*** (0.034)	0.052*** (0.003)	-0.002 (0.012)
Post-secondary	0.246*** (0.028)	0.026 (0.022)	-0.283*** (0.036)	0.096*** (0.020)	-0.000 (0.012)
Tertiary	0.342*** (0.024)	-0.001 (0.019)	-0.337*** (0.041)	0.107*** (0.038)	-0.016 (0.020)
Age	0.003 (0.003)	-0.002 (0.002)	-0.010*** (0.003)	0.006 (0.004)	0.000 (0.002)
Age squared ( $\times 0.01$ )	-0.002 (0.004)	0.002 (0.003)	0.009*** (0.003)	-0.006 (0.005)	-0.001 (0.003)
Female	-0.037*** (0.009)	0.029*** (0.003)	-0.008 (0.011)	-0.021*** (0.005)	-0.016*** (0.001)
Urban	0.029*** (0.006)	-0.027*** (0.007)	0.015 (0.018)	0.007 (0.009)	-0.013*** (0.004)

(Continued)

**Table 3** Continued

	(1)	(2)	(3)	(4)	(5)
<b>Variables</b>	<b>Formal wage employed</b>	<b>Upper-tier informal wage employed</b>	<b>Lower-tier informal wage employed</b>	<b>Formal self-employed</b>	<b>Upper-tier informal self-employed</b>
Country (base: Ghana)					
South Africa	0.122*** (0.006)	0.037*** (0.007)	0.005 (0.012)	-0.010 (0.006)	-0.020*** (0.003)
Tanzania	0.112*** (0.018)	-0.033*** (0.004)	-0.019*** (0.006)	0.017*** (0.002)	-0.036*** (0.005)
Uganda	-0.106*** (0.005)	0.099*** (0.010)	0.011 (0.007)	-0.018*** (0.002)	0.007*** (0.002)
Sample selection					
Panel retention from $t = 0$ to $t = 1$	-0.044** (0.020)	-0.019** (0.009)	-0.026* (0.015)	-0.007 (0.011)	-0.009 (0.013)
Employed (non-farm) in $t = 0$	0.011 (0.012)	-0.002 (0.021)	0.010 (0.022)	0.005 (0.014)	-0.025** (0.010)

Note: \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ . Standard errors clustered at the country level in parentheses.

Source: Authors' own calculations based on survey data from GSPS 2009/10-2013/14, NIDS 2014/15-2017, TZNPS 2010/11-2012/13, and UNPS 2010/11-2011/12.

GSPS, Ghana Socioeconomic Panel Study Survey; NIDS, National Income Dynamics Study; TZNPS, Tanzania National Panel Study; UNPS, Uganda National Panel Study.

In accordance with the descriptive evidence reported in Section 4.3, the econometric analysis confirms that workers in upper-tier informality have significantly higher chances of moving into formal employment than those in the lower tier, and this difference tends to be more pronounced among the wage employed. In the pooled regression, the same does not apply within self-employment, where lower-tier and upper-tier informal self-employment display a similar (not statistically different) conditional likelihood to formalize. However, in the interacted specification, we find significantly higher rates of business formalization among those in the upper-tier (see Figure C2d in Appendix C).

The regression results further indicate that even after controlling for differences in education, location, age, and initial work status, women are less likely than men to be formally employed. They are more likely to engage in lower-tier informal self-employment or to work in upper-tier informal wage jobs, which excludes them from the social protection benefits associated with formal wage employment. This may partly be explained by a higher preference for more flexible job arrangements but may also be attributable to the difficulty of females finding jobs in the formal economy. Our results, obtained using the ordered logit specification, confirm that women, on average, face a higher likelihood of dropping out of formal jobs and more often than men slip into upper-tier and particularly lower-tier informality (see Table B10 in Appendix B).

Moreover, we find that higher levels of education are associated with a higher likelihood of working formally (see Table 3). Interestingly, the correlation between educational attainment and formality status is stronger in wage employment than in self-employment. From this, we

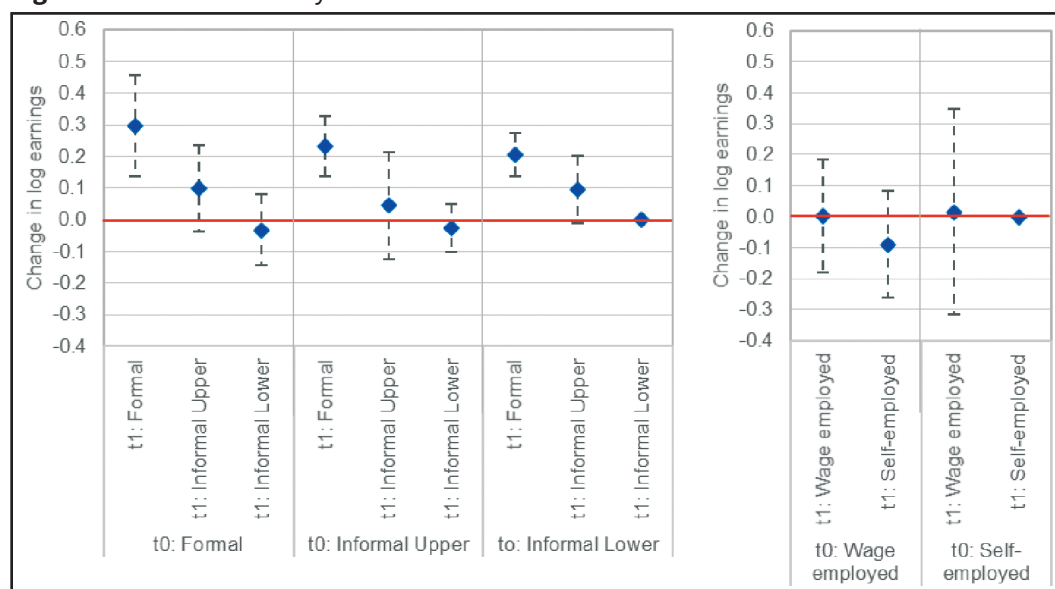
conclude that the lack of schooling presents an important barrier to attaining formal or upper-tier informal wage jobs, while on the business side other barriers such as access to credit may play an additional role.

### 5.3 Labor income dynamics

Last, we investigate changes in labor earnings by initial and destination work status. The main estimated effects are displayed in Figure 2 (see Table B11 in Appendix B for the full regression results). To keep the number of transition categories manageable for illustrative purposes, we separately control for formality status and tier (formal vs. upper-tier informal vs. lower-tier informal) and occupational position (wage vs. self-employment). The observed patterns are consistent with an alternative specification that splits the sample by initial work status, displayed in Figure C3 in Appendix C.

We estimate that workers who transition from lower-tier informal employment to upper-tier informal employment on average experience a 9.5% rise in earnings relative to those who stay in the same category. As expected, a larger positive earnings effect of 20.6% is found for those who move into formal employment. These effects are somewhat smaller than the coefficients estimates derived from the fixed effects panel regression model, displayed in Figure C4 in Appendix C.<sup>11</sup> Based on our findings, we cannot reject the hypothesis that the earnings premium (or penalty) associated with moving from one work status to another is symmetric to the penalty (or premium) associated with the reverse move.

**Figure 2** Labor income dynamics.



Note: Each point shows the estimated marginal effect on changes in log earnings by initial and destination employment state, with “Informal Lower” and “Wage Employed” being the base categories. The dashed lines show the 95% confidence intervals.

Source: Authors’ own calculations based on survey data from GSPS 2009/10-2013/14, NIDS 2014/15-2017, TZNSPS 2010/11-2012/13, and UNPS 2010/11-2011/12.

<sup>11</sup> These suggest that, on average, upper-tier informal employment is associated with an earnings premium of 18.3% and formal employment with an earnings premium of 31.4% compared to lower-tier informality.

Ceteris paribus, the largest inter-temporal change in earnings is experienced by those who were initially in formal employment and sustained this status over time. This may partly be attributable to unobserved individual characteristics of workers in this group but may also reflect a premium on experience in this labor market segment. Interestingly, while workers moving from formal to upper-tier informal employment experience a less favorable change in earnings compared to those who remain formal, they still tend to be better off than those who were already initially in upper-tier informal employment and maintained this status.

Furthermore, we observe that transitions from self- to wage employment are not significantly associated with an earnings premium. The transition from wage to self-employment tends to come with an earnings penalty, which however is not statistically significant.

## 6 Conclusion

Using panel data from Ghana, South Africa, Tanzania, and Uganda, this paper offered a comparative perspective on the composition of employment in four sub-Saharan African countries and documented the transition patterns between formal and informal employment and across different forms of informality, distinguishing between wage and self-employment.

Our analysis revealed that the distinction between lower-tier and upper-tier informal work is consequential in terms of both employment and earnings dynamics. We found that for most workers, informal work especially in the lower tier rather presents a “dead end” than a “steppingstone.” This particularly applies to workers in lower-tier informal self-employment, who often remain in this position of inferior pay and conditions. Lower-tier informal wage employment can present an entry point to formal jobs, the more developed the economy and the more formal jobs are available. In South Africa, about one out of four workers ascend from lower-tier informal to formal wage employment, compared to one out of twenty in Uganda.

Across countries, upper-tier informality presents a more dynamic state with a higher proportion of workers formalizing than in the lower-tier. Particularly workers in upper-tier informal wage jobs have significantly higher chances of moving into formal wage jobs than those in the lower tier. This result may partly be explained by formal employers using an informal employment relationship as a screening device before providing formal contracts. Moreover, we find that workers who transition from lower- to upper-tier informal employment on average experience a rise in earnings. This positive earnings effect is yet larger for those who move into formal employment.

As expected, employment stability tends to be highest among the formally wage employed. This can be attributed to these jobs being regulated and protected by existing legal standards. By contrast, formal self-employment is a much more dynamic state, with particularly high mobility into upper- or lower-tier informal self-employment being observed in Ghana, Tanzania, and Uganda. While these movements may partly be explained by reporting errors, business instability probably plays a major role. Interestingly, we find a relatively strong segmentation between wage and self-employment. That is, transitions between self- and wage employment are comparatively rare and mainly occur in the lower tier of informality. Exiting formal or upper-tier informal wage employment for self-employment is not common, particularly among better-educated workers.

This paper has provided a first, descriptive assessment of the patterns of mobility within and across informal and formal employment in the sub-Saharan African context. Future research should further investigate and test the dynamic mechanisms at play. Summarizing, two main findings emerge from this paper that have important policy implications. First, we find evidence of significant heterogeneity within the informal economy in sub-Saharan Africa, with clear differences in characteristics between lower-tier and upper-tier informal workers, whether in wage work or self-employment, as well as significant earnings gains for workers who make the transition from lower-tier employment to upper-tier employment. In addition, we find strong evidence of segmentation between wage employment and self-employment in the informal economy. Policymakers need to recognize this heterogeneity in informal work and devise policies that are not necessarily a “one size fits all” approach to the informal economy. Second, given the limited alternative job opportunities available, particularly to those in lower-tier informal self-employment, our findings suggest that specific policy measures that seek to enhance the livelihoods of workers in this most disadvantaged segment would be more relevant in the sub-Saharan context, as compared to policies that aim to reduce the regulatory barriers to formalization.

#### **Declarations**

##### **Availability of data and material**

The datasets analyzed in this study are available from the following sources:

ECG-ISSER. Ghana Socioeconomic Panel Study Survey (GSPS) 2009–2010, Wave 1 [dataset]. Version 01. Legon: ISSER, University of Ghana [producer]. New Haven, CT: ECG [producer/funding agency]. Washington, DC: World Bank [distributor]. <https://microdata.worldbank.org/index.php/catalog/2534>.

ECG-ISSER. Ghana Socioeconomic Panel Study Survey (GSPS) 2013–2014, Wave 2 [dataset]. Version 01. Legon: ISSER, University of Ghana [producer/distributor]. New Haven, CT: ECG [producer/funding agency]. Data are available on request (email Andre Nickow, a-nickow@kellogg.northwestern.edu).

NBS. Tanzania National Panel Survey (TZNPS) 2010–2011, Wave 2 [dataset]. Version 03. Dar es Salaam: NBS, Government of Tanzania [producer]. Washington, DC: World Bank [producer/funding agency/distributor]. <https://microdata.worldbank.org/index.php/catalog/1050>.

NBS. Tanzania National Panel Survey (TZNPS) 2012–2012, Wave 3 [dataset]. Version 01. Dar es Salaam: NBS, Government of Tanzania. [producer]. Washington, DC: World Bank [producer/funding agency/distributor]. <https://microdata.worldbank.org/index.php/catalog/2252>.

SALDRU. National Income Dynamics Study (NIDS) 2014–2015, Wave 4 [dataset]. Version 2.0.0. Pretoria: DPME [funding agency]. Cape Town: SALDRU [implementer]. Cape Town: DataFirst [distributor]. <https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/570>.

SALDRU. National Income Dynamics Study (NIDS) 2017, Wave 5 [dataset]. Version 1.0.0 Pretoria: DPME [funding agency]. Cape Town: SALDRU [implementer], 2018. Cape Town: DataFirst [distributor]. <https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/712>.

UBOS. Uganda National Panel Survey (UNPS) 2010–2011, Wave 2 [dataset]. Version 01. Kampala: UBOS, Government of Uganda [producer]. Washington, DC: World Bank [funding agency/distributor]. <https://microdata.worldbank.org/index.php/catalog/2166>.

UBOS. Uganda National Panel Survey (UNPS) 2011–2012, Wave 3 [dataset]. Version 01. Kampala: UBOS, Government of Uganda [producer]. Washington, DC: World Bank [funding agency/distributor]. <https://microdata.worldbank.org/index.php/catalog/2059>.

#### **Competing interests**

The authors declare that they have no competing interests.

#### **Funding**

This paper was prepared as part of the work under the UNU-WIDER project “Transforming informal work and livelihoods”.

The Institute is funded through income from an endowment fund with additional contributions to its work program from Finland, Sweden, and the United Kingdom as well as earmarked contributions for specific proj-

ects from a variety of donors. There has been no significant financial support for this work that could have influenced its outcome.

The findings, interpretations, and conclusions expressed in the paper are those of the authors and do not necessarily reflect the views of the Institute or the United Nations University, or the program/project donors.

### Authors' contributions

MD and KS were jointly responsible for the project administration and supervision. MD and SS contributed to the data curation. SS developed and performed the statistical analysis and was a major contributor in writing the manuscript. All authors contributed to the conceptual framework and the main argument underpinning the analysis and the interpretations of findings. All authors read, edited, and approved the final manuscript.

### Acknowledgments

The authors are grateful for valuable feedback and comments on the paper from Gary Fields and Tim Gindling as well as from participants at the UNU-WIDER Summer School and the fourth annual conference on "Inclusive Growth in Mozambique". The authors gratefully acknowledge Risto Rönkkö for excellent research assistance.

### References

- Aryeetey, E.; I. Osei-Akoto; R. Darko-Osei; C. Udry (2011): Ghana Socioeconomic Panel Survey: Report of the Baseline Survey. Yale/ISSER Technical Report. New Haven, CT and Accra: Yale University and Institute of Statistical, Social and Economic Research.
- Banerjee, A.V.; E. Duflo (2007): The Economic Lives of the Poor. *Journal of Economic Perspectives* 21(1), 141-168.
- Banerjee, S.; R.S.N. Raj; K. Sen (2016): Monitoring Costs, Credit Constraints and Entrepreneurship. *Manchester School* 84(5), 573-599.
- Basu, A.K.; N.H. Chau; G.S. Fields; R. Kanbur (2018): Job Creation in a Multi-sector Labour Market Model for Developing Economies. *Oxford Economic Papers* 71(1), 119-144.
- Beusch, E.; A. van Soest (2019): A Dynamic Multinomial Model of Self-employment in the Netherlands. [https://www.instituutgak.nl/wp-content/uploads/2019/03/2014-528\\_P01.pdf](https://www.instituutgak.nl/wp-content/uploads/2019/03/2014-528_P01.pdf). Accessed 13 May 2020.
- Bosch, M.; W. Maloney (2010): Comparative Analysis of Labor Market Dynamics Using Markov Processes: An Application to Informality. *Labour Economics* 17(4), 621-631.
- Buddelmeyer, H.; M. Wooden (2011): Transitions Out of Casual Employment: The Australian Experience. *Industrial Relations: A Journal of Economy and Society* 50(1), 109-130.
- Cappellari, L.; S.P. Jenkins (2004): Modelling Low Income Transitions. *Journal of Applied Econometrics* 19(5), 593-610.
- Chen, M.A. (2012): The Informal Economy: Definitions, Theories and Policies. WIEGO Working Paper 1. Cambridge, MA: Women in Informal Employment Globalizing and Organizing.
- Commander, S.; N. Isachenkova; Y. Rodionova (2013): Informal Employment Dynamics in Ukraine: An Analytical Model of Informality in Transition Economies. *International Labour Review* 152(3-4), 445-467.
- De Vreyer, P.; F. Roubaud (eds.) (2013): Urban Labor Markets in sub-Saharan Africa. Washington, DC: World Bank.
- ECG-ISSER (2016): 'Ghana Socioeconomic Panel Study Survey (GSPS) 2009–2010, Wave 1 [dataset]'. Version 01. Legon: ISSER, University of Ghana [producer]. New Haven, CT: ECG [producer/funding agency]. Washington, DC: World Bank [distributor].
- ECG-ISSER (2019): 'Ghana Socioeconomic Panel Study Survey (GSPS) 2013–2014, Wave 2 [dataset]'. Version 01. Legon: ISSER, University of Ghana [producer/distributor]. New Haven, CT: ECG [producer/funding agency].
- Echevin, D.; F. Murtin (2009): What Determines Productivity in Senegal? Sectoral Disparities and the Dual Labour Market. *The Journal of Development Studies* 45(10), 1707-1730.
- Fields, G.S. (1990): Labour Market Modelling and the Urban Informal Sector: Theory and Evidence'. [Electronic Version], in: Turnham, D.; B. Salomé; A. Schwarz (eds.), *The Informal Sector Revisited*. Paris: Organisation for Economic Co-operation and Development, 49-69.
- Fields, G.S. (2005): A Guide to Multisector Labor Market Models. Social Protection Discussion Paper Series 0505. Washington, DC: World Bank.
- Fields, G.S. (2019): *Employment and Development: How Work Can Lead from and into Poverty*. Oxford: Oxford University Press.
- Gindling, T.H.; D. Newhouse (2014): Self-Employment in the Developing World. *World Development* 56, 313-331.



- Gong, X., A. Van Soest; E. Villagomez (2004): Mobility in the Urban Labor Market: A Panel Data Analysis for Mexico. *Economic Development and Cultural Change* 53(1), 1-36.
- Grimm, M.; J. Krüger; J. Lay (2011): Barriers to Entry and Returns to Capital in Informal Activities: Evidence from West Africa. *Review of Income and Wealth* 57, S27-1352-68.
- Grimm, M.; P. Knorringer; J. Lay (2012): Constrained Gazelles: High Potentials in West Africa's Informal Economy. *World Development* 40(7), 1352-1368.
- Günther, I.; A. Launov (2012): Informal Employment in Developing Countries: Opportunity or Last Resort? *Journal of Development Economics* 97(1), 88-98.
- Harris, J. R.,; Todaro, M. P. (1970): Migration, unemployment and development: a two-sector analysis. *The American economic review*, 126-142.
- Hayami, Y.; A.K. Dikshit; S.N. Mishra (2006): Waste Pickers and Collectors in Delhi: Poverty and Environment in an Urban Informal Sector. *The Journal of Development Studies* 42(1), 41-69.
- Heckman, J. (1981): Statistical Models for Discrete Panel Data. In: Manski C.; D. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*. London: MIT Press, 114-178.
- Heckman, J. (1991): Identifying the Hand of the Past: Distinguishing State Dependence from Heterogeneity. *American Economic Review* 81(2), 75-79.
- ILO (2018): *Women and Men in the Informal Economy: A Statistical Picture, Third Edn.* Geneva: International Labour Office.
- ILO (2021): ILOSTAT Database. <https://ilostat.ilo.org/data/>. Accessed 9 April 2021.
- Kanbur, R. (2017): Informality: Causes, Consequences and Policy Responses. *Review of Development Economics* 21(4), 939-961.
- La Porta, R.; A. Shleifer (2014): Informality and Development. *Journal of Economic Perspectives* 28(3), 109-126.
- Lee, K.D.; S.H. Lee; J.I. Choe (2018): State Dependence, Individual Heterogeneity, and the Choice of Employment Status: Evidence from Korea. *Applied Economics* 50(8), 824-837.
- Lehmann, H. (2015): Informal Employment in Transition Countries: Empirical Evidence and Research Challenges. *Comparative Economic Studies* 57(1), 1-30.
- Lehmann, H.; N. Pignatti (2018): Informal Employment Relationships and the Labor Market: Is there Segmentation in Ukraine? *Journal of Comparative Economics* 46(3), 838-857.
- Lehmann, H.; A. Muravyev; K.F. Zimmermann (2012): The Ukrainian Longitudinal Monitoring Survey: Towards A Better Understanding of Labor Markets in Transition. *IZA Journal of Labor & Development* 1(1), 1-15.
- Liu, X. (2015): *Methods and Applications of Longitudinal Data Analysis*. Amsterdam: Elsevier.
- Maloney, W.F. (1999): Does Informality Imply Segmentation in Urban Labor Markets? Evidence from Sectoral Transitions in Mexico. *World Bank Economic Review* 13, 275-302.
- Maloney, W.F. (2004): Informality Revisited. *World Development* 32(7), 1159-1178.
- McKay, A.; J. Pirttilä; C. Schimanski (2018): The Elasticity of Formal Work in African Countries. Tampere Economic Working Paper 120. Tampere: Tampere University.
- NBS (2012): 'Tanzania National Panel Survey (TZNPS) 2010-2011, Wave 2 [dataset]'. Version 03. Dar es Salaam: NBS, Government of Tanzania [producer]. Washington, DC: World Bank [producer/funding agency/distributor].
- NBS (2015): 'Tanzania National Panel Survey (TZNPS) 2012-2012, Wave 3 [dataset]'. Version 01. Dar es Salaam: NBS, Government of Tanzania. [producer]. Washington, DC: World Bank [producer/funding agency/distributor].
- Oguzoglu, U. (2016): Disability and Multi-State Labour Force Choices with State Dependence. *Economic Record* 92(296), 28-46.
- Perry, G.; W. Maloney; O. Arias; P. Fajnzylber; A. Mason; J. Saavedra-Chanduvi (2007): *Informality: Exit and Exclusion*. Washington, DC: World Bank.
- Prowse, V. (2012): Modeling Employment Dynamics with State Dependence and Unobserved Heterogeneity. *Journal of Business & Economic Statistics* 30(3), 411-431.
- Radchenko, N. (2017): Informal Employment in Developing Economies: Multiple Heterogeneity. *The Journal of Development Studies* 53(4), 495-513.
- Rodrik, D. (2016): Premature Deindustrialization. *Journal of Economic Growth* 21(1), 1-33.
- SALDRU (2018a): 'National Income Dynamics Study (NIDS) 2014-2015, Wave 4 [dataset]'. Version 2.0.0. Pretoria: DPME [funding agency]. Cape Town: SALDRU [implementer]. Cape Town: DataFirst [distributor].
- SALDRU (2018b): National Income Dynamics Study (NIDS) 2017, Wave 5 [dataset]. Version 1.0.0 Pretoria: DPME [funding agency]. Cape Town: SALDRU [implementer], 2018. Cape Town: DataFirst [distributor].
- Sarkar, S.; S. Sahoo; S. Klasen, (2019): Employment Transitions of Women in India: A Panel Analysis. *World Development* 115, 291-309.

- Schotte, S.; R. Zizzamia; M. Leibbrandt (2018): A Poverty Dynamics Approach to Social Stratification: The South African Case. *World Development* 110, 88-103.
- Slonimczyk, F.; V. Gimpelson (2015): Informality and Mobility: Evidence from Russian Panel Data. *Economics of Transition* 23(2), 299-341.
- Teorell, J.; A. Sundström; S. Holmberg; B. Rothstein; P.N. Alvarado; C.M. Dalli. (2021): The Quality of Government Standard Dataset, version Jan 21. University of Gothenburg: The Quality of Government Institute. <https://www.gu.se/en/quality-government/qog-data/data-downloads/standard-dataset>. Accessed 15 April 2021.
- UBOS (2014a): Uganda National Panel Survey (UNPS) 2010-2011, Wave 2 [dataset]. Version 01. Kampala: UBOS, Government of Uganda [producer]. Washington, DC: World Bank [funding agency/distributor].
- UBOS (2014b): Uganda National Panel Survey (UNPS) 2011-2012, Wave 3 [dataset]. Version 01. Kampala: UBOS, Government of Uganda [producer]. Washington, DC: World Bank [funding agency/distributor].
- Vella, F. (1998): Estimating Models with Sample Selection Bias: A Survey. *Journal of Human Resources* 33(1), 127-169.
- World Bank. (2021): World Development Indicators 2021. Washington, DC: The World Bank. <https://databank.worldbank.org/source/world-development-indicators>. Accessed 5 April 2021.
- World Economic Forum. (2019): The Global Competitiveness Report 2019. <http://reports.weforum.org/global-competitiveness-report-2019>. Accessed 19 April 2021.

# Appendices

## Appendix A: Data Sources

This study is based on household survey data collected in four sub-Saharan African countries: Ghana, South Africa, Tanzania, and Uganda. For reasons of data availability and cross-country comparability, we focus the analysis on two waves of panel data in each of the four countries under study (listed in Table A1). Detailed study descriptions are provided below.

### Ghana: GSPS

The GSPS is a joint effort between the Economic Growth Centre (ECG) at Yale University and the Institute of Statistical, Social and Economic Research (ISSER) at the University of Ghana. Our study uses the first and second waves of the survey conducted in 2009/10 (GSPS I) and 2013/14 (GSPS II), respectively (ECG-ISSER 2016, 2019). The data from the first wave of the GSPS consists of a nationally representative sample of 5,010 households in 334 enumeration areas (EAs) containing 18,889 household members. The second wave covered a sample of 4,774 households containing 16,356 household members. The second wave also tracked the movement of households as well as an individual within a household. The GSPS collected data on the demographic characteristics of households, education, health, employment, migration, land information, agricultural production input, livestock and household tools, non-farm enterprise, housing characteristics of the household, financial assets, psychological measures, risk preference, and social status and responsibilities.

The GSPS data are regionally representative of the ten regions of Ghana. A two-stage stratified sample design was used for the survey. The first stage involved selecting geographical clusters from an updated master sampling frame constructed from the 2000 Ghana Population and Housing Census. In all, 334 clusters were selected from the list of EAs in each region, based on a simple random sampling technique. A complete household listing was carried out in 2009 in all the selected clusters to produce a sampling frame for the second-stage selection of households. The second stage of selection involved a simple random sampling of 15 of the listed households from each selected cluster (see Aryeetey et al., 2011).

**Table A1** Data sources

Country	Survey	t = 0	t = 1
Ghana	GSPS	Wave I (2009/10)	Wave II (2013/14)
South Africa	NIDS	Wave IV (2014/15)	Wave V (2017)
Tanzania	TZNPS	Wave II (2010/11)	Wave III (2012/13)
Uganda	UNPS	Wave II (2010/11)	Wave III (2011/12)

Source: Authors' own construct.

GSPS, Ghana Socioeconomic Panel Study Survey; NIDS, National Income Dynamics Study; TZNPS, Tanzania National Panel Study; UNPS, Uganda National Panel Study.

### **South Africa: NIDS**

The South African NIDS is funded by the Department of Planning, Monitoring and Evaluation (DPME) of the Government of South Africa and implemented by the Southern Africa Labour and Development Research Unit (SALDRU) at the University of Cape Town. NIDS started in 2008 with a nationally representative sample of over 28,000 individuals in 7,300 households. At present, there are five waves of data available, each of which is spaced approximately 2 years apart.

As the first national panel study of individuals in South Africa, NIDS provides a rich source of individual-level information on demographic characteristics, labor market participation, the nature of employment, working conditions, and earnings, along with information on other socioeconomic characteristics both at the individual and at the household levels. It attempts to track and re-interview respondents as they move out of their original households.

Importantly, the question to those in self-employment asking how many employees there are at the respondent's place of work, excluding themselves has only been included in the fourth survey wave. For this reason, we focus our analysis on the two last waves, collected in 2014/15 (NIDS IV) and 2017 (NIDS V), respectively (SALDRU, 2018a, 2018b).

### **Tanzania: TZNPS**

The TZNPS is a nationally representative household panel survey that collects information on a wide range of topics including household composition and characteristics, individual demographic characteristics, agricultural production, non-farm income-generating activities, and other socioeconomic activities. The data are collected by the Tanzania National Bureau of Statistics (NBS) following a stratified two-stage design (51 design strata corresponding to a rural/urban designation for each of the 26 regions, except for Dar es Salaam, which is purely urban and therefore constitutes only one stratum).

The TZNPS sample for the first round conducted in 2008/09 (TZNPS I) covered 3,265 households. The second and third rounds of the survey were conducted in 2010/11 (TZNPS II) and 2012/13 (TZNPS III), respectively. All original households were targeted for a revisit, including split-off tracking of adults who had relocated to a new location. In the fourth round, collected in 2014/15 (TZNPS IV), the sample was refreshed. Since several questions relevant for this study were only introduced in TZNPS II and most households were rotated out between TZNPS III and TZNPS IV, we focus our analysis on TZNPS II + III (NBS, 2012, 2015).

In Tanzania, the two questions that strictly identify informality have only been added in TZNPS III. Therefore, we use multiple correspondence analysis to create a formality index (separately for wage and self-employment), which draws on variables commonly associated with formality (details on the methodology are available from the authors upon request). Workers are ranked by their index score and the cut-off value (used to differentiate formal/informal work) is chosen to replicate the formality share observed in TZNPS III using the reference definition of formality. For consistency, we use the proxy definition in both TZNPS II and III.

### **Uganda: UNPS**

The UNPS, implemented by the Uganda Bureau of Statistics (UBOS), is representative at the national and regional levels. The baseline sample was collected in 2009/10 (UNPS I), covering 3,123 households that were distributed over 322 EAs. Within each stratum (rural/urban), the UNPS EAs were selected out of the 783 EAs that had been visited by the Uganda National Household Survey in 2005/06. This initial sample was re-visited in 2010/11 (UNPS II) and 2011/12 (UNPS III), where households or individuals that had permanently left the original households to known locations were tracked and interviewed. After this, in 2013/14 (UNPS IV), one-third of the original EAs were permanently rotated out of the sample and 100 new EAs were introduced (extracted from the updated sample frames developed by UBOS from the 2012 Census), resulting in a final sample of 3,119 households that were re-interviewed in 2015/16 (UNPS V). Since the panel was partly refreshed between UNPS III-IV and data for UNPS V has not yet been made publicly available, we focus on UNPS II + III (UBOS, 2014a, 2014b).

## Appendix B: Complementary Tables

**Table B1** Basic socioeconomic, labor market and institutional indicators, 2010–2017

	Ghana	South Africa	Tanzania	Uganda
<b>GDP per capita</b>				
Annual (constant 2017 PPP \$)	4,505	12,731	2,256	2,004
Growth rate (annual %)	4.4	0.4	3.4	1.7
<b>Sectoral contribution to GDP (% of GDP)</b>				
Agriculture	21.9	2.2	26.6	26.2
Industry	28.6	26.6	25.1	25.5
Services	42.8	61.1	40.6	42.0
<b>Poverty headcount (% of population)</b>				
\$1.90 a day (2011 PPP)	12.0	17.5	49.5	38.5
\$3.20 a day (2011 PPP)	30.1	36.5	78.2	68.5
<b>Educational attainment (% of population 25+ years) (cumulative)</b>				
At least completed primary	64.7	81.4	64.8	38.4
At least completed lower secondary	54.3	74.5	11.0	26.4
At least completed upper secondary	20.6	61.5	3.4	10.3
<b>Sector employment (% of total) (modeled ILO estimate)</b>				
Agriculture	41.6	5.0	68.3	70.1
Industry	16.2	23.7	6.1	7.2
Services	42.2	71.3	25.5	22.6
<b>Public sector employment</b>				
Public sector employment as a share of formal employment (%)	45.4	22.7	49.8	39.5
<b>Union membership</b>				
Share of private paid employees with union membership (%)	17.5	20.4	9.4	3.0
<b>Ease of doing business</b>				
Score (0 = lowest performance to 100 = best performance)	57.8	65.7	52.5	57.3
Number of start-up procedures to register a business	7.6	6.5	11.3	14.0
Days required to start a business	12.9	45.3	29.5	26.8
New business density (registrations per 1,000 people ages 15–64)	1.0	7.2	0.3	0.9
<b>Quality of institutions</b>				
Governance quality index (1–10)	6.5	6.1	5.2	5.7
Regulatory performance in paying taxes (1–100)	66.3	81.2	53.8	73.0
Workers' Rights (1–100)	79.0	86.0	71.0	70.0

*Note:* Calculated averages for the 2010–2017 study period. The reported public sector employment as a share of formal employment is in 2012 for Ghana, 2014 for South Africa, 2012 for Tanzania, and 2016 for Uganda.

*Source:* Authors' calculations based on data from World Bank (2021), ILO (2021), and Teorell et al. (2021).

GDP, gross domestic product; PPP, Purchasing power parity.

**Table B2** Distribution of workers by work status (%), extended definition

		Ghana	South Africa	Tanzania	Uganda	
a) Proportion of employment by work status (incl. family farms and unemployment)						
Wage employed	Formal	6.8	44.3	3.9	4.0	
	Informal	Upper	2.8	7.0	1.1	3.7
		Lower	9.8	16.7	9.5	8.8
Self-employed	Formal	4.6	3.1	3.0	1.1	
	Informal	Upper	6.2	4.2	1.3	2.0
		Lower	21.7	2.9	14.4	14.5
Family farms		43.4	0.9	65.5	65.0	
Unemployment		4.8	21.0	1.3	1.0	
TOTAL		100	100	100	100	
b) Proportion of formal vs. informal employment (incl. family farms as informal lower)						
Formal		12.0	59.9	7.0	5.1	
Informal	Upper	9.4	14.2	2.4	5.7	
	Lower	78.6	25.9	90.6	89.2	
TOTAL		100	100	100	100	

Note: For each country, summary statistics are compiled for the initial wave of panel study under study.

Source: Authors' own calculations based on survey data from GSPS 2009/10, NIDS 2014/15, TZNPS 2010/11, and UNPS 2010/11.

GSPS, Ghana Socioeconomic Panel Study Survey; NIDS, National Income Dynamics Study; TZNPS, Tanzania National Panel Study; UNPS, Uganda National Panel Study.

**Table B3** Change in distribution of workers by work status, balanced panel

		Ghana	South Africa	Tanzania	Uganda	
a) Change (ppts) in proportion of employment by work status						
Wage employed	Formal	1.7	-0.4	5.0	-2.7	
	Informal	Upper	-2.5	1.6	-2.4	0.0
		Lower	-3.8	-3.3	-0.4	0.7
Self-employed	Formal	0.8	2.0	-2.8	1.3	
	Informal	Upper	-1.5	-0.5	0.7	0.5
		Lower	5.3	0.6	-0.1	0.3
b) Change (ppts) in proportion of formal and informal employment						
Formal		2.5	1.5	2.3	-1.5	
Informal	Upper	-4.0	1.1	-1.7	0.5	
	Lower	1.5	-2.6	-0.5	1.0	
c) Change (ppts) in proportion of self-employment						
Self-employed		4.6	2.1	-2.2	2.1	

Source: Authors' own calculations based on survey data from GSPS 2009/10-2013/14, NIDS 2014/15-2017, TZNPS 2010/11-2012/13, and UNPS 2010/11-2011/12.

GSPS, Ghana Socioeconomic Panel Study Survey; NIDS, National Income Dynamics Study; TZNPS, Tanzania National Panel Study; UNPS, Uganda National Panel Study.

**Table B4** Average worker characteristics by work status

		Ghana	South Africa	Tanzania	Uganda	
<b>a) Average age (years)</b>						
Wage employed	Formal	43.5 (0.53)	37.5 (0.16)	40.5 (0.52)	36.1 (0.71)	
	Informal	Upper	36.4 (0.82)	35.6 (0.37)	33.4 (1.00)	34.7 (0.84)
		Lower	38.4 (0.46)	35.4 (0.25)	30.3 (0.35)	30.5 (0.53)
Self-employed	Formal	41.5 (0.64)	41.3 (0.72)	37.5 (0.54)	34.8 (1.29)	
	Informal	Upper	40.3 (0.53)	40.2 (0.54)	34.7 (0.83)	38.8 (0.95)
		Lower	40.8 (0.29)	38.6 (0.69)	31.0 (0.33)	34.8 (0.47)
TOTAL		40.5 (0.19)	37.2 (0.12)	32.7 (0.20)	34.1 (0.29)	
<b>b) Share of female workers (%)</b>						
Wage employed	Formal	33.7 (2.41)	41.5 (0.75)	32.5 (2.15)	41.3 (3.77)	
	Informal	Upper	30.0 (3.55)	57.0 (1.65)	35.5 (4.33)	34.3 (3.80)
		Lower	33.8 (1.91)	44.3 (1.09)	34.0 (1.55)	26.8 (2.12)
Self-employed	Formal	47.9 (3.24)	45.8 (3.00)	36.0 (2.83)	39.7 (6.60)	
	Informal	Upper	50.0 (2.67)	40.3 (2.25)	25.4 (3.73)	29.4 (4.25)
		Lower	81.7 (1.08)	58.4 (2.96)	63.9 (1.31)	47.9 (1.85)
TOTAL		56.8 (0.90)	44.2 (0.55)	46.7 (0.87)	38.9 (1.19)	
<b>c) Share of workers with secondary or tertiary education (%)</b>						
Wage employed	Formal	52.2 (3.02)	63.7 (0.74)	36.4 (2.29)	60.1 (3.88)	
	Informal	Upper	37.8 (4.04)	55.5 (1.66)	39.9 (4.76)	59.7 (4.09)
		Lower	19.1 (1.76)	28.0 (0.98)	1.2 (0.36)	11.8 (1.66)
Self-employed	Formal	18.6 (2.80)	69.3 (2.78)	2.4 (0.91)	42.9 (6.74)	
	Informal	Upper	10.0 (1.82)	28.6 (2.07)	3.3 (1.56)	37.3 (4.59)
		Lower	8.3 (0.95)	33.0 (2.82)	1.7 (0.38)	12.2 (1.29)
TOTAL		19.0 (0.83)	52.5 (0.55)	7.1 (0.47)	25.8 (1.13)	

Note: For each country, summary statistics are compiled for the initial wave of panel study under study. Standard errors of mean values in parentheses.

Source: Authors' own calculations based on survey data from GSPS 2009/10, NIDS 2014/15, TZNPS 2010/11, and UNPS 2010/11.

GSPS, Ghana Socioeconomic Panel Study Survey; NIDS, National Income Dynamics Study; TZNPS, Tanzania National Panel Study; UNPS, Uganda National Panel Study.



**Table B5** Transition matrices across work status groups, including additional destination states (family farms under lower-tier informal self-employment and unemployment)

			WAVE t = 1						Unempl.	Share of stayers	
			Wage employed			Self-employed					
			Formal	Informal		Formal	Informal				
				Upper	Lower		Upper	Lower			
a) Ghana											
WAVE t = 0	Wage employed	Formal		59.0	6.2	8.7	3.9	3.2	16.7	2.2	7.8
		Informal	Upper	25.7	9.2	18.5	2.0	2.5	40.5	1.6	0.5
			Lower	15.5	3.8	22.9	5.4	4.5	41.2	6.8	4.3
	Self-employed	Formal		1.1	1.7	6.2	14.9	8.0	64.2	4.0	1.3
		Informal	Upper	2.2	1.3	8.2	8.0	24.2	52.4	3.8	2.9
			Lower	0.8	0.8	6.5	2.8	2.6	83.5	3.1	34.9
TOTAL			11.9	2.6	10.5	5.2	6.4	59.7	3.7	51.7	
b) South Africa											
WAVE t = 0	Wage employed	Formal		78.9	6.7	5.5	2.1	1.0	0.7	5.1	44.7
		Informal	Upper	42.0	21.0	12.4	5.0	0.9	2.7	16.1	1.9
			Lower	22.0	10.9	39.8	3.9	3.2	3.9	16.2	8.5
	Self-employed	Formal		10.8	8.0	4.2	42.0	11.0	8.5	15.5	1.7
		Informal	Upper	9.9	9.0	15.8	13.1	19.1	14.3	18.8	1.0
			Lower	10.0	4.2	11.4	4.7	16.7	21.2	31.9	0.8
TOTAL			56.9	8.9	13.4	4.7	3.1	3.0	10.1	58.6	
c) Tanzania											
WAVE t = 0	Wage employed	Formal		71.9	4.0	8.8	1.4	0.9	12.7	0.4	8.4
		Informal	Upper	40.5	5.1	29.2	5.7	2.5	15.9	1.1	0.2
			Lower	11.1	0.7	46.7	2.2	2.6	35.9	0.8	13.4
	Self-employed	Formal		1.6	0.0	4.1	25.8	14.8	53.6	0.0	2.3
		Informal	Upper	3.8	0.0	18.5	12.9	19.6	45.2	0.0	0.8
			Lower	3.8	1.0	9.9	5.4	2.7	74.8	2.5	32.5
TOTAL			16.5	1.3	20.4	6.4	4.5	49.6	1.3	57.5	
d) Uganda											
WAVE t = 0	Wage employed	Formal		57.7	26.8	8.1	5.2	0.0	2.2	0.0	6.8
		Informal	Upper	17.8	43.1	12.6	5.1	5.5	16.0	0.0	4.7
			Lower	3.2	4.2	46.1	1.8	2.9	41.4	0.5	12.0
	Self-employed	Formal		2.6	0.0	10.1	19.8	14.3	53.3	0.0	0.6
		Informal	Upper	0.0	3.7	4.0	11.2	32.1	49.1	0.0	1.8
			Lower	1.2	0.5	6.3	2.3	3.8	85.4	0.5	36.4
TOTAL			10.2	9.1	16.7	3.9	5.4	54.3	0.3	62.3	

Note: Each row indicates work status in the base period, and each column in transition matrices indicates work status in the next period; transition matrix rows sum to 100. The likelihood of staying in the same employment status conditional on the base year employment status is highlighted in gray. The share of stayers (proportion of workers who remain in their work status) is calculated as the product of highlighted diagonals and initial size.

Source: Authors' own calculations based on survey data from GSPS 2009/10-2013/14, NIDS 2014/15-2017, TZNPS 2010/11-2012/12, and UNPS 2010/11-2011/12.

GSPS, Ghana Socioeconomic Panel Study Survey; NIDS, National Income Dynamics Study; TZNPS, Tanzania National Panel Study; UNPS, Uganda National Panel Study.

**Table B6** (Non-farm) employment and panel retention rates (%)

	Ghana		South Africa		Tanzania		Uganda	
	Employed (non-farm)	Panel retention	Employed (non-farm)	Panel retention	Employed (non-farm)	Panel retention	Employed (non-farm)	Panel retention
Percentage share	51.7	70.8	78.1	75.9	33.2	69.1	34.0	71.9

*Note:* The dummy variable capturing (non-farm) employment equals one if the individual is working in formal or informal wage or self-employment and zero if working on a family farm or is unemployed. For individuals who responded to the employment module in  $t = 0$ , the dummy variable capturing panel retention equals one if the individual was successfully re-interviewed in the employment module in  $t = 1$ , and zero otherwise.

*Source:* Authors' own calculations based on survey data from GSPS 2009/10-2013/14, NIDS 2014/15-2017, TZNPS 2010/11-2012/13, and UNPS 2010/11-2011/12.

GSPS, Ghana Socioeconomic Panel Study Survey; NIDS, National Income Dynamics Study; TZNPS, Tanzania National Panel Study; UNPS, Uganda National Panel Study.

**Table B7** Probability of employment (non-farm) in the baseline and panel retention

Variables	Ghana		South Africa		Tanzania		Uganda	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
	Employed (non-farm)	Panel retention	Employed (non-farm)	Panel retention	Employed (non-farm)	Panel retention	Employed (non-farm)	Panel retention
Level of education (base: no schooling)								
Primary	0.164*	-0.044	-0.040	0.013	0.091***	0.014	-0.052**	0.004
	(0.086)	(0.080)	(0.025)	(0.024)	(0.013)	(0.013)	(0.022)	(0.020)
Post-primary	0.216**	-0.034	-0.047*	0.013	0.332***	0.014	0.108***	-0.033
	(0.085)	(0.079)	(0.024)	(0.023)	(0.019)	(0.018)	(0.029)	(0.026)
Secondary	0.328***	-0.044	0.001	0.010	0.602***	0.127***	0.089**	-0.030
	(0.088)	(0.082)	(0.024)	(0.024)	(0.038)	(0.032)	(0.035)	(0.031)
Post-secondary	0.506***	-0.018	0.069***	-0.013	0.556***	0.024	0.356***	-0.041
	(0.093)	(0.090)	(0.025)	(0.025)	(0.045)	(0.045)	(0.040)	(0.037)
Tertiary	0.521***	-0.081	0.160***	-0.022	0.593***	0.145***	0.124**	-0.111*
	(0.091)	(0.089)	(0.028)	(0.032)	(0.063)	(0.040)	(0.063)	(0.059)
Age	0.033***	0.052***	0.026***	0.010***	0.004*	0.031***	0.010***	0.027***
	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Age squared	-0.042***	-0.064***	-0.024***	-0.013***	-0.010***	-0.039***	-0.016***	-0.030***
(×0.01)	(0.005)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Female	0.189***	-0.015	-0.089***	0.061***	-0.001	-0.074***	-0.115***	0.019
	(0.017)	(0.014)	(0.008)	(0.008)	(0.011)	(0.010)	(0.015)	(0.013)
Urban	0.264***	-0.032**	0.113***	-0.011	0.375***	-0.092***	0.340***	-0.050***
	(0.012)	(0.015)	(0.008)	(0.008)	(0.007)	(0.011)	(0.015)	(0.016)
Head of household	0.268***		0.106***		0.108***		0.140***	
	(0.018)		(0.008)		(0.012)		(0.017)	
Sample member in the previous wave		n.a.		0.234***		0.044***		0.073***
		n.a.		(0.008)		(0.012)		(0.016)
Wald test of rho = 0	15.216		10.086		17.313		9.811	
F-statistic (p-value)	0.000		0.002		0.000		0.002	
Observations	6,718		21,498		7,870		4,315	

Note: \*\*\*p < 0.01, \*\*p < 0.05, \*p < 0.1. Standard errors clustered at the country level in parentheses.

Source: Authors' own calculations based on survey data from GSPS 2009/10-2013/14, NIDS 2014/15-2017, TZNPS 2010/11-2012/13, and UNPS 2010/11-2011/12.

GSPS, Ghana Socioeconomic Panel Study Survey; NIDS, National Income Dynamics Study; TZNPS, Tanzania National Panel Study; UNPS, Uganda National Panel Study.

**Table B8** Excludability of instruments from main employment dynamics equation

	<b>Head of household</b>	<b>Sample member in the previous wave</b>
<b>F-statistic</b> (p-value)	0.1800	0.9647

*Source:* Authors' own calculations based on survey data from GSPS 2009/10-2013/14, NIDS 2014/15-2017, TZNPS 2010/11-2012/13, and UNPS 2010/11-2011/12.

GSPS, Ghana Socioeconomic Panel Study Survey; NIDS, National Income Dynamics Study; TZNPS, Tanzania National Panel Study; UNPS, Uganda National Panel Study.

**Table B9** Employment dynamics, including inactivity/unemployment as an additional destination state

Multinomial logistic regression		Number of obs = 9,851				
Average marginal effects on work status in $t = 1$		Log-likelihood = -12,048				
Base outcome: Lower-tier informal wage employed		Pseudo $R^2 = 0.284$				
Variables	(1)	(2)	(3)	(4)	(5)	(6)
	Formal wage employed	Upper-tier informal wage employed	Lower-tier informal wage employed	Formal self-employed	Upper-tier informal self-employed	Inactive/unemployed
Work status in $t = 0$ (base: lower-tier informal self-employed)						
(1) Formal wage employed	0.584*** (0.016)	0.051*** (0.009)	-0.002 (0.020)	-0.070*** (0.009)	-0.052*** (0.018)	-0.271*** (0.012)
(2) Upper-tier informal wage employed	0.264*** (0.011)	0.168*** (0.007)	0.084*** (0.019)	-0.048*** (0.009)	-0.038** (0.017)	-0.210*** (0.022)
(3) Lower-tier informal wage employed	0.130*** (0.008)	0.053*** (0.013)	0.228*** (0.023)	-0.054*** (0.010)	-0.028* (0.015)	-0.137*** (0.019)
(4) Formal self-employed	-0.006 (0.012)	0.006 (0.013)	-0.022* (0.012)	0.195*** (0.024)	0.051* (0.027)	-0.168*** (0.017)
(5) Upper-tier informal self-employed	0.008 (0.010)	0.023** (0.011)	0.008 (0.023)	0.027** (0.013)	0.151*** (0.044)	-0.101*** (0.038)
Level of education (base: no schooling)						
Primary	0.035** (0.016)	0.001 (0.007)	-0.065** (0.027)	0.018*** (0.005)	0.002 (0.004)	-0.003 (0.018)
Post-primary	0.093*** (0.017)	0.012 (0.020)	-0.113*** (0.032)	0.040*** (0.004)	0.006 (0.005)	-0.021 (0.019)
Secondary	0.157*** (0.024)	0.011 (0.017)	-0.157*** (0.033)	0.042*** (0.003)	0.001 (0.009)	-0.036** (0.015)
Post-secondary	0.219*** (0.028)	0.031* (0.019)	-0.204*** (0.035)	0.082*** (0.017)	0.005 (0.010)	-0.072*** (0.014)
Tertiary	0.329*** (0.029)	0.013 (0.014)	-0.249*** (0.040)	0.088*** (0.032)	-0.009 (0.018)	-0.099*** (0.023)
Age	0.015*** (0.004)	0.002 (0.001)	-0.001 (0.001)	0.006* (0.003)	0.002 (0.002)	-0.030*** (0.006)
Age squared ( $\times 0.01$ )	-0.018*** (0.004)	-0.003 (0.002)	-0.001 (0.001)	-0.006* (0.004)	-0.003 (0.002)	0.038*** (0.006)
Female	-0.050*** (0.010)	0.016*** (0.003)	-0.028* (0.016)	-0.023*** (0.003)	-0.019*** (0.001)	0.078*** (0.024)
Urban	0.012*** (0.004)	-0.022*** (0.005)	0.015 (0.016)	0.006 (0.007)	-0.009*** (0.004)	0.018 (0.013)

(Continued)

**Table B9** Continued

Variables	(1)	(2)	(3)	(4)	(5)	(6)
	Formal wage employed	Upper-tier informal wage employed	Lower-tier informal wage employed	Formal self-employed	Upper-tier informal self-employed	Inactive/unemployed
Country (base: Ghana)						
South Africa	0.085*** (0.007)	0.020*** (0.003)	-0.035*** (0.008)	-0.020*** (0.005)	-0.032*** (0.002)	0.107*** (0.006)
Tanzania	0.162*** (0.017)	-0.021*** (0.004)	0.030*** (0.006)	0.026*** (0.002)	-0.024*** (0.004)	-0.188*** (0.005)
Uganda	-0.053*** (0.003)	0.100*** (0.009)	0.042*** (0.007)	-0.009*** (0.002)	0.016*** (0.002)	-0.151*** (0.009)
Sample selection						
Panel retention	-0.037 (0.039)	-0.020* (0.012)	-0.033*** (0.012)	-0.012 (0.009)	-0.015* (0.008)	0.031 (0.030)
Employed (non-farm) in $t = 0$	-0.034 (0.033)	-0.009 (0.013)	-0.003 (0.022)	0.001 (0.013)	-0.022*** (0.007)	0.087 (0.056)

Note: \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ . Standard errors clustered at the country level in parentheses.

Source: Authors' own calculations based on survey data from GSPS 2009/10-2013/14, NIDS 2014/15-2017, TZNPS 2010/11-2012/13, and UNPS 2010/11-2011/12.

GSPS, Ghana Socioeconomic Panel Study Survey; NIDS, National Income Dynamics Study; TZNPS, Tanzania National Panel Study; UNPS, Uganda National Panel Study.

**Table B10** Employment dynamics, using ordered logistic regression

Ordered logistic regression, average marginal effects, split sample						
	(1)		(2)		(3)	
Formality status in $t = 0$	Lower-tier informal		Upper-tier informal		Formal	
Formality status in $t = 1$	Formal	Upper-tier informal	Formal	Lower-tier informal	Upper-tier informal	Lower-tier informal
Variables						
Employment status (base: Wage employed)						
Self-employed	-0.036 (0.026)	-0.013 (0.012)	-0.122** (0.057)	0.114** (0.055)	0.081*** (0.007)	0.185*** (0.034)
Level of education (base: no schooling)						
Primary	0.033 (0.020)	0.019** (0.008)	0.022 (0.021)	-0.033 (0.035)	-0.034* (0.020)	-0.125* (0.073)
Post-primary	0.121** (0.053)	0.056*** (0.011)	0.117 (0.080)	-0.145 (0.109)	-0.069*** (0.016)	-0.197*** (0.055)
Secondary	0.208*** (0.074)	0.077*** (0.005)	0.238*** (0.068)	-0.245*** (0.093)	-0.092*** (0.020)	-0.232*** (0.055)
Post-secondary	0.281*** (0.043)	0.084*** (0.004)	0.328*** (0.076)	-0.300*** (0.096)	-0.114*** (0.024)	-0.263*** (0.059)
Tertiary	0.575*** (0.064)	0.048** (0.023)	0.554*** (0.101)	-0.395*** (0.084)	-0.133*** (0.024)	-0.286*** (0.054)
Age	0.003 (0.004)	0.001 (0.001)	0.013 (0.010)	-0.011 (0.009)	-0.004** (0.002)	-0.007** (0.003)
Age squared ( $\times 0.01$ )	-0.002 (0.004)	-0.001 (0.001)	-0.014 (0.009)	0.012 (0.009)	0.004* (0.002)	0.006* (0.004)
Female	-0.066** (0.031)	-0.023*** (0.005)	-0.014 (0.048)	0.013 (0.044)	0.019*** (0.005)	0.034*** (0.009)
Urban	0.021* (0.012)	0.007 (0.006)	0.046** (0.019)	-0.042** (0.018)	0.001 (0.007)	0.003 (0.013)
Country (base: Ghana)						
South Africa	0.102*** (0.015)	0.036*** (0.009)	0.068*** (0.010)	-0.061*** (0.010)	-0.058*** (0.012)	-0.122*** (0.004)
Tanzania	0.038 (0.031)	0.016** (0.007)	0.027 (0.035)	-0.026 (0.033)	-0.049*** (0.004)	-0.106*** (0.008)
Uganda	-0.040*** (0.010)	-0.021* (0.012)	-0.080*** (0.014)	0.090*** (0.015)	0.013*** (0.001)	0.045*** (0.005)
Sample selection						
Panel retention	-0.144* (0.082)	-0.049*** (0.014)	-0.024 (0.044)	0.021 (0.040)	0.026* (0.014)	0.045 (0.030)
From $t = 0$ to $t = 1$						
Employed (non-farm) in $t = 0$	0.013 (0.023)	0.004 (0.009)	0.090*** (0.030)	-0.082*** (0.026)	0.025* (0.014)	0.043* (0.024)
Observations	2,869		1,166		3,781	
Pseudo R <sup>2</sup>	0.0728		0.0614		0.1220	

Note: \*\*\*p < 0.01, \*\*p < 0.05, \*p < 0.1. Standard errors clustered at the country level in parentheses. We group together workers in wage and self-employment and define the individual's formality status,  $F_i$ , on an ordinal scale, taking on three possible values: lower-tier informal employment,  $F_i = 1$ ; upper-tier informal employment  $F_i = 2$ ; and formal employment,  $F_i = 3$ . We split the sample by the formality status at time  $t = 0$  and express the formality status at time  $t = 1$  as a function of the same set of initial worker attributes  $X_{i0}$ .

Source: Authors' own calculations based on survey data from GSPS 2009/10-2013/14, NIDS 2014/15-2017, TZNPS 2010/11-2012/13, and UNPS 2010/11-2011/12.

GSPS, Ghana Socioeconomic Panel Study Survey; NIDS, National Income Dynamics Study; TZNPS, Tanzania National Panel Study; UNPS, Uganda National Panel Study.

**Table B11** Change in log labor earnings

Variables	All countries		Excl. South Africa	
	(1)	(2)	(3)	(4)
Log labor earnings ( $t = 0$ )	-0.335*** (0.057)	-0.334** (0.058)	-0.375* (0.123)	-0.375* (0.124)
Formality status in $t = 0$ (base: Lower informal in $t = 0$ )				
Formal ( $t = 0$ )	-0.033 (0.037)	-0.033 (0.035)	0.018 (0.045)	0.016 (0.044)
Upper informal ( $t = 0$ )	-0.023 (0.023)	-0.025 (0.024)	0.016 (0.031)	0.021 (0.030)
Formality status in $t = 0$ and $t = 1$ (base: Lower informal in $t = 0$ and $t = 1$ )				
Formal ( $t = 0$ ) × Formal ( $t = 1$ )	0.337** (0.081)	0.330** (0.077)	0.240 (0.101)	0.243 (0.110)
Formal ( $t = 0$ ) × Upper informal ( $t = 1$ )	0.136 (0.076)	0.133 (0.073)	0.014 (0.080)	0.013 (0.087)
Upper informal ( $t = 0$ ) × Formal ( $t = 1$ )	0.257*** (0.029)	0.256*** (0.030)	0.265** (0.035)	0.260** (0.045)
Upper informal ( $t = 0$ ) × Upper informal ( $t = 1$ )	0.071 (0.057)	0.070 (0.056)	0.109 (0.118)	0.099 (0.121)
Lower informal ( $t = 0$ ) × Formal ( $t = 0$ )	0.206*** (0.022)	0.206*** (0.021)	0.211* (0.054)	0.212* (0.054)
Lower informal ( $t = 0$ ) × Upper informal ( $t = 0$ )	0.082* (0.029)	0.095* (0.033)	0.067 (0.062)	0.069 (0.045)
Employment status in $t = 0$ (base: Wage employed in $t = 0$ )				
Self-employed ( $t = 0$ )		0.045 (0.106)		-0.091** (0.014)
Employment status in $t = 0$ and $t = 1$ (base: Wage employed in $t = 0$ and $t = 1$ )				
Self-employed ( $t = 0$ ) × Self-employed ( $t = 0$ )		-0.090 (0.054)		-0.021 (0.041)
Wage employed ( $t = 0$ ) × Self-employed ( $t = 0$ )		0.015 (0.104)		-0.067 (0.087)
Characteristics in $t = 0$				
Level of education (base: No schooling)				
Primary	0.058 (0.036)	0.061 (0.031)	0.076 (0.036)	0.076 (0.029)
Post-primary	0.103** (0.022)	0.106*** (0.018)	0.104* (0.033)	0.108* (0.032)
Secondary	0.205*** (0.035)	0.207*** (0.032)	0.212* (0.054)	0.219* (0.063)
Post-secondary	0.328*** (0.039)	0.332*** (0.037)	0.227** (0.045)	0.236** (0.055)
Tertiary	0.542*** (0.089)	0.544*** (0.088)	0.474* (0.143)	0.482* (0.154)

(Continued)



**Table B11** Continued

Variables	All countries		Excl. South Africa	
	(1)	(2)	(3)	(4)
Age	0.005 (0.004)	0.004 (0.003)	0.004 (0.009)	0.003 (0.008)
Age squared ( $\times 0.01$ )	-0.004 (0.004)	-0.004 (0.004)	-0.002 (0.009)	-0.001 (0.008)
Female	-0.134** (0.029)	-0.135** (0.030)	-0.152 (0.082)	-0.157 (0.083)
Urban	0.076** (0.023)	0.076** (0.023)	0.095 (0.064)	0.100 (0.062)
Country (base: Uganda)				
South Africa	0.110 (0.052)	0.102 (0.067)		
Tanzania	0.073 (0.033)	0.067 (0.036)	0.096 (0.094)	0.097 (0.099)
Uganda	-0.014 (0.018)	-0.021 (0.023)	0.013 (0.058)	0.013 (0.065)
Sample selection				
Panel retention	-0.021 (0.029)	-0.018 (0.024)	-0.136 (0.149)	-0.140 (0.140)
From $t = 0$ to $t = 1$				
Employed (non-farm) in $t = 0$	0.005 (0.010)	0.005 (0.011)	-0.044 (0.115)	-0.038 (0.114)
Constant	1.500** (0.269)	1.504** (0.269)	1.765* (0.490)	1.805* (0.484)
Observations	7,240	7,240	2,760	2,760
Adj. R <sup>2</sup>	0.371	0.376	0.382	0.384

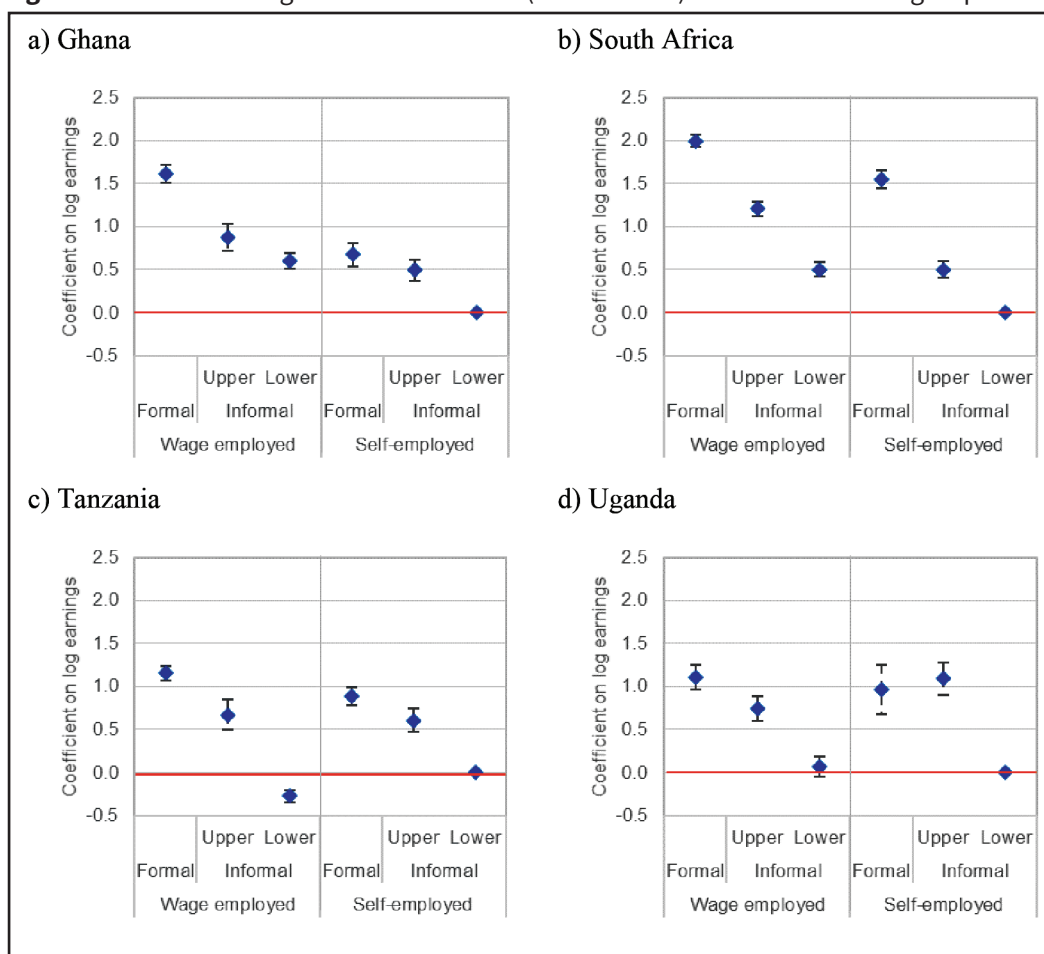
Note: \*\*\* $p < 0.01$ , \*\*  $p < 0.05$ , \* $p < 0.1$ . Standard errors clustered at the country level in parentheses.

Source: Authors' own calculations based on survey data from GSPS 2009/10-2013/14, NIDS 2014/15-2017, TZNPS 2010/11-2012/13, and UNPS 2010/11-2011/12.

GSPS, Ghana Socioeconomic Panel Study Survey; NIDS, National Income Dynamics Study; TZNPS, Tanzania National Panel Study; UNPS, Uganda National Panel Study.

### Appendix C: Complementary Figures

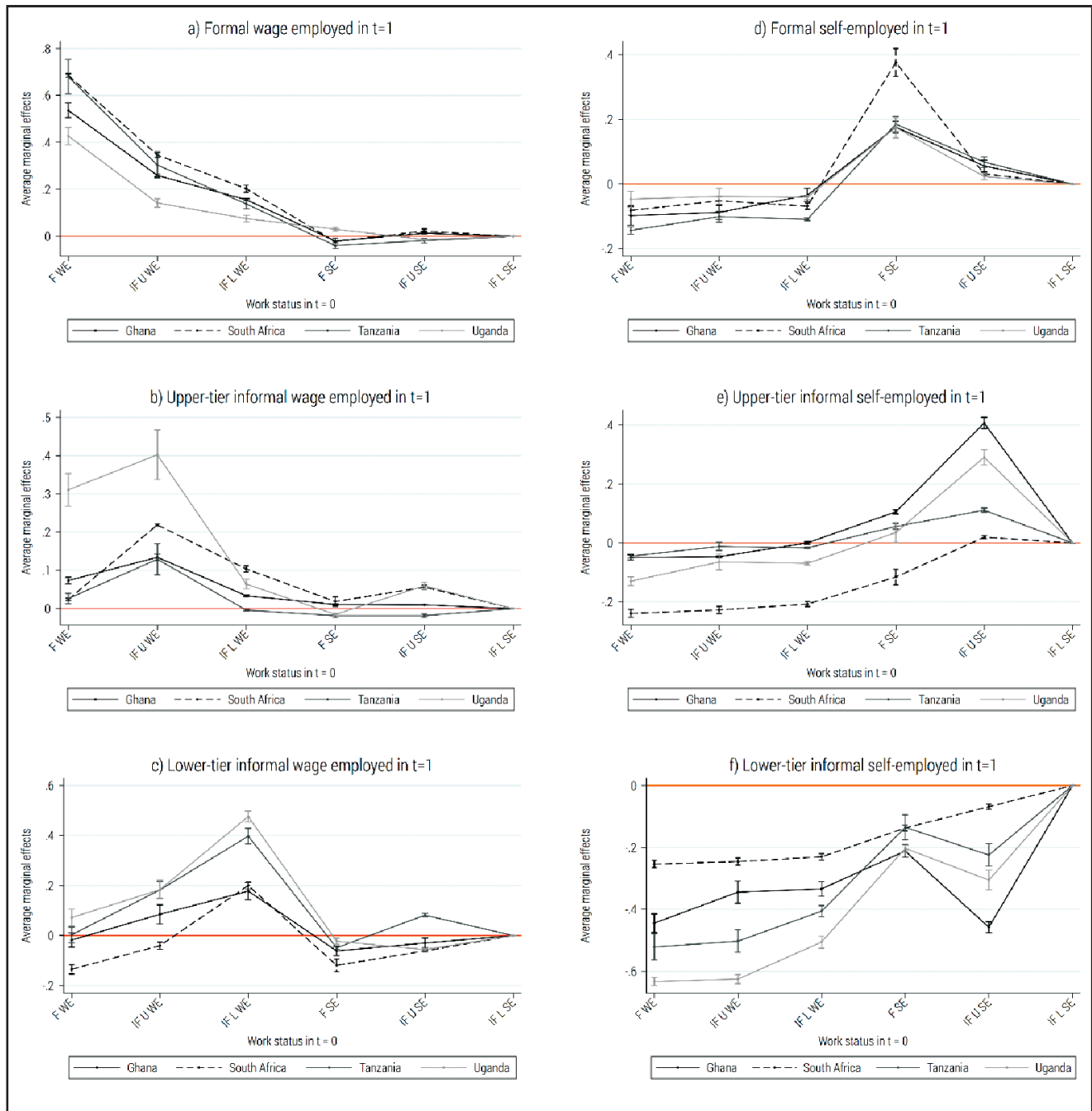
**Figure C1** ANOVA of log mean labor income (in 2005 PPPs) across work status groups.



*Note:* Coefficient estimates and 95% confidence interval of fitted one-way ANOVA model of log mean labor income using formal self-employment as the base category, controlling for wave fixed effects.

*Source:* Authors' own calculations based on survey data from GSPS 2009/10-2013/14, NIDS 2014/15-2017, TZNPS 2010/11-2012/13, and UNPS 2010/11-2011/12. GSPS, Ghana Socioeconomic Panel Study Survey; NIDS, National Income Dynamics Study; PPP, Purchasing power parity; TZNPS, Tanzania National Panel Study; UNPS, Uganda National Panel Study.

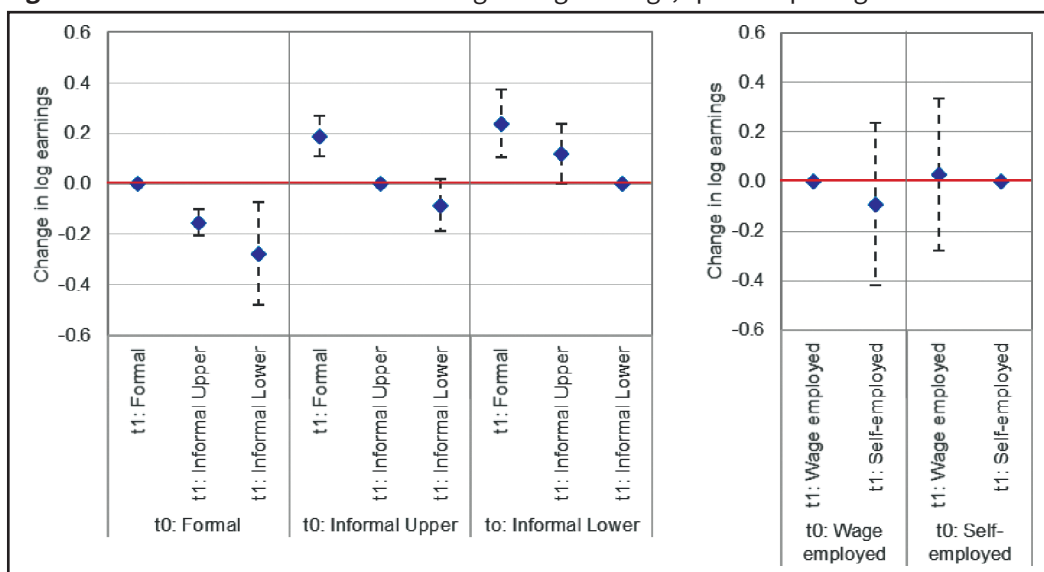
**Figure C2** Marginal effect on employment transition probabilities, by country and destination state.



*Note:* Average marginal effects estimated from same model specification and controls as in Table 3 but including interactions between the initial work status and country fixed effects. Each point shows the estimated marginal effect on the employment transition probability by country and destination state, with “lower-tier informal self-employed” being the base category in each country. The dashed lines show the 95% confidence intervals. Work status abbreviations: F, formal, IF U, informal upper-tier, IF L, informal lower-tier; SE, self-employed; WE, wage employed.

*Source:* Authors’ own calculations based on survey data from GSPS 2009/10-2013/14, NIDS 2014/15-2017, TZNPS 2010/11-2012/13, and UNPS 2010/11-2011/12. GSPS, Ghana Socioeconomic Panel Study Survey; NIDS, National Income Dynamics Study; TZNPS, Tanzania National Panel Study; UNPS, Uganda National Panel Study.

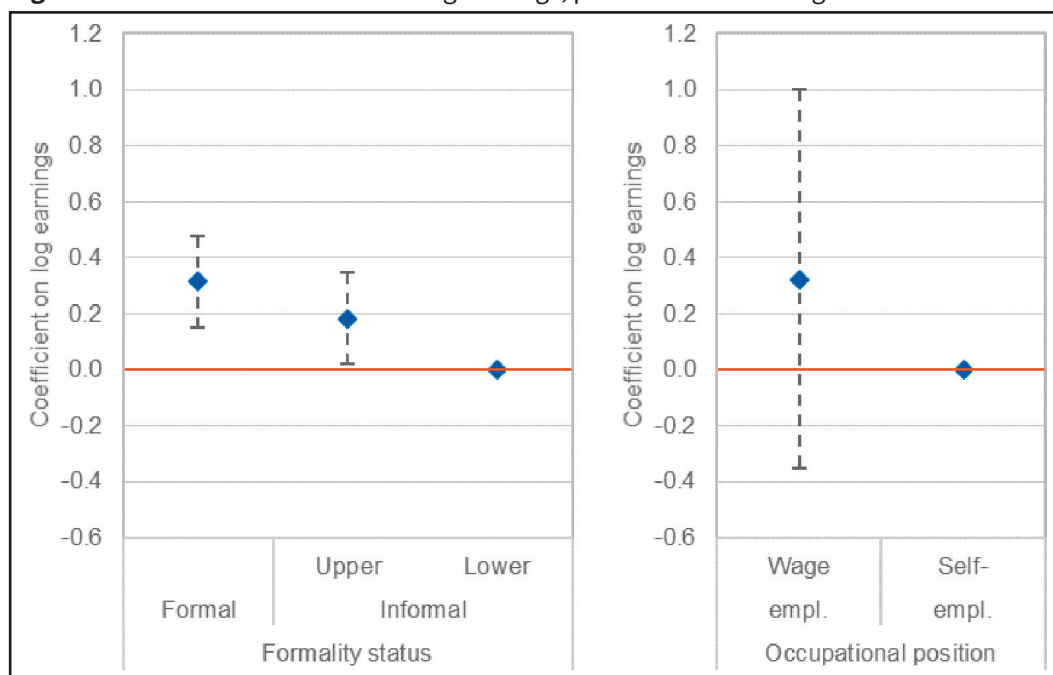
**Figure C3** Coefficient estimate on change in log earnings, split-sample regressions.



*Note:* Coefficient estimates based on same model specification and controls as in Table B10, but splitting the sample by initial formality status (formal, upper-tier informal, lower-tier informal) and initial occupational position (wage employed, self-employed). Standard errors have been clustered at the country level. Each point shows the estimated marginal effect on log earnings by employment state, with “Informal Lower” and “Wage Employed” being the base categories. The dashed lines show the 95% confidence intervals.

*Source:* Authors’ own calculations based on survey data from GSPS 2009/10-2013/14, NIDS 2014/15-2017, TZNPS 2010/11-2012/13, and UNPS 2010/11-2011/12. GSPS, Ghana Socio-economic Panel Study Survey; NIDS, National Income Dynamics Study; TZNPS, Tanzania National Panel Study; UNPS, Uganda National Panel Study.

**Figure C4** Coefficient estimate on log earnings, panel fixed effects regression.



*Note:* Coefficient estimates based on fixed effects panel regression, controlling for individual’s age and age squared and country-time fixed effects. Standard errors have been clustered at the country level. Each point shows the estimated marginal effect on log earnings by employment state, with “Informal Lower” and “Wage Employed” being the base categories. The dashed lines show the 95% confidence intervals.

*Source:* Authors’ own calculations based on survey data from GSPS 2009/10-2013/14, NIDS 2014/15-2017, TZNPS 2010/11-2012/13, and UNPS 2010/11-2011/12. GSPS, Ghana Socio-economic Panel Study Survey; NIDS, National Income Dynamics Study; TZNPS, Tanzania National Panel Study; UNPS, Uganda National Panel Study.