

Job Creation and Wages in Least Developed Countries: Evidence from Sub-Saharan Africa

Short Title: Job Creation and Wages

Juan Pablo Rud Ija Trapeznikova *

Abstract

Least developed economies are characterised by poorly functioning labour markets: only a small fraction of workers is in paid employment, where productivity and wages are low. We incorporate a standard search framework into a two-sector model of development to assess the importance of different obstacles to job creation and productivity. The model provides new insights in the characterisation of poorly developed labour markets that are observed in the data, such as high wage dispersion. We estimate the model using micro data for six countries in Sub-Saharan Africa and highlight the empirical relevance of labour market frictions, entry costs and skills.

Keywords: development, labour markets, job creation, wages, entry costs, search and matching, skills, inequality

*Corresponding author: Juan Pablo Rud, Royal Holloway, University of London, Egham, Surrey, TW20 0EX, UK. Email: juan.rud@rhul.ac.uk. We are grateful to the editor, Barbara Petrongolo, and to three anonymous referees whose comments have greatly improved the manuscript. We thank Jesper Bagger, Manolis Galenianos, Dan Hamermesh, John Kennan, Philipp Kircher, Michael Koelle, Melanie Lührmann, Fernando Martin, Santiago Oliveros, Elias Papaioannou, Victor Rios-Rull, Mark Roberts, Rob Shimer, Linas Tarasonis, Jon Temple, Gianluca Violante, Yves Zenou and seminar participants at Alicante, Chicago Fed, Edinburgh, Essex, Federal Reserve Board, QMUL, RHUL, St. Gallen, St. Louis Fed, Sussex, UTDT, 2nd African SAM Workshop, CSAE, EEA, IFS-EdePo, Novafrika, Sandbjerg and SED Conference for useful comments and suggestions. We are grateful to Taye Mengistae at World Bank for helping us procure the data.

1 Introduction

Least developed economies are characterized by poorly functioning labour markets: only a small fraction of the labour force is in paid employment, and both productivity and wages are very low.¹ In many Sub-Saharan African (SSA) countries, for example, up to nine out of ten workers are engaged in own-account work or helping family activities for no pay, predominantly in subsistence farming and petty trade. Labour productivity in SSA is, on average, fourteen times lower than in advanced economies and four times lower than in Latin America (ILO, 2012). A key question is what prevents labour markets from adjusting. In other words, why are jobs not being created through lower wages, as competitive labour market models would predict?

Recent studies in development economics have established both theoretically and empirically a number of factors that constrain wage sector growth in poor economies, including firm entry costs, labour market inefficiencies, low skills, poor infrastructure, and low aggregate productivity.² In this paper, we propose a simple unifying framework that allows us to analyse the role of these factors for determining job creation, productivity and wages simultaneously. We then estimate the model using micro data from a number of countries in Sub-Saharan Africa to examine the empirical relevance of various constraints.

The main contribution of this paper is to develop a model that can account for

¹See, for example, Fields (2011) and Banerjee and Duflo (2007) for a comprehensive summary.

²See, for example, Banerjee and Newman, 1993 and Ghatak and Nien-Huei Jiang, 2002 for the role of credit constraints, Djankov, La Porta, Lopez-de Silanes and Shleifer, 2002 and Herrendorf and Teixeira, 2011 for barriers to firm entry, Hsieh and Klenow, 2009, Vollrath, 2014 and Hsieh and Klenow, 2010) for misallocation of resources.

cross-country differences among least developed countries in underlying productivity, entry barriers, labour market frictions, and workers' outside option within a single framework. The integrated model allows us to analyse the relative importance of each channel and the interactions between them.

We model a dual labour market with a frictional wage sector and a frictionless subsistence sector. In particular, we incorporate the tools of a standard search and matching framework (Mortensen and Pissarides, 1994) with firm and worker heterogeneity into a traditional two-sector model of development (e.g. Harris and Todaro, 1970, Lewis, 1954, Robinson, 1976, and Banerjee and Newman, 1993). The wage sector is populated by heterogeneous firms that use labour for production. Workers differ in their ability and search intensity. To enter the market and realize their productivity, firms have to pay a one-time entry cost. Labour market frictions imply that it takes time and resources for firms and workers to match with each other. Workers that are unsuccessful in their job search end up in subsistence work. Finally, a wage bargaining process links wages to firms' productivity and the workers' outside option (i.e. income from low productivity self-employment).

The model generates a rich characterization of labour markets in least developed countries. First, we show that entry barriers limit the reallocation of workers from the subsistence sector to wage employment, and reduce average productivity and wages. Second, under some mild distributional assumptions on underlying firm productivity, the model implies that a higher degree of frictions results in a larger wage dispersion. The intuition for this result is straightforward: entry costs reduce firm entry and result in lower competition and, as a consequence, low productivity firms are more likely to survive. Hence, the wage sector size, average wages and productivity fall, at the same time as the variance of wages increases. The relationship between the first and second moments of wages helps us identify

the relative importance of various constraints. Based on this result, we can show that differences in the underlying productivity distribution alone are not sufficient to explain the observed differences in wage distributions across countries and that frictions play a large role in shaping labour market outcomes.

The fact that wage dispersion is greater in countries with smaller wage sectors and lower average wages has not been documented in the literature. Using household-level data for twelve countries in SSA, we find that this prediction is borne out in the data. While this relationship seems to be at odds with the well-established fact that there is a positive cross-country correlation between average income and inequality among the poorest countries³, we show that high *wage* inequality and low *income* inequality can coexist in a country where a large fraction of the workforce is engaged in low-income non-wage activities. This evidence suggests that not only are there barriers to firm entry that prevent job creation, but that other obstacles, such as labour market frictions or low skills, prevent workers from moving to the highest-paying firms within the wage sector. Hence, standard policies aimed solely at expanding the wage sector may be insufficient.

We estimate the model for six economies in SSA - Niger, Uganda, Tanzania, Ethiopia, Nigeria, and South Africa. We use individual-level panel data with detailed information on workers' demographic and employment characteristics that allow us to construct transitions between the private sector and self-employment activities. Our main set of results suggests that a reduction in labour market frictions has a large impact on job creation: a one percent increase in labour market efficiency leads to about 0.7 percent increase in wage employment in South Africa and almost two percent increase in the poorer economies in our sample; whereas the reduction in entry costs is only half as effective. Moreover, we also document im-

³Namely, the left section of the Kuznets curve.

portant complementarities between policies: the effectiveness of reducing labour market frictions on improving wages and wage inequality is amplified in the presence of lower entry costs.

More generally, our results suggest that a unifying model is a valuable analytical tool to inform debates about policy effectiveness, their complementarities and the trade-offs they generate. For example, while a fall in the degree of market frictions induces job creation, it may also lead to an increase in income inequality, particularly in countries where the majority of workers are self-employed. Similarly, while programmes aiming at increasing productivity in the subsistence sector improve wages, income, and inequality, they may come at the expense of reducing firm entry and job creation. Finally, policies aiming at increasing workers' skills may have large effects on job creation, as more firms post vacancies to benefit from this productivity boost, but small effects on earnings among wage earners when workers' bargaining power or outside options are low. These insights can be used in conjunction with existing reduced-form studies to better assess the general equilibrium effects of randomized field experiments that focus on improving labour market outcomes.⁴

The paper is organised as follows. In Section 2, we discuss a set of stylized facts that characterize labour markets in developing countries. In Section 3, we develop the model and derive its main predictions. In Section 4, we present empirical support for our model. Section 5 describes how the model can be used to quantify barriers to job creation and other labour market outcomes. We discuss the estimation strategy and empirical moments used to estimate the model for a

⁴See, for example, recent studies on reducing search costs (Abebe, Caria, Fafchamps, Falco, Franklin and Quinn, Forthcoming and Franklin, 2018), improving firm entry (de Mel, McKenzie and Woodruff, 2012 and de Mel, McKenzie and Woodruff, 2013), worker training and skills upgrade (Alfonsi, Bandiera, Bassi, Burgess, Rasul, Sulaiman and Vitaliy, 2019) or enhancing productivity in the home sector through asset transfers (Banerjee, Duflo, Goldberg, Karlan, Osei, Parienté, Shapiro, Thuysbaert and Udry, 2015 and Blattman, Fiala and Martinez, 2013).

set of countries in Sub-Saharan Africa. We present our estimates and a number of simulations to quantify the role of different channels for job creation, the levels and the dispersion of wages and income. In Section 6, we briefly discuss how existing studies and policies can be linked back to our model. In Section 7, we conclude. Further information on the data sources used in the paper, key model assumptions and proofs, as well as the estimation details can be found in the online Appendix.

2 Labour markets in least developed countries

It has been well documented in the literature that labour markets in least developed countries are strikingly different from those in richer economies. In this section, we use micro data from household surveys for a number of Sub-Saharan African countries to briefly describe the stylized facts that inform our modelling assumptions in the subsequent sections (see online Appendix A for data description). We focus on SSA countries because, on average, the region has been the worst performing in a series of indicators in the last 20 years, such as labour productivity, shares of wage employment and wage growth rates (ILO, 2012).

For illustration purposes, Table 1 presents the key summary statistics for five countries in the SSA region that we compare to South Africa. Three facts are evident from this comparison. First, labour force participation in least developed countries is very high: close to 80% (with the exception of Nigeria, where it is closer to 70%) compared to 55% in South Africa and the average of 60% for the OECD countries. At the same time, the unemployment rates are below 5%, while one in four workers is unemployed in South Africa. Virtually non-existent unemployment is not a sign of healthy labour markets and the abundance of jobs; on the contrary, it reflects the idea that in poor economies adults generally cannot afford not to participate in income generating activities.

Table 1: *Labour Markets in Sub-Saharan African Countries*

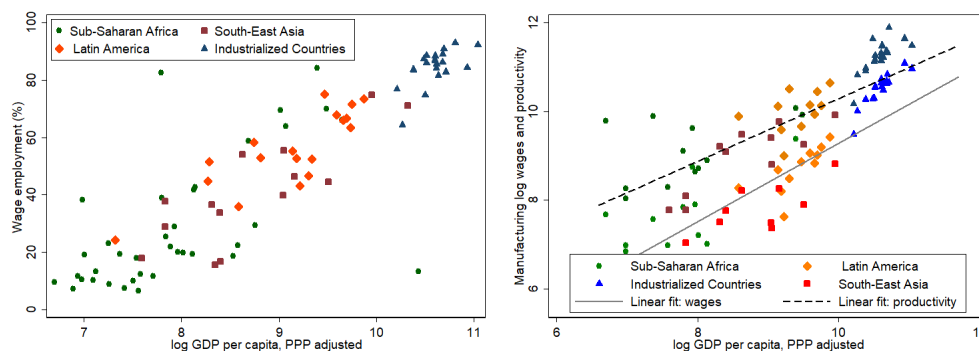
Country	Ethiopia	Niger	Nigeria	Tanzania	Uganda	South Africa
Labor Force Participation Rate, %	88.8*	78.6	65.7	79.8	82.9	55.4
Unemployment Rate, %	n/a	0.3	4.8	1.5	1.1	25.0
<i>Main activity among those working (%)</i>						
Working for someone for a pay	9.5	8.1	16.2	16.4	18.6	59.0
Household business (inc. farm)	88.0	90.8	75.2	81.1	80.7	15.1
<i>Sector among those working for a pay (%)</i>						
Private	47.1	70.4	35.5	78.3	77.3	78.5
Government	43.4	26.4	46.9	15.7	15.7	21.5
Formal/Employer pays taxes	n/a	20.5	39.4 [†]	n/a	22.1	70.3
<i>Monthly household expenditure per person (in 2010 PPP dollars)</i>						
No wage employee in HH	49.3	69.0	99.6 [‡]	84.2	94.8	110.1
At least 1 employee in HH	79.4	115.1	157.1 [‡]	183.8	159.9	243.5
Ratio	1.61	1.67	1.58	2.18	1.69	2.21

Note: The sample is limited to 15-65 year old individuals. *There is no question on job seekers in Ethiopia. We measure labour force participation based on economic activity in the last 12 months, that include everyone who is working for a pay, on household farm or business, having a casual or temporary work (including unpaid) and those in government employment program (Productive Safety Net Program). [†]For Nigeria, the formal sector share refers to firms paying pension contributions for their employees and is available for 2015 only. [‡]Consumption data for Nigeria refers to 2011 and 2012. Source: Authors' computations using the following datasets: Ethiopia ESS 2013 and 2015; Niger ECVMAI (2011) and ECVMAII (2014); Nigeria General Household Surveys 2011, 2012, 2015; Tanzania National Panel Survey Waves 1 (2008), 2 (2010), and 3 (2012); Uganda National Panel Survey 2010, 2011, 2012, and 2014; South Africa General Household Survey 2007 or Labor Force Survey 2007.

This brings us to the second fact: the share of people working for a wage is very low, relative to middle income countries. For example, while in South Africa 59% of economically active individuals (78% of those working) are in wage employment, fewer than 20% are in paid employment in the other five countries. Most people are engaged in occupations where the household is the main producer (i.e. working in the family farm or business or in low-productivity self-employment activities) using almost exclusively household labour.⁵ The left panel of Figure 1 shows that there is a clear positive relationship between the size of the wage sector and the level of development.

⁵Even non-agricultural household enterprises do not hire much: in Niger, only 2% of household firms employ at least one person outside their household, while in Tanzania and Uganda, around 15%.

Figure 1: Wage sector, productivity and development



Source: World Bank World Development Indicators Database (the left panel) and INDSTAT2 (Industrial Statistics Database), the United Nations Industrial Development Organization (the right panel). Note: Wage employment in the left panel includes wage and salaried workers, while the remainder is comprised of employers, own-account workers, members of producers' cooperatives, and contributing family workers. Wages and productivity in the right panel refer to the manufacturing sector only.

Previous studies have documented that own-account workers in least developed countries are “forced” self-employed, choosing to run their businesses not because of their entrepreneurial drive but because they cannot find a steady well-paid job in the wage sector.⁶ While it is difficult to measure self-employed income as it often includes production for own consumption, in-kind payments, barter, etc., we can compare consumption expenditures for households with no wage employees to households with a least one employee. Table 1 shows that the latter tend to be 1.5-2 times richer.⁷

Third, the wage sector in the poorest countries is characterized by relatively low levels of productivity and earnings. For example, GDP per person employed in Sub-Saharan Africa is, on average, fourteen times lower than in advanced economies

⁶For example, Fields (2011) state that workers in poor economies cannot afford to remain unemployed and to search for wage sector jobs and hence choose to create their own self-employment opportunities. Banerjee and Duflo (2007) write: “If they [petty entrepreneurs] could only find the right salaried job, they might be quite content to shut their business down”.

⁷The difference is even larger if at least 2 members work for a wage. Note that this measure includes monetary and non-monetary expenditures.

and four times lower than in Latin America (ILO, 2012).⁸ Even when focusing on the relatively high-productivity manufacturing sector, labour productivity as measured by PPP value added per employee is significantly greater for industrialized countries, by a factor of 4 (see the right panel of Figure 1). Average wages follow a similar pattern. Moreover, the labour share in total value added is almost twice as large in developed economies than in the SSA countries.

Finally, a number of papers on labour markets in developing countries highlight the distinction between formal and informal sectors.⁹ The formal sector in SSA is synonymous with the public sector, whereas almost all workers in the private sector would be classified as informal, including those working for a wage. For example, in contrast to South Africa where above 70% of people working for a wage are employed in the formal sector, the share of workers reporting that their employers pay income tax is about 20% in Niger and Uganda, amounting to 2% and 4% of all economically active individuals. Similarly, only 1.2% and 3% of non-agricultural household enterprises are registered for income or value added tax in the latter two countries. Thus, we argue that for least developed countries understanding the margin of formality-informality is not as relevant as focusing on why job creation is so low.

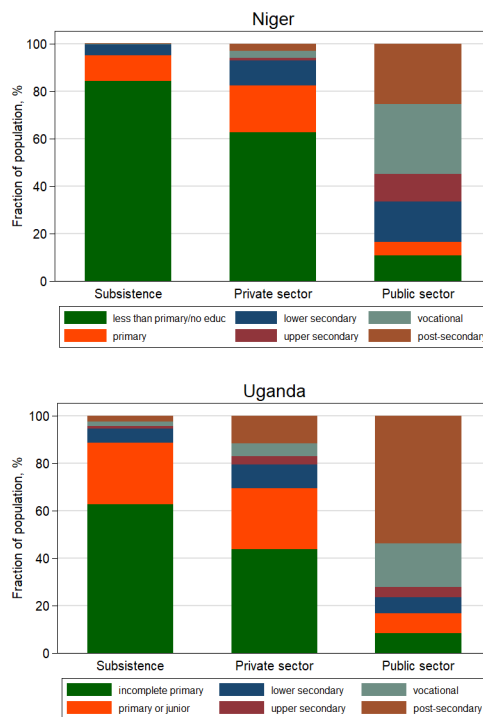
The distinction between public and private sectors is important not only because private firms might face very different constraints in developing countries than government administration and state enterprises, but the type of workers

⁸These differences in productivity and pay cannot be explained solely by the composition of the labour force, such as workers' education, skills, etc. For example, Clemens, Montenegro and Pritchett (2008) estimate the wage gain obtained by foreign workers who arrive to work in the United States relative to their country of origin. They find that the same person would earn on average 7 times more when relocating from Ghana to the US and 3 times more if coming from South Africa.

⁹The distinction between formality and informality is relevant for middle income countries such as Brazil, where informal wage employment accounts for around 20% of the labour force (see Meghir, Narita and Robin, 2015 and Ulyssea, 2010). For a comprehensive review of the evidence on informality in developing countries see La Porta and Shleifer (2014).

that are hired by these sectors is very different. For example, the majority of those working for the government have secondary and post-secondary education, while workers in the private sector and in the subsistence sector are more similar, with predominantly primary education or less (see Figure 2 for Niger and Uganda and Figure A.1 in the online Appendix for other countries). That is, the overlap in educational attainment between the two sectors - private firms and self-employment - is very large; while the opposite is true for the public sector. For this reason, we exclude the public sector from our analysis and focus on private job creation.¹⁰

Figure 2: Education distribution by sector



Source: Authors' computations using the following datasets:
Niger ECVMAI (2011) and ECVMAII (2014), Uganda NPS 2010, 2011, 2012, and 2014.

¹⁰Girsberger and Meango (2018) show the extent to which the labour market works differently in relation to the public sector: educated workers are willing to remain unemployed as they wait for better-paying and more stable public sector jobs.

Although the composition of workers in the private and subsistence sectors is similar, education and other demographic factors play a role in selection into paid employment. Table A.1 in the online Appendix shows how the probability to get a private sector job (as opposed to working on a household farm or business) depends on worker's characteristics based on a linear probability model. As expected, workers with vocational and post-secondary education (in some countries also secondary) are more likely to be employed for a wage. However, note that more educated workers comprise less than 20% of private sector employees (30% in Nigeria). In brief, we conclude that it is unlikely that selection on worker characteristics can explain most of the patterns in labour markets in least developed countries, including the lack of relocation from self-employment to the wage sector. However, this pattern may be symptomatic of poorly functioning labour and product markets as low skills can hurt job creation and firm productivity. Our analysis will allow for this channel to play a role in labour market outcomes.

In sum, an extremely small wage sector in Sub-Saharan Africa, coupled with the evidence that income from household production activities tends to be lower than wages, is a clear indicator of underperforming labour markets. The development economics literature has established, both theoretically and empirically, a number of factors that prevent job creation and productivity growth in developing countries, including limits to labour mobility, lack of skills, and barriers to firm entry.¹¹

In the next section, we develop a model that accounts for these various constraints

¹¹A key aspect in the early literature focused on constraints to labour mobility in a traditional two-sector model of development (e.g. Harris and Todaro, 1970, Lewis, 1954, and Robinson, 1976). More recently, the literature has also explored other channels. For example, credit constraints can affect the occupational choice of individuals and determine both the size of the modern sector and the level of wages (e.g. Banerjee and Newman, 1993 and Ghatak and Nien-Huei Jiang, 2002). Similarly, regulatory barriers to firm entry have been associated with higher employment in non-wage activities (Djankov *et al.*, 2002 and Herrendorf and Teixeira, 2011, among others). Finally, the misallocation of resources has also been identified as a constraint to job creation and as an important determinant of wages (see Hsieh and Klenow, 2009, Vollrath, 2014 and Hsieh and Klenow, 2010).

within a unifying framework and estimate their relative importance for restricting job growth.

3 Model

We propose an integrated model of home production, or subsistence sector, and the wage sector that allows for both entry barriers and labour market frictions. In particular, we incorporate the tools of a standard search and matching framework (Mortensen and Pissarides, 1994) into a two-sector model of development, as in Harris and Todaro (1970) and Lewis (1954). In the Harris-Todaro model migrants move until income in the traditional sector is equalized to the expected income in the modern/urban sector (i.e. probability of getting a job times income). The search framework captures this fundamental aspect of Harris-Todaro as wages and unemployment are jointly determined. It also adds the notion that firms and workers spend time and resources before the match is created and that labour market inefficiencies matter for job creation, wages and productivity. Similarly to the Lewis model, a large surplus of labour in the home production sector means low marginal returns, which make jobs in the wage sector more attractive. That is, workers are willing to queue and to pay a search cost to find wage sector jobs. In line with the observation that more educated workers are more likely to populate wage sector jobs and that their skills can affect job creation and firm productivity, we also add worker heterogeneity. In addition, our model incorporates a standard assumption of costly market entry under firm heterogeneity, so that a firm's success or exit is linked to its idiosyncratic productivity, in equilibrium.¹² With these ingredients, our model explores how entry barriers and labour market frictions limit the

¹²This framework, first developed by Hopenhayn (1992), has been used in a variety of settings. See Aw, Chung and Roberts (2003), Melitz (2003) and Bartelsman, Haltiwanger and Scarpetta (2013), for example.

reallocation of workers from subsistence to the wage sector, and reduce average productivity and wages.

3.1 Setup

This is a continuous time model of two labour markets - the wage sector and the self-employment, or home production, sector. There is a continuum of infinitely lived workers, with a mass normalized to one, that supply labour to firms. There are two types of workers in the labour market - fraction α of workers have ability a_L , while $1 - \alpha$ have ability a_H , with $a_H > a_L$.¹³

(a) Wage sector

The wage sector is populated by heterogeneous firms that differ in their productivity level p . There are infinitely many potential firms that may enter the market and open a job after paying fixed cost k . Firm productivity is revealed upon entry and is constant over the firm's lifetime. The technology exhibits constant returns to scale and uses labour as input.¹⁴

In the wage sector, firms and workers are brought together pairwise through a sequential and random matching process. To recruit, firms post a vacancy at cost c per unit of time. We assume that workers with higher ability search for a job with higher relative search intensity $s_H > s_L = 1$.¹⁵ Alternatively, the assumption that $s_H > s_L$ can be interpreted as higher ability workers being more successful in getting a wage job, which might be due to them using more effective search methods, having a broader (formal or informal) job contact network, or being better

¹³The model can be extended to an arbitrary number of worker types; however, two types are sufficient to gain the intuition.

¹⁴Given the constant returns to scale production function the size of a firm is undetermined. Without loss of generality, we can think of each firm consisting of a single job. Hence, we use 'jobs' and 'firms' interchangeably.

¹⁵As we have seen in Section 2, more educated workers are more likely to be working in the wage sector than their less educated counterparts. Allowing for differential search intensity by worker type is one way of modelling this selection.

able to signal their ability to potential employers.

Reflecting search frictions, the offer arrival and vacancy filling rates are exogenous to workers and firms but are determined in equilibrium. The matching function $M(v, u)$ is assumed to be increasing, concave, and homogenous of degree one in both arguments – aggregate vacancies v and job seekers $u = s_L u_L + s_H u_H$, where u_j is the mass of job seekers of type j and s_j is their relative search intensity, with s_L normalized to one. As is standard in the literature, we assume a Cobb-Douglas form, i.e.

$$M(v, u) = m v^\eta u^{1-\eta}, \quad 0 < \eta < 1, \quad (1)$$

where m is a matching efficiency parameter. Given the constant returns to scale assumption, we can express the job finding and job filling rates as functions of market tightness, $\theta = \frac{v}{u}$. That is, when workers search for a job they receive an offer at Poisson arrival rate $\lambda = \frac{M(v, u)}{u} = m\theta^\eta$ per search efficiency unit; the vacancy filling rate is given by $q = \frac{M(v, u)}{v} = m\theta^{\eta-1}$ and $s_j u_j / u$ is the probability that a randomly met worker is of type j .

When a firm with productivity p and a worker with ability a_j form a match, the job produces output pa_j . Jobs are subject to an exogenous destruction shock that arrives at rate δ . Competition and entry costs endogenously determine the number of firms in the market. Similarly, the ability distribution among the wage sector workers is endogenous and is driven by the differential search intensity between the two types. Wages are determined through a bargaining process between the firm and its workers. Both workers and firms are risk neutral and they discount the future at rate r .

(b) Home sector

Workers without a job end up in home production. Unlike in industrialized countries, the unemployment rate in least developed economies is very low or vir-

tually non-existent; therefore, self-employment income is a more relevant outside option for workers. The aggregate production in the home sector Y_H is assumed to be an increasing concave function of home sector labour measured in efficiency units, i.e. $Y_H = AL_H^\gamma$, where A captures aggregate self-employment productivity (reflecting other factors of production that are assumed to be fixed, such as land or aggregate capital), $L_H = a_{LuL} + a_{HuH}$ is the aggregated ability of self-employed workers, and $0 < \gamma < 1$ is a returns to scale parameter. Note that all self-employed workers are assumed to be looking for a wage job (we discuss the significance of this assumption in online Appendix B.1).

The home production sector is assumed to be competitive so that a j -type worker's earned income in home production is equal to her marginal product:¹⁶

$$h_j = \gamma AL_H^{\gamma-1} a_j. \quad (2)$$

That is, income earned in the home sector is proportional to worker's ability and can be written as $h_j = \tilde{h} a_j$, where $\tilde{h} = \gamma AL_H^{\gamma-1}$ is common to both types. Given that $\gamma < 1$, this setup implies that a larger self-employment sector is associated with lower incomes.¹⁷

¹⁶We adopt this approach as it is standard in the literature (see for example Zenou, 2008). It implicitly assumes that landlords or other owners of fixed factors (captured by parameter A), get the surplus not earned by the self-employed. Although the latter are not explicitly included in our analysis, this is not crucial for the results. As an alternative, we can assume that self-employed workers receive the average labour product. While it leads to higher estimates of A in our numerical exercises, none of the policy experiments or other results are affected.

¹⁷For the agricultural sector, for example, this could be interpreted as the amount of land being fixed as in Matsuyama (1992). Alternative explanations include a decrease in productivity due to a fall in either land or labour quality. Lagakos and Waugh (2013), for example, propose a Roy model where a small non-agricultural sector implies a larger agricultural sector populated with relatively unproductive workers.

3.2 Worker's problem

The value of employment of a worker with ability a_j at a firm with productivity p , $W_j(p)$, satisfies the following Bellman equation:

$$rW_j(p) = w_j(p) + \delta(U_j - W_j(p)), \quad (3)$$

where r is the common firms' and workers' discount rate and U_j is the value of search to a j -type worker. The right-hand side of the equation is the sum of income flow from working, w_j , and the expected capital loss if the job is destroyed and the worker becomes self-employed and searching. The latter event happens at constant Poisson rate δ . The value of working can be re-written as

$$W_j(p) = \frac{w_j(p) + \delta U_j}{r + \delta}. \quad (4)$$

Assuming that wages are increasing in firm's productivity and worker's ability, which we show further below, the value of employment is also strictly increasing in p and a .¹⁸

Job search is a costly process that involves direct search costs and time away from home production. Hence, we postulate that a self-employed job seeker with ability a_j obtains consumption flow $h_j - z_j$ by means of home production less search costs z_j , and she has an option of finding a job in the wage sector. The value of search U_j then solves the following Bellman equation for a j -type worker:

$$rU_j = h_j - z_j + s_j \lambda(\theta) \int (\max\{W_j(p), U_j\} - U_j) d\bar{F}(p), \quad (5)$$

¹⁸Note that in addition to individual worker's type j and firm's productivity p all value functions depend on equilibrium aggregate variables (such as the market tightness θ). To simplify the exposition, we omit them from the value function notations.

where $s_j\lambda(\theta)$ is the job offer arrival rate that depends on market tightness θ and $\tilde{\Gamma}(\cdot)$ is the cumulative distribution function of firms that operate in the market.

Using equations (4) and (5), we can solve for the worker's reservation wage w_{Rj} that equates the value of search with that of working:

$$w_{Rj} = h_j - z_j + \frac{s_j\lambda(\theta)}{r + \delta} \int_{p_{Rj}} (w_j(p) - w_{Rj}) d\tilde{\Gamma}(p), \quad (6)$$

where p_{Rj} is defined as the productivity of the marginally acceptable firm for a j -type worker, i.e. $w_j(p_{Rj}) = w_{Rj}$.

3.3 Firm's problem

Firms operate a constant returns to scale technology in labour and differ in their productivity level p . The value of a job in a firm with productivity p hiring a worker with ability a_j , $J_j(p)$, solves the following Bellman equation:

$$rJ_j(p) = pa_j - w_j(p) + \delta(V(p) - J_j(p)). \quad (7)$$

The first term on the right-hand side of equation (7) is the firm's profit flow, $pa_j - w_j(p)$. The second term is the expected capital loss related to the possibility that the job is destroyed, in which case the firm ends up with the value of an open vacancy, $V(p)$.

To hire a worker, the firm needs to post a vacancy that is then randomly matched with job seekers. The hiring rate, $q(\theta)$, is derived from the matching function and depends on aggregate market tightness; while probability that the randomly met worker is of type j is equal to $s_j u_j / u$. The value of an open vacancy can be found as

$$rV(p) = -c + q(\theta) \left(\frac{u_L}{u} \max\{J_L(p) - V(p), 0\} + \frac{s_H u_H}{u} \max\{J_H(p) - V(p), 0\} \right), \quad (8)$$

where c is a vacancy posting cost and $u = u_L + s_H u_H$ is the mass of the job seekers. The value of a vacant job, $V(p)$, and a filled job, $J_j(p)$, are strictly increasing in p .

3.4 Wage determination

Once a match is formed, the firm and the worker bargain over the wage. Bargaining between each worker-firm pair takes place in sequence of rounds and we assume that the threat point of a worker is the value of delay.¹⁹ During a potential delay, the worker engages in home production and receives the flow value of $h_j = \tilde{h}a_j$, while the firm is idle during that period as the firm cannot replace the worker instantaneously. Then, the wage paid to the worker is a solution to the following bargaining problem:

$$w_j(p) = \arg \max_w (pa_j - w)^{1-\beta} (w - \tilde{h}a_j)^\beta, \quad (9)$$

where $0 < \beta < 1$ represents the worker's bargaining power. Taking the first order condition, we obtain the following equation for the wage as a function of productivity:

$$w_j(p) = (\beta p + (1 - \beta)\tilde{h})a_j, \quad (10)$$

conditional on $p > \tilde{h}$.

¹⁹Hall and Milgrom (2008) point out that a permanent breakdown of negotiations (and hence of a match) is not a credible threat point in a highly frictional market. Instead, they suggest to use a temporary disruption of negotiations, and the resulting forgone production during the delay, as a threat point. We follow Elsby and Gottfries (2019) who assume continuous renegotiation. Since wages are renegotiated every period, the probability of a breakdown is infinitely small, and hence the wage is a function of only the flow surplus. Alternatively, we can interpret this as a wage-setting mechanism, where wages are renegotiated every day and either the worker or the employer is selected to make a take-it-or-leave-it offer setting the wage for that day. Then β can be thought of as the probability that the worker makes the offer.

3.5 Labor market clearing

Firms are identical ex ante and their type is revealed upon entry.²⁰ Productivity of potential entrants is assumed to be drawn randomly from a given distribution $\Gamma(\cdot)$ with the support $[\underline{p}, \infty)$. Firms have to pay fixed cost k per job upon entry reflecting credit constraints and other entry impediments. Hence, the free entry condition implies that

$$-k + \int_{\hat{p}} V(p) d\Gamma(p) = 0, \quad (11)$$

which means that the value of an open vacancy in expectation should be equal to the entry cost. The lowest productivity level, for which a firm would post a vacancy, is denoted by \hat{p} and is such that $V(\hat{p}) = 0$. This condition is referred to as the zero profit condition. That is, firms with productivity below \hat{p} exit the market immediately after entry and receive the value of zero. Here, we implicitly assume that the wage offer paid by the firm with the lowest productivity level is accepted by all workers, i.e. $w_j(\hat{p}) \geq w_{Rj}$ for all worker types j .

Substituting for the wage function $w(p)$ into the value of a job given in equation (7), we can rewrite the flow value of a vacancy as follows

$$rV(p) = -\frac{r + \delta}{r + \delta + q(\theta)}c + \frac{q(\theta)}{r + \delta + q(\theta)}(1 - \beta)(p - \tilde{h})\bar{a}, \quad (12)$$

where $\bar{a} = \frac{u_L}{u}a_L + \frac{s_H u_H}{u}a_H$ is the average ability weighted by the probability of meeting a job seeker of a particular type. Note that the wage rate in the home production sector, \tilde{h} , and the average ability of a randomly met worker, \bar{a} , are determined endogenously in equilibrium and are functions of market tightness θ . From now on, we will use $\bar{a}(\theta)$ and $\tilde{h}(\theta)$ notations.

It is useful to rewrite the free entry condition in (11) as the expected gain relative

²⁰As in Hopenhayn (1992), Aw *et al.* (2003), Melitz (2003) and Bartelsman *et al.* (2013).

to the outside option of exiting the market, $V(\hat{p})$:

$$rk = \int_{\hat{p}} (rV(p) - rV(\hat{p}))d\Gamma(p) = \frac{q(\theta)(1-\beta)\bar{a}(\theta)}{r+\delta+q(\theta)} \int_{\hat{p}} (p - \hat{p})d\Gamma(p), \quad (\text{FE})$$

where we use the fact that $V(\hat{p})$ is equal to zero. Define the surplus function as $\varphi(\hat{p}) = \int_{\hat{p}} (p - \hat{p})d\Gamma(p)$, i.e. the average productivity gain in excess of the reservation productivity in the market. Integrating by parts we can show that the surplus function $\varphi(\hat{p}) = \int_{\hat{p}} (1 - \Gamma(p))dp$, with $\varphi'(\hat{p}) = \Gamma(\hat{p}) - 1 < 0$ and $\varphi''(\hat{p}) = \Gamma'(\hat{p}) > 0$. Below we show that $\bar{a}(\theta)$ is a decreasing function of θ , as is the vacancy filling rate $q(\theta)$, which means that the free entry condition implies a decreasing relationship between \hat{p} and θ .

The reservation productivity level \hat{p} is derived from setting $V(\hat{p}) = 0$, i.e.

$$\frac{c}{q(\theta)} = (1-\beta) \frac{(\hat{p} - \tilde{h}(\theta))\bar{a}(\theta)}{r+\delta}. \quad (13)$$

This equation shows that at the threshold the expected cost of keeping an open vacancy (the flow cost c multiplied by the average duration of an opening $\frac{1}{q(\theta)}$) should be equal to $(1-\beta)$ times the share of the present discounted value of the match surplus (output flow p less home production \tilde{h} multiplied by the expected ability of a hired worker). Rearranging this equation, we get the following zero profit condition

$$\hat{p} = \tilde{h}(\theta) + \frac{c(r+\delta)}{q(\theta)(1-\beta)\bar{a}(\theta)}. \quad (\text{ZP})$$

The link between \hat{p} and θ is given by the vacancy filling rate q , income in the home sector \tilde{h} (which depends negatively on the size of the home sector L_H) and the average ability of a randomly met worker \bar{a} . To determine this, we consider a steady state equilibrium in which the composition of labor between the two sectors is constant and the sum is equal to one. Hence, the outflow from the home

production sector should be equal to the outflow from the wage sector. That is,

$$\lambda(\theta)u_L = \delta(\alpha - u_L) \Rightarrow u_L = \frac{\delta\alpha}{\delta + \lambda(\theta)}, \quad (14)$$

$$s_H\lambda(\theta)u_H = \delta(1 - \alpha - u_H) \Rightarrow u_H = \frac{\delta(1 - \alpha)}{\delta + s_H\lambda(\theta)}, \quad (15)$$

where α is the share of L -type of workers in the population. Then the average ability of a randomly met worker is given by

$$\bar{a}(\theta) = a_L + (a_H - a_L) \frac{(1 - \alpha)s_H(\delta + \lambda(\theta))}{(\alpha\delta + s_H\lambda(\theta) + (1 - \alpha)\delta s_H)}, \quad (16)$$

which is decreasing in θ . The intuition for this result is that, as the wage sector grows and more jobs are created, high ability workers are more likely to leave self-employment than low ability workers due to their more intensive job search. As a result, the composition of the pool of job seekers starts shifting towards low ability workers.

The efficiency units of labour engaged in home production is equal to

$$L_H(\theta) = a_L u_L(\theta) + a_H u_H(\theta) = \frac{\delta\alpha a_L}{\delta + \lambda(\theta)} + \frac{\delta(1 - \alpha)a_H}{\delta + s_H\lambda(\theta)}, \quad (17)$$

which is decreasing in θ as the number of job seekers of both types is falling in market tightness. Therefore, income gained in the home sector \tilde{h} is increasing in θ , due to the decreasing returns to scale in home production.

In sum, we have two equations - the zero profit (ZP) condition and the free entry (FE) condition- and two unknowns: \hat{p} and θ . The zero profit condition is upward sloping, while the free entry condition is downward sloping, resulting in a unique equilibrium. Note that this solution relies on two assumptions: (i) the lower bound of the productivity distribution is determined by the firm's side (i.e. labour

demand) as opposed to the workers' reservation wage, and (ii) all self-employed workers search for a job. In online Appendix B.1, we discuss these assumptions in detail and show that they are not restrictive.

3.6 Model predictions

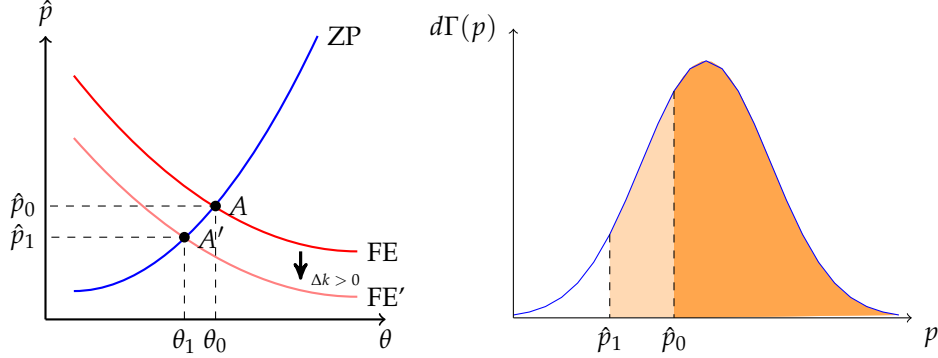
There are a number of parameters that shape labour market outcomes in this model. Some of the key parameters are entry costs k , labour market efficiency m , home sector productivity A , workers' bargaining power β , ex-ante productivity distribution, $\Gamma(p)$, and the ability distribution in the population, captured by α . In this subsection, we use changes in the entry costs k to illustrate that the model can deliver a rich characterization of labour markets in least developed countries, while the effects of other variables are described in detail in online Appendix B.2.

Prediction 1: There is a positive relationship between wage employment, productivity and wages.

Consider an increase in entry costs k . Holding the reservation productivity constant, in order to recover the now-higher fixed costs, the level of competition (i.e. market tightness) needs to be lower, thus shifting the Free Entry curve downwards. As a result, both the reservation productivity and market tightness fall (see Figure 3). As it becomes more difficult to enter the market, the number of firms falls and so does the vacancy-to-unemployment ratio. The decrease in market tightness θ reduces the job finding rate λ and, as a consequence, the size of the wage sector. Perhaps counter-intuitively, the increase in the entry costs also leads to a drop in average productivity: conditional on entry, low productivity firms are more likely to survive as fewer firms enter. As a consequence, the survival threshold \hat{p} is now lower.

In addition to hindering job creation and lowering average productivity, entry

Figure 3: The effect of an increase in entry costs k



barriers have a direct implication in terms of wage levels. Recall that wages are a linear combination of productivity and self-employment income for each worker type. Therefore, the parameter differences that lower the productivity threshold and market tightness simultaneously will reduce mean wages through both of these channels: a lower average productivity and a lower value of the outside option through a decreasing marginal product of labor in the home sector. Thus differences in entry costs, all else equal, will generate a positive relationship between the wage sector size, average productivity and wages.

Prediction 2: There is a negative relationship between mean wages and wage dispersion.

Conditional on the same mean, a greater wage dispersion reflects a higher degree of market inefficiencies and hence is more illuminating about constraints to job creation and wage growth. The total variance of log wages in the model is the sum of the variance arising from firm heterogeneity, p , across the same workers (within-group component) and the between-group component due to differences in workers' abilities, a . Denote the firm component of wages by $\tilde{w}(p) = \beta p + (1 - \beta)\tilde{h}$ so that $w(p, a) = \tilde{w}(p)a$. Then,

$$\text{Var}(\ln w) = \text{Var}(\ln \tilde{w}(p)) + \frac{e_L e_H}{(e_L + e_H)^2} (\ln a_H - \ln a_L)^2, \quad (18)$$

where $e_L = \alpha - u_L$ and $e_H = 1 - \alpha - u_H$, the mass of employed workers of types L and H , respectively. Using Taylor approximation of $\ln \tilde{w}$ around $\ln \tilde{h}$ to linearise log wages, we get

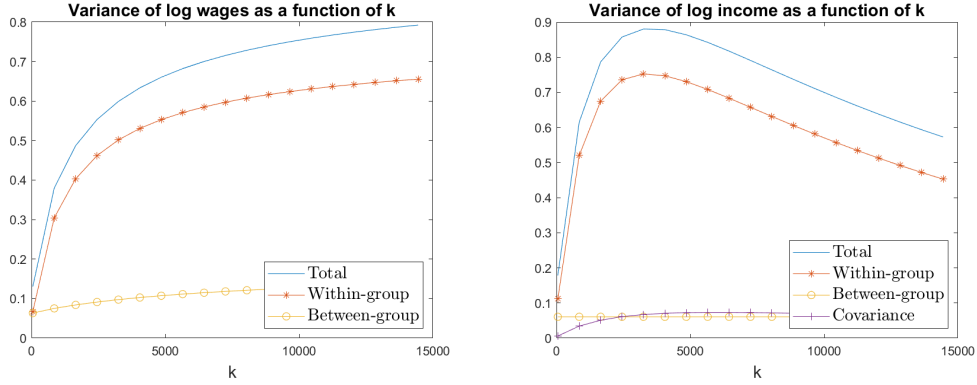
$$\text{Var}(\ln w) = \frac{\beta^2}{\tilde{h}^2} \text{Var}(p \mid p \geq \hat{p}) + \frac{e_L e_H}{(e_L + e_H)^2} (\ln a_H - \ln a_L)^2.$$

Heckman and Honoré (1990) show that the conditional variance is decreasing in \hat{p} if the productivity distribution belongs to the family of log-concave density functions, as shown in Figure 3.²¹ It follows that if productivity is distributed log-concave, our model delivers additional implications in terms of wage distributions. More precisely, higher entry costs k decrease equilibrium productivity threshold \hat{p} and market tightness θ , and as a result, reduce self-employment income \tilde{h} . Hence, the within-group variance of log wages is now higher due to a rise in the variance of ex-post firm productivity and a fall in \tilde{h} (see the within-group component on the left panel of Figure 4).

The between-group component of the variance in equation (18) can be written as $\pi_H(1 - \pi_H)(\ln a_H - \ln a_L)^2$, where $\pi_H = \frac{e_H}{e_L + e_H}$ is the share of high-type workers among wage employees. This between component is maximized when $\pi_H = \pi_L = 0.5$. The differential job finding probability between the two types of workers implies that, as the entry costs increase and the market tightness falls, the share of high-type workers among the employed, π_H , increases. Intuitively, in bad markets only those searching more intensely manage to find jobs. We find π_H to

²¹Proposition 1 in Heckman and Honoré (1990) shows that for log-concave distributions, $\frac{\partial \text{Var}(p \mid p \geq \hat{p})}{\partial \hat{p}} \leq 0$. The opposite is true for log-convex densities. We assume a log-concave density to fit the empirical relationship between wage levels and wage dispersion. Log-concave distributions include normal, exponential, logistic, gamma (for shape parameter greater or equal than 1), beta, extreme value, among others.

Figure 4: Wage and income inequality as a function of k



be close to 20 percent in our data, which means that a rise in k increases both the within- and between-group components of the log wage variance.

Figure 4 also shows that while worker heterogeneity might explain a large share of total wage variation in advanced economies, where entry costs are low, it is dwarfed by firm heterogeneity in least developed countries. As a result, our model predicts that higher entry barriers are associated with lower wage levels but higher wage dispersion, driven mostly by firm, rather than worker, heterogeneity.

Prediction 3: There is a non-monotone relationship between income and income inequality.

The determination of the size of the wage sector and wages is also intrinsically linked to income inequality, through reallocation of people between modern and traditional production and the resulting wage distribution. To see this, the variance of (log) income $\ln I$ can be written as

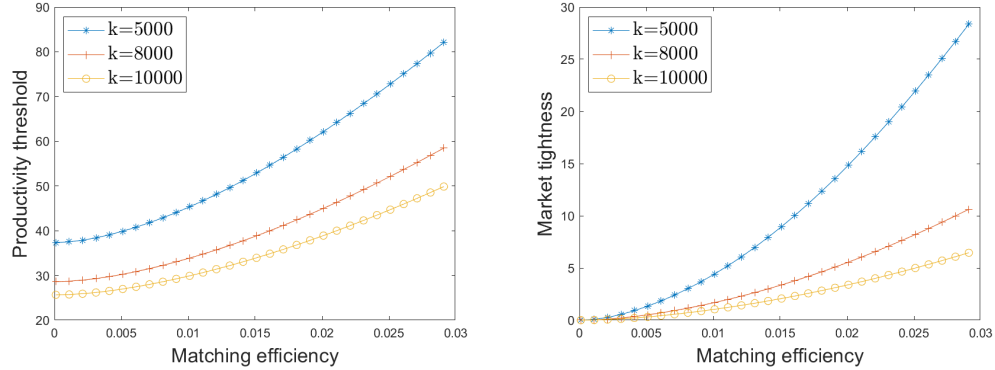
$$\begin{aligned} \text{Var}(\ln I) &= (e_H + e_L)\text{Var}(\ln \tilde{w}) + (e_H + e_L)(u_H + u_L)(E \ln \tilde{w} - \ln \tilde{h})^2 \\ &+ \alpha(1 - \alpha) \left[(\ln a_H - \ln a_L)^2 + 2(\ln a_H - \ln a_L)(E \ln \tilde{w} - \ln \tilde{h}) \left(\frac{u_L}{\alpha} - \frac{u_H}{1 - \alpha} \right) \right]. \end{aligned}$$

The variance of log income is comprised of four components. The first term arises

from firm heterogeneity among those employed. The second term is due to the percentage income gap between the wage workers and the self-employed. These two terms comprise the within-group component as they are the same for each worker type. The third component is the between-group variance that is due to differences in ability between the two types of workers, with the corresponding weights α and $1 - \alpha$. The last component is the covariance between the average log ability of workers and their corresponding income, depending on their labour status. This term reflects worker selection into the wage sector. That is, since low-ability workers are more likely to work in the home sector than their more able counterparts, the share of job seekers among the L -type workers is higher than among the H -type, i.e. $\frac{u_L}{\alpha} > \frac{u_H}{1-\alpha}$. This implies that the covariance is positive and income inequality is higher.

To understand what happens to income inequality when entry costs change, consider two extreme scenarios. First, suppose that there are no entry costs and no frictions so that everyone is employed in the wage sector. Only the most productive firms survive in this economy, hence there is no heterogeneity in productivity, nor in wages within the worker type. Income variance is close to zero in this case. The opposite case is when the frictions are so high that no firm enters the market and everyone is self-employed, receiving $\tilde{h}a_j$. Also in this case, the only source of income dispersion is the difference in worker ability. Departing from either extreme will result in a rise of income inequality, generating an inverse U-shape relationship between the levels of income and inequality (see the right panel of Figure 4). For Sub-Saharan African countries, with a very small wage sector, it implies that moving people away from self-employment would increase income inequality. This is similar in spirit to early arguments explaining the Kuznets curve, such as Robinson (1976).

Figure 5: Complementarities between entry costs k and matching efficiency m



An increase in the matching efficiency m has a similar effect on equilibrium variables as a reduction in entry barriers. Although the model does not allow for a full analytical characterization in this case (see online Appendix B.2 for details), our estimated parameter values suggest that a reduction in labour market frictions leads to a larger wage sector, greater average productivity and wages, and a lower wage dispersion.

Prediction 4: There are complementarities between parameters k and m

We simulate the model using the baseline parameters that are obtained from the data on Uganda (see Section 5 for details) and vary matching efficiency and entry cost parameter values. For each set of parameters, we solve for equilibrium market tightness and the productivity threshold. Figure 5 shows that improvements in labour market efficiency m generate a larger increase in the productivity threshold \hat{p} and in the market tightness θ when the entry costs are lower (the line with blue stars on the graph). Analogously, entry barriers are more detrimental for the wage sector in countries with a higher degree of labour market frictions.

4 Supporting empirical evidence

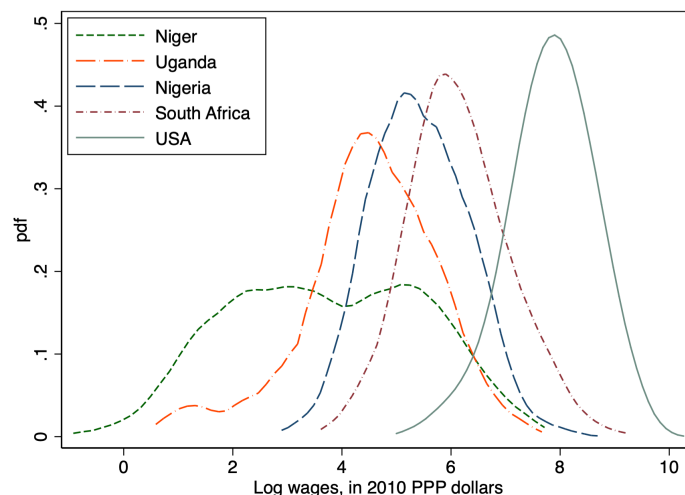
Our model predicts that entry barriers limit the reallocation of workers from self-employment to the wage sector, reduce average productivity and wages, and sustain a high degree of wage dispersion in the market. In this section, we use household survey data for a number of countries in Sub-Saharan Africa to provide empirical support for these predictions (see online Appendix A for data description). The evidence below is based on cross-country comparisons, whereas in online Appendix C.2 we demonstrate for the case of Uganda that these relationships can hold also across regions within a country, provided that there is enough inter-regional variation.

(a) Paid employment, wages and wage dispersion

For illustration purposes, we plot log wage densities for four SSA countries and the US. Figure 6 shows that as the distributions move to the right, they become more compressed, suggesting that wage variance decreases with mean wages (and GDP per capita).

Some of the observed variation in wages can be explained by worker's demographics, such as education, experience, work industry and their place of residence. For example, it has been well documented in the literature that there exists a large and persistent productivity gap across sectors (especially between agriculture and manufacturing) in developing countries (see Gollin, Lagakos and Waugh, 2014 for recent evidence). Hence, in the next step we run a wage regression for each country to control for observed worker characteristics, urban and rural areas, regions, and industries and then report residual wage dispersion. Figure 7 shows that wage dispersion remains substantial after controlling for observables and, more importantly, the negative correlation between mean wages and residual wage dispersion

Figure 6: *Log wage densities*



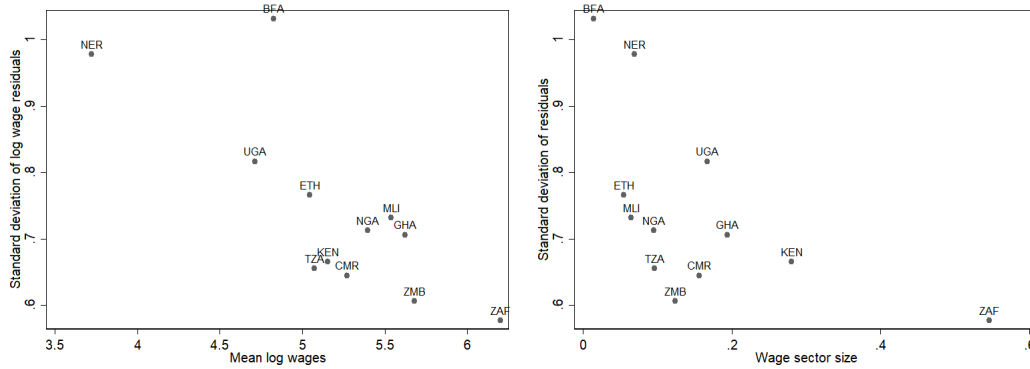
Source: Authors' computations based on household surveys for Niger 2011, Uganda 2010, Nigeria 2010, South Africa 2008 and USA 2010 (see online Appendix A for details). Monthly wages deflated using yearly CPI (with 2010=100) and converted into 2010 PPP dollars. The sample consists of workers aged between 15 and 65, excluding public sector employees. Note that Niger's GDP per capita in 2010 PPP dollars was \$798, Uganda's was \$1,509, Nigeria's was \$5,046, while South Africa's and the USA's was \$11,786 and \$48,374, respectively.

holds.²² Moreover, the highest wage dispersion seems to be coming from countries with smaller wage sectors.

One hypothesis for why residual wage dispersion is higher in poorer countries might be a higher variance in unobserved workers' characteristics. For example, if the quality of education is different within a country then wages will vary across workers even after controlling for workers' education. We know from Section 2 that better educated workers are more likely to get a job in the wage sector and they might be responsible for a large fraction of observed dispersion. However, we do not find strong evidence in favour of this hypothesis. First, after we exclude the workers with upper secondary, vocational and post-secondary education from

²²Figure C.1 in online Appendix C.1 shows (for a larger number of countries) that even within the manufacturing sector alone there appears to be a strong negative relationship between country's wage inequality and the level of GDP per capita or the average wage.

Figure 7: Residual (log) wage dispersion, mean wages and the size of the wage sector



Source: Authors' computations based on household surveys (see online Appendix A for details). Note: For wage and wage sector size measures, the sample is limited to 15-65 year old private sector employees (as opposed to self-employed and unpaid family members). In South Africa, self-employed also include unemployed individuals. Monthly wages are deflated using CPI and expressed in constant 2010 PPP dollars. Residual wage dispersion is obtained from the residuals of a wage regression that controls for demographics (gender, age, age squared, marital status, education), regions, urban status, and industry.

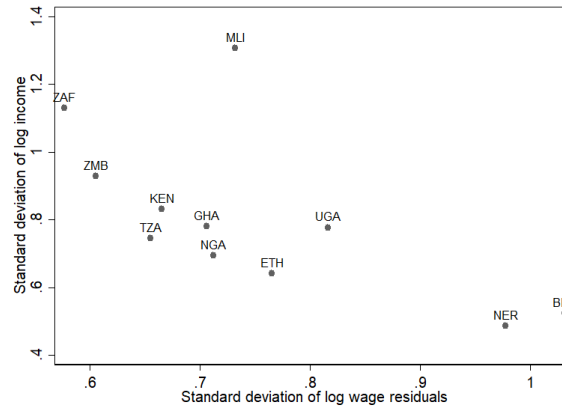
the analysis (based on the selection equation in Table A.1 in online Appendix A), the resulting reduction in wage dispersion is negligible, being below 5% in most countries. Second, as the wage sector expands, we would expect to see an inflow of less productive workers into wage employment, leading to an increase in the dispersion of abilities and, thus, wages. This contradicts our empirical finding that a larger wage sector is associated with a lower wage dispersion. This empirical evidence suggests that firm heterogeneity plays a larger role than worker heterogeneity in determining wage dispersion in least developed countries.²³

(b) *Wage and income inequality*

The model illustrates how a negative correlation between wage levels and wage dispersion can co-exist with a positive correlation between average income and

²³This is in line with recent empirical evidence that shows that changes of earnings inequality in many countries, either increasing or decreasing, can be accounted for primarily by changes between and not within firms, suggesting a larger role for firm (as opposed to worker) heterogeneity. See, for example, Song, Price, Guvenen, Bloom and von Wachterk (2018) for the US and Alvarez, Benguria, Engbom and Moser (2018) for Brazil.

Figure 8: *Wage and income inequality*



Source: Authors' computations based on household surveys (see online Appendix A for details).
Note: The dispersion of log income is calculated based on household consumption diaries.

income inequality for poor countries (reflecting the left part of the Kuznets curve). In order to check whether this theoretical prediction holds in the data, we compute the standard deviation of log income using monthly household consumption data from the same household surveys (where that information is available).²⁴ Figure 8 shows that countries that exhibit high levels of (residual or raw) wage dispersion are the ones with low income inequality.

The evidence presented above shows a general pattern in least developed countries: the wage sector in poor economies is characterized by high levels of wage dispersion. Moreover, wage dispersion is inversely related to the mean wage, the wage sector size, and income inequality.

(c) *Measurement error*

One concern is that in poorer countries the differences in wage dispersion may be driven by non-classical measurement error, for example due to poor survey

²⁴Since it is very difficult to measure household income in poor countries, where the majority of the workforce is employed in (often irregular) self-employment activities, it is typical to use consumption diaries to proxy for income. The same methodology is used by the World Bank in computing the Gini index for income.

quality. If that were the case, systematic errors should be present in wages as well as in household income, thus leading to a positive relationship between income and wage dispersion. This hypothesis is disputed by the data, as Figure 8 shows that countries with greater wage dispersion exhibit, on average, lower income dispersion. Another source of non-classical measurement error in wages could be the frequency of payment, as documented in Borjas (1980). If countries differ substantially in how wage earners report their earnings (e.g. daily vs monthly payments) and those frequencies are more subject to measurement error, then the observed differences in residual wage dispersion could be explained by measurement, rather than by entry barriers or labour market frictions.

In Table C.1 in online Appendix C.1, we observe some country differences in frequency of payment among wage earners, even though in all cases most workers report monthly wages (from almost 42% in Niger to around 80% in Nigeria). The rest is, in most cases, divided between daily, weekly or fortnightly wages. At the bottom of the table we show that controlling for the frequency of payment does reduce residual wage dispersion (in some cases by more than 10%). However, the ordering of countries does not change and the remaining dispersion remains substantial.

In Figure C.2, also in online Appendix C.1, we show wage distributions by frequency of payment for the two countries with greater wage dispersion (namely, Niger and Uganda), which also have more variation in the reported frequency of payment (e.g. a large number of workers report daily wages). When controlling for demographics, industry, region, urban/rural areas, we find that the residual wage distributions by frequency of payment overlap considerably, particularly in Uganda when there seem to be virtually no differences according to how workers report their earnings.

Table 2: *Entry costs in selected countries*

Country	GDP per capita in 2010 PPP	Cost of		Bank branches
		Starting a business	Electricity connection	
Niger	812	118.7	7,996	0.9
Ethiopia	1,075	268.4	4,913	1.4
Uganda	1,585	84.4	7,022	2.5
Tanzania	2,228	121.7	2,861	2.0
Nigeria	5,085	73.8	1,436	6.6
South Africa	11,973	5.9	875	9.9
Sub-Saharan Africa	4,739	97.6	6,075	6.4
USA	49,479	0.7	17	35.4

Source: World Bank's Doing Business database 2010. Cost of starting a business includes all official fees and fees for legal or professional services if such services are required by law. Cost of getting electricity is the cost required for a business to obtain a permanent electricity connection for a newly constructed warehouse. Costs of entry and electricity connection are recorded as a percentage of the economy's income per capita. Bank branches is the number of commercial bank branches per 100,000 adults. GDP per capita is in 2011 international dollars.

Taken together, we interpret these pieces of evidence as suggestive that the pattern of wage distributions cannot be ascribed to non-classical measurement error.

(d) *Entry costs*

Entry costs k in the model can be interpreted as any barriers that prevent firms from entering the market and thus reduce competition, such as red-tape regulations, borrowing constraints (e.g. the collateral required to get a credit), or access to advanced technology. The effects of entry barriers in the model on wages and employment are fairly intuitive. The entry costs endogenously determine the number of firms in the market. That is, more binding entry constraints reduce the wage sector size in the economy and put downward pressure on wages.²⁵ To illustrate this relationship in the data, we use a series of different indicators of entry costs and access to credit drawn from the World Bank's Doing Business survey for 2010.

The first measure presented in Table 2 is legal costs of starting a business. We

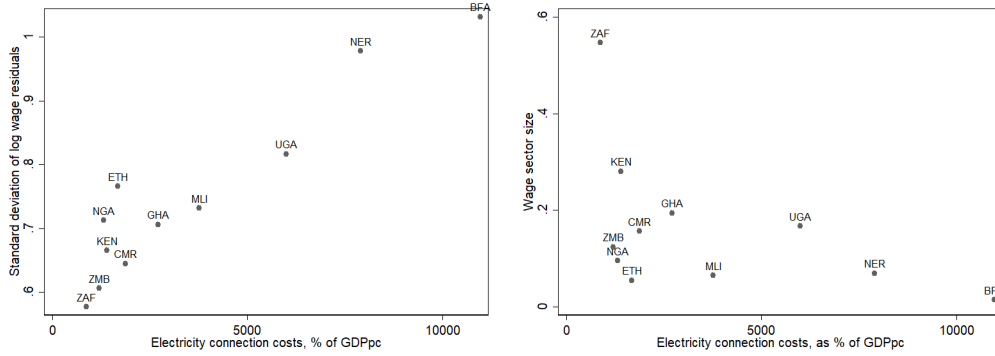
²⁵This finding is in line with other empirical and theoretical work (see for instance, Ghatak and Nien-Huei Jiang (2002), Ayyagari, Demirgüç-Kunt and Maksimovic (2008), Djankov *et al.* (2002) and Herrendorf and Teixeira (2011), among others). Consistent with the mechanism we propose in our model, McKenzie and Paffhausen (2019) look at firm exit in developing countries and find that richer countries tend to have greater firm death rate and that exiting firms tend to be less productive.

find that the entry costs in Sub-Saharan Africa are almost 100 times higher than in the US and countries that exhibit a relatively low measure of entry costs tend to have higher GDP per capita. Figure C.1 in Appendix C.1 shows that a negative relationship between the entry costs and a level of income is a general pattern that holds for other countries as well.

The legal fees might not be the best indicator for firm entry costs in least developed countries. On one hand, the legal fees may overstate the actual costs as many of the enterprises in SSA are informal and hence are not subject to many government regulations. On the other hand, starting a business might involve bribes and unofficial expenses that will not be captured in the legal fees. Empirical evidence (based on Enterprise Survey data) suggests that poor electricity supply is one of the major obstacles to firm entry and operation in least developed countries and under a half of the firms adopt a stand-alone power generator to cope with power outages, implying that accessing the market involves a substantial fixed cost. Hence, we use the costs of getting electricity as an alternative measure of entry costs and find a similar pattern. The third measure we use is the prevalence of commercial bank branches as a measure of credit access. We find that a higher degree of financial development is associated with higher incomes (and a larger wage sector as well).

Finally, our model predicts that a higher level of entry costs not only reduces the size of the wage sector, but also increases wage dispersion, all else equal. Figure 9, which plots electricity connection costs versus residual wage dispersion and the wage sector size for the same twelve countries, confirms this prediction.

Figure 9: Correlation between wage employment, wage dispersion and entry costs



Source: Authors' computations based on household surveys (see online Appendix A for details). The electricity connection costs are obtained from the World Bank's Doing Business Survey for 2010-2016.

5 Using the model to quantify barriers to job creation and other labour market outcomes

5.1 Estimation

In this section we illustrate how our structural model can be used to quantify the impact of different factors on wages and job creation in a way that a reduced form approach cannot capture. To do so, we estimate the model using the indirect inference approach (see Gourieroux, Monfort and Renault (1993)), which essentially minimizes a distance criterion between key moments from actual and simulated data. We estimate the model for six countries, for which we have panel data: Ethiopia, Niger, Nigeria, Tanzania, Uganda, and South Africa. Below, we briefly outline the estimation procedure and provide intuition for parameter identification, while more details on estimation, the choice of empirical moments and the sensitivity analysis can be found in online Appendix D.

The model is solved under the assumption that the economy is in steady state. The equilibrium of the model can be fully characterized by a vector of 13 parameters: $(r, \delta, \eta, c, k, m, \sigma, \beta, A, \gamma, \alpha, a_H, s_H)$, where σ is a standard deviation of the ex

ante firm productivity $\Gamma(p)$. Due to the lack of available data, we make a number of assumptions to reduce the dimensions of the model. First, we set the interest rate r at 1.25% (where a unit of time is one month), implying an annual rate of approximately 15%. The interest rate is relatively high to reflect the fact that borrowing constraints are more significant in the SSA countries for both firms and workers. Second, it is impossible to separately identify the matching function elasticity with respect to vacancies, η , from matching efficiency, m , without data on vacancies. Hence, we assume that $\eta = 0.5$, as is common in the literature.²⁶ Finally, the location parameter of the underlying productivity distribution $\Gamma(p)$ cannot be identified separately from the outside option \tilde{h} using only wage data. Instead of normalizing the mean of Γ distribution, we choose to use an exponential distribution that is characterized by only one parameter, σ , equal to both its mean and standard deviation.

We normalize a_L to one and interpret a_H as relative productivity of a high type. To identify a more productive type in the data, we use a selection regression presented in Table A.1 in online Appendix A. In particular, we group all education categories that have a statistically significant positive effect on the probability to be employed in the wage sector. The H -type includes predominantly workers with vocational and post-secondary education and in some cases also those with secondary education (the exact composition is shown in Table A.2 in online Appendix A).²⁷

Home production income in the model is characterized by two parameters: aggregate productivity A and returns to scale γ . In the data, it is difficult to measure

²⁶The matching function elasticity with respect to vacancies is usually estimated in the range of 0.3 – 0.5 (see Petrongolo and Pissarides, 2001). In online Appendix D.4 we show how the estimates would change if we assumed $\eta = 0.3$ instead.

²⁷While primary education also increases the likelihood to get a wage job in Niger (relative to no education), the coefficient is relatively small. For consistency with other countries, we assign primary education to the low skill group.

income from home production as it often includes in-kind payments or working for a household business or farm without pay. Given that the majority of self-employed activities in developing countries are related to subsistence farming, we use aggregate agricultural production to estimate γ . In particular, we regress agricultural value added on the number of workers employed in agriculture across SSA countries to obtain the elasticity of labour. We run a number of specifications using different output and control variables and our preferred value is $\hat{\gamma} = 0.246$.²⁸ This value of γ might seem to be very low, yet it is appropriate for the types of self-employment occupations that we have in mind (traditional farming, casual jobs, petty retail, etc.) and similar values have been reported in the literature (see Aragón and Rud, 2015). We assume that the returns to scale parameter γ is the same across the six countries we analyse, while the overall home sector productivity A is allowed to vary and is estimated within the model.

The remaining vector of parameters $\vartheta = (c, k, m, A, \sigma, \beta, \delta, \alpha, a_H, s_H)'$ is estimated by matching the following ten moments: the fraction of low-educated workers in the population, self-employment rate and mean log wages for the two types of workers, the standard deviation of log wage residuals, the transition rate from wage employment into the home sector, the labour share in value added from the Enterprise Survey data, the entry costs per worker from Doing Business, and the average hiring costs of one month of wages (see Table 3, where countries are ordered by their mean wages). While it is not possible to associate all individual parameters with individual moments as they are determined together in the model, below we provide intuition for identification.

The variance of log wages within a worker type, $Var(\ln \tilde{w})$, is determined pri-

²⁸More details on the exact specification can be found in online Appendix D.1. In addition, in online Appendix D.4 we run a robustness check for our counterfactual experiments with alternative values of γ and find virtually the same results.

Table 3: *Empirical targets*

	Niger	Uganda	Tanzania	Ethiopia	Nigeria	South Africa
Fraction of L -type, α	0.946	0.936	0.944	0.955	0.886	0.607
Self-employment rate						
L -type workers, $\frac{u_L}{\alpha}$	0.948	0.850	0.938	0.913	0.910	0.440
H -type workers, $\frac{u_H}{1-\alpha}$	0.806	0.557	0.740	0.560	0.701	0.295
Average log wages						
L -type workers, $E(\ln w a_L)$	4.09	4.59	4.93	5.01	5.22	5.83
H -type workers, $E(\ln w a_H)$	5.56	5.49	5.56	5.42	5.62	6.66
Std. dev. of res., $sd(\ln \tilde{w})$	0.962	0.799	0.701	0.762	0.695	0.579
Job separation rate, δ	0.036	0.031	0.019	0.018	0.020	0.024
Labor share, $E\left(\frac{w}{p}\right)$	0.136	0.219	0.247	0.187	0.232	0.424
Entry costs, $\frac{k}{\gamma}$	53.88	73.10	12.55	6.41	5.98	1.83
Number of observations, N	14,548	19,798	24,479	16,945	38,160	63,176

Note: The two worker types are identified from the selection equation into the wage sector. Table A.2 in online Appendix A shows which educational groups comprise the high type for each country. The self-employment share in South Africa includes unemployed workers. Mean log wages are presented in constant 2010 international dollars. Log wage residuals are obtained after controlling for workers' age, age squared, sex, marital status, education, regions, urban and rural areas, and industry. Job separation rate is equal to the share of wage workers moving into self-employment between the survey waves, converted into monthly transition rates. Labour share $\frac{w}{p}$ is obtained as the firm-level ratio of labour costs to value added based on the Enterprise Survey data. The costs of electricity connection (see Table D.1 in online Appendix D.1) are expressed in per worker terms and converted into monthly income. The number of observations is the number of workers in wage employment or self-employment drawn from the household surveys. The last target (not included in the table) is the average hiring costs, $\frac{c}{q}$, that is chosen to be equal to one month of wages. Moments are calculated based on the following datasets: Ethiopia Socioeconomic Surveys 2013 and 2015; Niger National Surveys on Household Living Conditions and Agriculture 2011 and 2014; Nigeria General Household (Post-Planting and Post-Harvest) Surveys 2010, 2012 and 2015; South Africa Labour Force Surveys March and September 2007; Tanzania National Panel Survey 2008, 2010 and 2012; Uganda National Household Surveys 2010, 2011, 2012 and 2014.

marily by the variance of underlying productivity, σ^2 . Average wages for the low type help us identify the outside option \tilde{h} , that in turn determines A . The difference in mean log wages between the two types of workers identifies their relative productivity, a_H . The transitions between wage employment and self-employment determine the job destruction rate, δ . The matching efficiency, m , is used to fit the self-employment share of less educated workers, while the self-employment rate among better educated workers determines their relative search intensity, s_H . Workers' bargaining power, β , is related to the labour share in the model. Finally, we back out the entry costs, k , from electricity connection costs, and the vacancy

Table 4: *Estimated parameters in the model*

Country	Estimates									
	σ	A	β	δ	m	α	a_H	s_H	k	c
Niger	761.7 (38.5)	24.6 (8.7)	0.111 (0.006)	0.036 (0.002)	0.0023 (0.0001)	0.946 (0.002)	4.33 (0.37)	4.37 (0.43)	4259.8 -	0.37 -
Uganda	734.0 (19.1)	71.8 (21.5)	0.155 (0.004)	0.032 (0.003)	0.0082 (0.0007)	0.936 (0.002)	2.47 (0.09)	4.50 (0.27)	12912.2 -	2.03 -
Tanzania	902.5 (22.3)	152.0 (46.3)	0.154 (0.005)	0.019 (0.001)	0.0009 (0.0000)	0.944 (0.001)	1.88 (0.08)	5.32 (0.36)	1482.9 -	0.15 -
Ethiopia	1366.9 (42.1)	130.1 (39.4)	0.120 (0.005)	0.018 (0.001)	0.0008 (0.0000)	0.955 (0.002)	1.52 (0.07)	8.26 (0.64)	1218.8 -	0.08 -
Nigeria	1301.0 (27.4)	200.3 (60.0)	0.141 (0.003)	0.021 (0.002)	0.0010 (0.0001)	0.886 (0.002)	1.49 (0.03)	4.30 (0.16)	1351.6 -	0.13 -
South Africa	1028.0 (5.5)	323.5 (83.1)	0.265 (0.001)	0.024 (0.000)	0.0056 (0.0001)	0.607 (0.002)	2.30 (0.01)	1.88 (0.03)	2202.0 -	0.65 -

Note: Asymptotic standard errors are given in parentheses. The standard errors for k and c are not available, as the empirical moments they are based on - the electricity connection costs and the ratio of the average hiring costs to wages - do not exhibit variation in the data.

cost parameter, c , from the average hiring costs.

5.2 Results

5.2.1 Estimated parameters

Table 4 shows the estimated parameters and their standard errors. One noticeable feature is our estimates for σ . A natural way to think about differences in the degree of dispersion in wages and productivity across countries is that they stem primarily from differences in underlying productivity. This, however, is not supported by our findings. For example, Ethiopia and South Africa are not that dissimilar in terms of their ex-ante productivity distribution but are miles away in terms of their ex-post productivity and wage distributions due to the varying degrees of market frictions. This result is in line with other studies that show that misallocation of resources due to frictions lowers aggregate productivity and growth.²⁹

²⁹See, for example, Hsieh and Klenow (2009). Bartelsman *et al.* (2013) provide empirical evidence on importance of distortions for within-industry productivity dispersion based on the firm-level data for the US, UK, Germany, France, Netherlands, Hungary, Romania, and Slovenia. They show that distortions not only affect the allocation of resources across firms, but also the selection of firms

While the relationship between underlying firm productivity dispersion and mean wages (or GDP per capita) is not monotone in our sample, the estimated values of home sector productivity, A , are increasing in the level of development. We compare our implied values of h to poverty statistics for each country. For example, our estimates generate self-employment income of about \$7, \$21, and \$35 a month for Niger, Uganda and Ethiopia, respectively, which aligns well with the fact that about 50%, 45%, and 34% of the corresponding country's population live below \$1.90 a day (World Development Indicators Database).

We find very low estimates of the matching efficiency parameter, m . For comparison, a recent study by Sahin, Song, Topa and Violante (2014) estimates the aggregate matching efficiency parameter in the US to be 0.94, while Albrecht, Robayo-Abril and Vroman (2017) use 0.25 for Colombia. Our parameter values imply that, with the exception of Uganda, South Africa has two to six times more efficient labour market than the poorer economies in our sample. The implied job finding probability for low-skilled workers, λ , ranges from 2% to 7% yearly in the poorest countries in our sample. That is, less than three out of a hundred own-account workers get a paid job in a year's time in Niger, Ethiopia and Tanzania, less than four in Nigeria and less than seven in Uganda. While the job finding rates for the high-skilled workers are 4-6 times higher, given the overall small fraction of this group of workers in the population the overall outflow from self-employment is extremely low. Even for South Africa, where the transition rates are much higher than in the other SSA countries, the job finding rate (out of self-employment and unemployment together) is about 3% a month for less educated workers and 5% for their more educated counterparts, almost 10 times lower than in the US.³⁰ These

producing in each market.

³⁰Note that, while actively searching unemployed workers in South Africa have a higher job-finding rate than the self-employed do, the rates are not substantially different. The overall (over two education types) annual transition rate into wage employment is 61% and 57%, respectively.

results suggest that labour market inefficiencies are prevalent in this region and labour mobility is extremely low.

The estimates of workers' bargaining power parameter β range from 0.11 in Niger to 0.27 in South Africa. These values reflect the fact that workers' bargaining position is relatively weak in poor countries due to a lower degree of unionisation³¹ and relatively low levels of workers' human capital.

In the following subsections, we use these parameters to understand the role of entry costs and labour market frictions in shaping labour market outcomes.

5.2.2 The effect of frictions on labour market outcomes

(a) Wage variance decomposition: the role of k and m .

The differences in wage inequality across countries stem from various sources, such as underlying differences in firm productivity, worker bargaining power, educational composition, home sector productivity and market frictions. Our estimates allow us to quantify the relative importance of each of these channels and the interactions between them. First, our results suggest that most of the variation in wages is driven by firms - educational composition explains from as little as 5% in Ethiopia and up to 30% in South Africa. This is consistent with our model's prediction that, while worker heterogeneity might explain a large share of total wage variation in advanced economies, it is dominated by firm heterogeneity in the least developed countries where market frictions are severe (see Figure 4). Therefore, for the rest of our analysis we focus on the drivers of within-worker type variation in wages.

To illustrate how the model can be used for variance decomposition, we choose

³¹The Global Wage Report 2010/2011 (ILO) shows that the share of unionised workers as a fraction of the workforce is 1.1 percent in Niger and Uganda, 2.2 percent in Tanzania, and 25 percent in South Africa.

Table 5: *Within-type wage variance decomposition*

	Variance of log wage rates, $V(\ln \tilde{w}(p))$				
	Baseline	With k_{SA}	With m_{SA}	With $k_{SA} \& m_{SA}$	All SA pars
Niger	0.926	0.892	0.870	0.810	0.335
Uganda	0.638	0.468	0.660	0.522	0.335
	Percentage of the relative variance gap explained by differences in				
		k	m	$k \& m$	Other pars
Niger		5.7%	9.5%	19.6%	80.4%
Uganda		56.3%	-7.2%	38.3%	61.7%

South Africa as a benchmark economy and simulate the counterfactual variance of log wage rates, $\tilde{w}(p)$, that would arise in other countries if they had SA's parameters. We focus on Uganda and Niger as these countries have the largest wage variance gap compared to South Africa.

Table 5 shows how the gap in wage dispersion relative to South Africa can be accounted for by differences in market frictions versus differences in other parameters. For example, the variance of log wages in Niger decreases from 0.926 to 0.892 if it had the same entry costs as South Africa. The corresponding fall in the variance is much larger in Uganda, given its prohibitively high levels of entry costs: down from 0.638 to 0.468. Changing all parameters to South Africa's values would reproduce the variance of log wage rates in the benchmark economy of 0.335.

In Uganda, around 38% of the relative variance gap can be explained by differences in k and m relative to South Africa, while the remaining 62% are attributed to differences in other parameters. In Niger, more than one fifth can be explained by frictions and around 80% by other parameters. The role of the entry costs for wage inequality is in stark contrast between Niger and Uganda, explaining around 6% of the variance gap in the former and around 56% in the latter country. The opposite is true for matching frictions that account for almost 10% of the variance gap in Niger,

but contribute negatively in Uganda. That is, the matching efficiency in Uganda is actually higher than in South Africa and thus moving to SA's level does not constitute an improvement. Furthermore, Table 5 highlights the fact that the effects of k and m are not linear and the combination of these parameters explains more than the sum of their individual effects. For example, we can attribute the remaining 4.4% of the gap in Niger ($19.6\% - 9.5\% - 5.7\% = 4.4\%$) to the interaction between market frictions. The sensitivity analysis in online Appendix D.4 suggests that our finding that about a half of the relative variance gap in wage rates in Uganda and one quarter in Niger can be explained by differences in market frictions is robust to different values of pre-determined parameters (see Table D.5).

(b) Productivity and wage gains are constrained by frictions

In the model, as the frictions are reduced and the wage sector becomes more competitive, less productive firms are forced out of the market, thus increasing the average levels of productivity and wages. This mechanism is similar in spirit to the basis of creative destruction models of, for instance, Aghion and Howitt (1992) and Grossman and Helpman (1991) that suggest that productivity growth is driven primarily by entering firms that adopt new technologies and replace less productive older firms. Using the estimated model parameters, we can compare the ex-ante productivity distribution, $\Gamma(p)$, and the ex-post productivity distribution, $\Gamma(p|p > \hat{p})$, to quantify the gains that arise due to firm entry and competition.

Similarly, we can compare ex-ante and ex-post wage distributions as wages are a linear combination of firm productivity and worker's outside option. Again, we focus on the wage rate as average wages increase proportionally for both worker types. For ex-ante wage rate distribution we assume that in the limit - when the degree of frictions is the highest - almost all the workforce are employed in home production, i.e. $L_H \rightarrow \alpha a_L + (1 - \alpha)a_H$. Then, the outside option of a worker

Table 6: *Ex-ante and ex-post distributions of firm productivity and wages*

	Average productivity			Average wage rate		
	ex-ante	ex-post	% Δ	ex-ante	ex-post	% Δ
Niger	762	772	1.4	89	90	1.7
Uganda	734	760	3.6	128	134	5.1
Tanzania	902	948	5.0	169	178	5.3
Ethiopia	1367	1408	3.0	192	199	3.9
Nigeria	1301	1363	4.7	224	237	5.8
South Africa	1028	1176	14.4	316	406	28.7

becomes $\tilde{h} \rightarrow \gamma A(\alpha a_L + (1 - \alpha)a_H)^{\gamma-1}$.

Table 6 presents the percentage difference in average firm productivity, p , and wage rate, $\tilde{w}(p)$, between each country's ex-ante and ex-post distributions. Again, our results indicate that frictions play a large role in shaping labour market outcomes. Average productivities and wage gains under current labour market conditions are modest in the poorer countries in the sample averaging from 2 to 6 percent. With the market frictions being significantly lower in South Africa, it is not surprising to find a 14 percent increase in average productivity and a 29 percent gain in average wages relative to the initial distributions.

(c) The responsiveness of outcomes to changes in parameters.

In this subsection, we use the model to simulate changes in the entry costs, k , the labour market efficiency, m , self-employment productivity, A , and underlying productivity dispersion, σ , and the fraction of less educated workers in the population, α , to illustrate what happens to the size of the wage sector, wage levels and dispersion, as well as the overall income and income inequality. Recall that individual wages are a product of the wage rate, $\tilde{w}(p)$, and workers' ability level, a . Therefore, it is sufficient to focus on the average wage rate as the average wages for each type of workers will move in tandem with it (if the ability premium a_H is un-

Table 7: Elasticities of outcome variables with respect to changes in parameters

	Elasticity with respect to	Empl. share	Average wage rate	St. dev of log wage rate	Average income	St. dev of log income
Uganda	k	-0.94	-0.03	0.06	-0.55	-0.30
	m	1.50	0.04	-0.06	0.87	0.46
	A	-0.03	0.14	-0.24	0.45	-0.36
	σ	0.98	0.88	0.20	1.11	0.66
	α	-4.43	-0.01	0.01	-4.34	-2.94
South Africa	k	-0.35	-0.15	0.18	-0.38	0.14
	m	0.68	0.27	-0.33	0.73	-0.28
	A	-0.04	0.30	-0.35	0.37	-0.32
	σ	0.39	0.84	0.18	1.01	0.18
	α	-0.55	-0.02	0.03	-1.19	-0.16

Note: The elasticity with respect to α shows the percentage change in the outcome variable if α increases by one percentage point.

changed). Similarly, we look at the effect on within-worker type wage dispersion, $SD(\ln \bar{w})$. The results are expressed in terms of elasticities. Table 7 summarizes our estimates for Uganda and South Africa for the exposition brevity, while the results for other countries can be found in online Appendix D.

In line with our analytical results, a reduction in frictions (a reduction in k or a rise in m) leads to a larger wage sector, higher wages and a lower residual wage dispersion (i.e. controlling for worker's ability). While many of our results confirm the existing wisdom in the development literature, below we focus on what we consider to be new insights that come directly from our use of a unifying modelling framework. Note that these results are qualitatively robust to the choice of calibrated parameters (see online Appendix D.4).

1. *A reduction in labour market frictions has a larger impact on wage employment than a fall in entry costs.* A one percent rise in the matching efficiency parameter leads to 0.68 percent increase in the size of the wage sector in South Africa and 1.5-1.8 percent increase in the other countries. A reduction in the entry costs is about half as effective.

2. *A non-linear relationship between the reduction in frictions and income inequality.* This finding is consistent with our theoretical results: a fall in the degree of market frictions (be it a reduction in k or an increase in m) causes reallocation of workers from low-earning self-employment to the higher-earning wage sector. Since the majority of workers are self-employed in the poorer countries in our sample, this leads to an increase in income inequality in Uganda, while the opposite is true for South Africa.

3. *A substantial impact of self-employment productivity on mean wages, income, and inequality.* Productivity in self-employment activities is generally lower in poorer countries and exogenous positive shocks in that sector have been used to explain structural change (see for example Lewis, 1954 and Matsuyama, 1992). Even in South Africa, where the share of self-employed workers (those directly affected by changes in A) is relatively low, the rise in workers' outside option has a substantial effect on mean income (0.37 percent increase) and income dispersion (0.32 percent drop). The magnitude of its impact on mean income is about 1.2 times higher in Uganda. In addition, it has a large effect on wages: a one percent increase in A causes about 0.1-0.3 percent rise in the mean wage across countries and about a 0.15-0.35 percent drop in the standard deviation of log wage rates.

Although higher self-employment productivity has a negative impact on job creation³², quantitatively these effects are close to zero.

4. *The increase in underlying productivity dispersion increases average wages, wage employment and income, at the expense of higher inequality.* Consider an increase

³²This is similar to multiple equilibria models of development, such as Banerjee and Newman (1993) and Ghatak and Nien-Huei Jiang (2002), where high relative productivity in the self-employment sector may be associated with an equilibrium dominated by a self-sufficient agricultural sector and cottage industries that curb the growth of the modern sector.

in the underlying productivity dispersion that might be driven by changes in capital intensity, technology adoption, or opening to trade. A one percent increase in σ has a similar effect to a one percent reduction in k on the size of the wage sector, but it is more effective for increasing average wages and income. At the same time, it generates higher wage and income inequality, with the latter effect being especially large for poorer countries.

5. *The lack of skills is a major barrier to job creation.* The fall in wage employment associated with a 1% point increase in the share of low skilled workers, α , varies from 2.6% in Nigeria and 4.4% in Uganda to 8% in Niger. The deficiency of highly educated workers, which are both more productive and more successful in finding a job, makes it less likely for firms to encounter high-type workers and discourages them from posting vacancies. In South Africa, where over one third of the population have at least upper secondary education level, wage employment is less sensitive to changes in workers' composition. Thus, our model suggests that in countries where high-skilled labour is scarce, increasing access to education or providing training is an important channel for job creation and raising overall income.³³ In addition, this policy generates a positive spillover effect on less educated workers, as an increase in vacancies means that they too are more likely to find a paid job. One caveat is that the wage premium a_H and the relative search intensity of high-skilled workers s_H are unchanged in the simulation. If the supply of highly educated workers changes significantly we would expect the educa-

³³This is consistent with findings from the literature assessing the effects of vocational training programmes in developing countries. Alfonsi *et al.* (2019) find that a training intervention among young workers in Uganda increases the likelihood of employment by around 15-20%. In relatively richer countries, the effects are more modest. For example, for Colombia, Attanasio, Kugler and Meghir (2011) find effects of 7% among women (and no effects among men), while Hirshleifer, McKenzie, Almeida and Ridao-Cano (2016) and Card, Ibararán, Regalia, Rosas-Shady and Soares (2011) find negligible employment effects in Turkey and Dominican Republic, respectively.

tion premium to fall and the overall effect on the wage sector size and income to be smaller.

6 Policy discussion

In this section, we discuss what policy instruments can be used to reach certain development targets - be it the size of the private sector wage employment, the overall income, or income inequality - and how they can be linked back to our theoretical framework. In general, a structural model is useful for extrapolating the results of existing studies that use reduced-form estimation or field experiments, as well as for measuring potential general equilibrium effects of a given policy.

An obvious candidate policy is improving firms' productivity through, for example, international trade or foreign direct investment (FDI) inflows. We can represent it in our model through an increase in the underlying productivity dispersion parameter, σ . The conjecture is that foreign entry or opening to trade may result in knowledge spillovers to domestic firms within the wage sector, thus improving their profitability (possibly having a larger effect on firms in the right tail of the distribution). That will lead to an increase in average productivity directly, as well as through the exit of less productive firms due to intensified competition. The overall effect on the economy is an increase in productivity, wages, income and job creation, which is in line with a number of empirical studies that find a positive effect of trade or FDI inflows on domestic firms' productivity.³⁴

In what follows, however, we want to focus on three alternative strategies that might be considered as potential substitutes for technological improvement: (i)

³⁴Interestingly, the effect of a TFP shock on job creation is likely to be different. Consider our model in a special case where hiring costs and entry costs are proportional to wages or GDP. In this case, a sector-neutral shock that increases productivity in both home and wage sectors by the same amount, while raising the level of wages and income, will fail to induce reallocation of workers from self-employment to the wage sector.

active labour market policies, (ii) a reduction in entry barriers , and (iii) an increase in home sector productivity.

In the previous section we have found that, in least developed economies, both labour market frictions and the lack of skills have large impacts on job creation and wages. Many active labour market policies usually try to address either of these issues or both simultaneously, even though results have usually been underwhelming. A thorough analysis of job creation in the developing countries by the World Bank concluded that “there is no consensus on what the content of labor policies should be” (World Bank, 2013). Their review of active labour market policies (such as training) shows that effects are modest at best and that regulation (e.g. minimum wage or job security) has little impact on employment.³⁵

More recently, an expanding literature has focused on policies addressing skill shortages, by testing employment effects of vocational training among young workers. There are two papers that stand out, as they allow us to separate the effects of reducing labour market frictions from interventions aimed at increasing skills, through the design of field experiments in Sub-Saharan Africa. In both cases, skill acquisition and its observability (e.g. through certificates) have improved worker’s long term labour market outcomes in terms of employment and earnings over and above interventions that subsidize either search or firms.

Abebe *et al.* (Forthcoming) set up an experiment in Addis Ababa, Ethiopia, to test the labour market effects of two supply side constraints: high search costs, re-

³⁵The evidence in favour of active labour market policies is stronger in developed countries. Card, Kluve and Weber (2010) provide a literature review on the microeconomic evaluation of active labour market policies in developed countries. Their analysis suggests that subsidized public sector employment programs are relatively ineffective, while job search assistance and related programs have generally favourable impacts, especially in the short run. Moreover, a follow-up study by Card, Kluve and Weber (2017) show that the effectiveness of active labour market policies and job assistance is stronger for long-term unemployed and in countries with higher unemployment rates and lower GDP growth rates. Similarly, Pallais (2014) shows that employment outcomes improve when employers have more information about workers’ abilities.

duced through a transport subsidy, and informational frictions, that are mitigated by inviting young workers to a job application workshop that trains them how to signal observable skills (such as education or past job experience) and offers them a certificate of less observable skills through cognitive and other tests. The authors find that, while subsidising job search increases the short-run job finding rate (by increasing search intensity), these workers do not perform better than the control group in the medium to long run.³⁶ After four years, the transport subsidy has no discernible effect on employment. On the other hand, outcomes among workshop participants improve in terms of finding better (i.e., formal) and more durable jobs, and higher earnings. These effects are persistent after four years and suggest that policies that provide information about workers' skills are more effective at improving the quality and efficiency of matches between workers and firms.

Alfonsi *et al.* (2019) introduced a series of interventions targetting young workers and firms to try to improve urban labour markets in Uganda. Young workers received vocational training that improved sector-specific skills that were backed by a certificate. In the second intervention, firms were offered a subsidy to train new workers. Workers that were assigned to the vocational training were substantially more likely to comply, suggesting that, despite the subsidy, small and medium firms found it difficult to engage with the process. Those workers who did receive either training have accumulated sector-specific skills and earned similar wages, conditional on employment. However, the authors found a substantial difference in average effects resulting from the fact that skill certification for workers with vocational training allowed them to exit unemployment at a much faster rate, even over workers that acquired similar skills through firm training. This pa-

³⁶Franklin (2018) also shows that weekly transport costs in Ethiopia average about 20 percent of median total expenditure and that transport subsidies increase the probability of finding permanent employment by 30 percent in the short run. Moreover, the subsidies reduce participation in temporary and casual work.

per shows that, consistent with our finding that the lack of skills on the supply side is a major obstacle to the creation of stable jobs with higher productivity, policies aiming at providing certifiable skills before looking for jobs improves both the quantity and quality of matches. It also suggests that firm subsidies to increase employment and skills may not work in the context of least developed economies as the typical firm would struggle to put time and resources in training.

These papers highlight that skills matter substantially more when they are certified and that these returns seem larger in poorer countries, where frictions may be more severe. Other papers looking at vocational training interventions in middle income countries show more modest results. Attanasio *et al.* (2011) look at an intervention in Colombia and find that it has been effective at increasing employment and wages among young women, but not men. Hirshleifer *et al.* (2016) find that an experiment at scale for the unemployed in Turkey has had no discernible effects on employment and earnings, even though it did increase job quality (i.e. 2pp greater access to formal jobs among treated workers). Similarly, Card *et al.* (2011) find small effects on employment outcomes in Dominican Republic. Other papers have also found evidence that providing information about workers' abilities increases employment outcomes. Abel, Burger and Piraino (forthcoming) show that reference letters from previous employers increase the chances of interview call-backs by 60% in South Africa, while Bassi and Nansambaz (2019) show that certificates on non-cognitive skills increase worker and firm assortative matching.

In terms of entry barriers, the sunk cost that must be incurred by a market entrant is a key parameter in many models looking at productivity distribution of incumbent firms. However, existing literature has been less precise in defining these costs, even in counterfactual analysis. For example, Bartelsman *et al.* (2013) talk about a combination of factors, including entrepreneur's effort and administrative

fees; Aw *et al.* (2003) simply mentions regulatory and technological differences in entry costs in two countries (Taiwan and South Korea) and Ulyssea (2010) refers to technological determinants of entry for firms in Brazil. Direct micro evidence on entry costs has also failed to provide a good empirical answer to what entry costs are. For example, de Mel *et al.* (2012) and de Mel *et al.* (2013) show that for small firms in Sri Lanka, information about the registration process and reimbursement of direct costs are not as effective as one-off cash transfers for entry and survival of firms. McKenzie (2017) shows that large business grants in Nigeria are associated with greater firm entry, more survival and higher employment. In our estimation, we suggest that a good way of looking at entry costs in least developed countries is to concentrate on a key input for production, namely electricity. There is a large body of evidence that shows that appropriate access, quality and pricing of electricity can increase firm entry and industrial output (Rud, 2012a), and firm productivity and growth (Abeberese, 2017 and Allcott, Collard-Wexler and O'Connell, 2016). There is also evidence that, in the presence of shortages, larger firms in developing countries cope with electricity shocks, either by buying captive power generators (Reinikka and Svensson, 2002 and Rud, 2012b) or by adjusting labour and wages (Hardy and McCasland, 2019). As a consequence, infrastructure policies reducing the cost of access or improving its quality seem to be a good proxy for a reduction in entry costs.

Finally, improvements in the productivity of self-employed (which can be interpreted as an increase in A in our model) has been the target of a growing experimental literature in development. For example, Banerjee *et al.* (2015) in a comprehensive intervention that covered six countries and included asset transfers, consumption support, training in technical skills and access to savings showed substantial and persistent increases in consumption levels and measures of financial

inclusion and assets. Similarly, Blattman *et al.* (2013) shows that a program asking participants to submit grant proposals for vocational training and business start-up for self-employment increases earnings and assets. These, and a number of other papers, show that asset transfers (from livestock and other animals to start-up capital and cash) are an effective way of boosting productivity in self-employment activities.

7 Conclusion

Labour markets in least developed countries are characterized by a small proportion of workers in wage employment. Furthermore, the wage sector in developing countries tends to generate jobs that are relatively unproductive compared to similar jobs in industrialized and middle-income economies. As a consequence, pay is low on average. Despite these characteristics, wage employment in developing countries is still preferred by workers and has been identified by international organizations as key in generating economic growth and reducing poverty. This is because most of the labour force end up in less desirable and even less productive self-employment occupations (e.g. subsistence farming) or helping family activities for no pay.

We propose a unifying framework that endogenously generates the link between the size of the wage sector, mean productivity and wages. In particular, we incorporate channels identified by both the development and the labour literature - such as underlying productivity differences across countries (e.g. driven by lower capital intensity, inferior technology, etc.), barriers to entry (such as regulations, financial constraints, access to infrastructure) that prevent firms from entering the market and reduce competition, differences in workers' skills, bargaining power, and outside options (e.g. subsistence level farming), and labour market inefficien-

cies - that can interact to generate these outcomes.

We provide new empirical evidence on wage distributions using household level data for a number of Sub-Saharan African countries. Namely, the wage sector in developing countries, despite being very small in size, is characterized by very high levels of wage dispersion. We show that there exists a negative relationship between mean wage and wage dispersion and that this relationship holds even after controlling for workers' demographics, industries and regions. That is, wages vary substantially in developing countries even across similar individuals in similar occupations.

We also perform a numerical exercise that shows that the variation in the entry costs and labour market frictions can qualitatively and quantitatively describe labour markets in least developed countries. Differences in the underlying productivity dispersion, on the other hand, are not sufficient to explain differences in wage distributions across countries. Our results also reveal that there are significant complementarities between policy variables: for example, the effect of a change in labour market frictions on wage inequality is amplified in the presence of higher barriers to entry.

Our results demonstrate the power of estimating an integrated model of labour markets in developing countries. First, it allows us to combine different barriers to growth within a single framework and to examine their relative importance and interactions between them. Second, we can use it to analyse a great number of policies from relaxing entry constraints to improving self-employment productivity in order to identify priority areas in enhancing job creation and reducing inequality, which is a key step to designing more efficient policies that generate growth and reduce poverty.

Royal Holloway, University of London and Institute for Fiscal Studies

Royal Holloway, University of London

References

- Abebe, G., Caria, S., Fafchamps, M., Falco, P., Franklin, S. and Quinn, S. (Forthcoming). 'Anonymity or distance? job search and labour market exclusion in a growing african city', .
- Abeberese, A.B. (2017). 'Electricity cost and firm performance: Evidence from india', *The Review of Economics and Statistics*, vol. 99(5), pp. 839–852, doi: 10.1162/REST_a_00641.
- Abel, M., Burger, R. and Piraino, P. (forthcoming). 'The value of reference letters: Experimental evidence from south africa', *American Economic Journal: Applied*.
- Abowd, J.M. and Kramarz, F. (2003). 'The costs of hiring and separations', *Labour Economics*, vol. 10(5), pp. 499–530.
- Aghion, P. and Howitt, P. (1992). 'A model of growth through creative destruction', *Econometrica*, vol. 60(2), pp. 323–351.
- Albrecht, J., Robayo-Abril, M. and Vroman, S. (2017). 'Public-sector employment in an equilibrium search and matching model', *The Economic Journal*, doi: 10.1111/eoj.12533.
- Alfonsi, L., Bandiera, O., Bassi, V., Burgess, R., Rasul, I., Sulaiman, M. and Vitaliy, A. (2019). 'Tackling youth unemployment: Evidence from a labor market experiment in uganda', *Mimeo*.
- Allcott, H., Collard-Wexler, A. and O'Connell, S.D. (2016). 'How do electricity shortages affect industry? evidence from india', *American Economic Review*, vol. 106(3), pp. 587–624, doi:10.1257/aer.20140389.
- Alvarez, J., Benguria, F., Engbom, N. and Moser, C. (2018). 'Firms and the decline in earnings inequality in brazil', *American Economic Journal: Macroeconomics*, vol. 10(1), pp. 149–89, doi:10.1257/mac.20150355.

- Aragón, F.M. and Rud, J.P. (2015). 'Polluting industries and agricultural productivity: Evidence from mining in ghana', *The Economic Journal*, pp. n/a–n/a, ISSN 1468-0297, doi:10.1111/eoj.12244.
- Attanasio, O., Kugler, A. and Meghir, C. (2011). 'Subsidizing vocational training for disadvantaged youth in colombia: Evidence from a randomized trial', *American Economic Journal: Applied Economics*, vol. 3(3), pp. 188–220, doi: 10.1257/app.3.3.188.
- Aw, B.Y., Chung, S. and Roberts, M.J. (2003). 'Productivity, output, and failure: a comparison of taiwanese and korean manufacturers', *The Economic Journal*, vol. 113(491), pp. F485–F510, ISSN 1468-0297, doi:10.1046/j.0013-0133.2003.00166.x.
- Ayyagari, M., Demirgüç-Kunt, A. and Maksimovic, V. (2008). 'How important are financing constraints? the role of finance in the business environment', *World Bank Economic Review*, vol. 22(3), pp. 483–516.
- Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Parienté, W., Shapiro, J., Thuysbaert, B. and Udry, C. (2015). 'A multifaceted program causes lasting progress for the very poor: Evidence from six countries', *Science*, vol. 348(6236), ISSN 0036-8075, doi:10.1126/science.1260799.
- Banerjee, A.V. and Duflo, E. (2007). 'The economic lives of the poor', *Journal of Economic Perspectives*, vol. 21(1), pp. 141–168, doi:10.1257/jep.21.1.141.
- Banerjee, A.V. and Newman, A.F. (1993). 'Occupational choice and the process of development', *Journal of Political Economy*, vol. 101(2), pp. 274–298, doi: 10.2307/2138820.
- Bartelsman, E., Haltiwanger, J. and Scarpetta, S. (2013). 'Cross-country differences in productivity: The role of allocation and selection', *American Economic Review*, vol. 103(1), pp. 305–34, doi:10.1257/aer.103.1.305.
- Bassi, V. and Nansambaz, A. (2019). 'Screening and signalling non-cognitive skills: experimental evidence from uganda', .

- Blattman, C., Fiala, N. and Martinez, S. (2013). 'Generating skilled self-employment in developing countries: Experimental evidence from uganda*', *The Quarterly Journal of Economics*, doi:10.1093/qje/qjt057.
- Borjas, G. (1980). 'The relationship between wages and weekly hours of work: The role of division bias', *Journal of Human Resources*, vol. 15(3), pp. 409–423.
- Card, D., Ibarrarán, P., Regalia, F., Rosas-Shady, D. and Soares, Y. (2011). 'The labor market impacts of youth training in the dominican republic', *Journal of Labor Economics*, vol. 29(2), pp. 267–300, ISSN 0734306X, 15375307.
- Card, D., Kluve, J. and Weber, A. (2010). 'What works? a meta analysis of recent active labor market program evaluations', *The Economic Journal*, vol. 120, pp. F452–F477.
- Card, D., Kluve, J. and Weber, A. (2017). 'What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations', *Journal of the European Economic Association*, vol. 16(3), pp. 894–931, ISSN 1542-4766, doi:10.1093/jeea/jvx028.
- Clemens, M., Montenegro, C.E. and Pritchett, L. (2008). 'The place premium: Wage differences for identical workers across the u.s. border', Center for Global Development.
- de Mel, S., McKenzie, D. and Woodruff, C. (2012). 'One-time transfers of cash or capital have long-lasting effects on microenterprises in sri lanka', *Science*, vol. 335(6071), pp. 962–966.
- de Mel, S., McKenzie, D. and Woodruff, C. (2013). 'The demand for, and consequences of, formalization among informal firms in sri lanka', *American Economic Journal: Applied Economics*, vol. 5(2), pp. 122–50, doi:10.1257/app.5.2.122.
- Djankov, S., La Porta, R., Lopez-de Silanes, F. and Shleifer, A. (2002). 'The regulation of entry', *The Quarterly Journal of Economics*, vol. 117(1), pp. 1–37, doi: 10.1162/003355302753399436.
- Elsby, M.W.L. and Gottfries, A. (2019). 'Firm dynamics, on-the-job search and labor market fluctuations', *Mimeo*.

- Fields, G.S. (2011). 'Labor market analysis for developing countries', *Labour Economics*, vol. 18, Supplement 1(0), pp. S16–S22, doi: <http://dx.doi.org/10.1016/j.labeco.2011.09.005>.
- Franklin, S. (2018). 'Location, search costs and youth unemployment: Experimental evidence from transport subsidies', *The Economic Journal*, vol. 128(614), pp. 2353–2379.
- Ghatak, M. and Nien-Huei Jiang, N. (2002). 'A simple model of inequality, occupational choice, and development', *Journal of Development Economics*, vol. 69(1), pp. 205–226, doi:[http://dx.doi.org/10.1016/S0304-3878\(02\)00059-7](http://dx.doi.org/10.1016/S0304-3878(02)00059-7).
- Girsberger, E.M. and Meango, R. (2018). 'The puzzle of educated unemployment in west africa', *Mimeo*.
- Gollin, D., Lagakos, D. and Waugh, M.E. (2014). 'The agricultural productivity gap', *The Quarterly Journal of Economics*, vol. 129(2), pp. 939–993.
- Gourieroux, C., Monfort, A. and Renault, E. (1993). 'Indirect Inference', *Journal of Applied Econometrics*, vol. 8, pp. S85–S118.
- Grossman, G. and Helpman, E. (1991). *Innovation and growth in the global economy*, Cambridge, MA: MIT Press.
- Hall, R.E. and Milgrom, P.R. (2008). 'The limited influence of unemployment on the wage bargain', *American Economic Review*, vol. 98(4), pp. 1653–74, doi: [10.1257/aer.98.4.1653](https://doi.org/10.1257/aer.98.4.1653).
- Hardy, M. and McCasland, J. (2019). 'Lights off, lights on: The effects of electricity shortages on small firms', *The World Bank Economic Review*, ISSN 0258-6770, doi: [10.1093/wber/lhz028](https://doi.org/10.1093/wber/lhz028), lhz028.
- Harris, J.R. and Todaro, M.P. (1970). 'Migration, unemployment and development: A two-sector analysis', *The American Economic Review*, vol. 60(1), pp. 126–142, doi:[10.2307/1807860](https://doi.org/10.2307/1807860).
- Heckman, J.J. and Honoré, B.E. (1990). 'The empirical content of the roy model', *Econometrica*, vol. 58(5), pp. 1121–1149, doi:[10.2307/2938303](https://doi.org/10.2307/2938303).

- Herrendorf, B. and Teixeira, A. (2011). 'Barriers to entry and development', *International Economic Review*, vol. 52(2), pp. 573–602, ISSN 1468-2354, doi: 10.1111/j.1468-2354.2011.00639.x.
- Hirshleifer, S., McKenzie, D., Almeida, R. and Ridao-Cano, C. (2016). 'The impact of vocational training for the unemployed: Experimental evidence from turkey', *The Economic Journal*, vol. 126(597), pp. 2115–2146, doi:10.1111/ecoj.12211.
- Hopenhayn, H.A. (1992). 'Entry, exit, and firm dynamics in long run equilibrium', *Econometrica*, vol. 60(5), pp. 1127–1150, doi:10.2307/2951541.
- Hsieh, C.T. and Klenow, P.J. (2009). 'Misallocation and manufacturing tfp in china and india', *The Quarterly Journal of Economics*, vol. 124(4), pp. 1403–1448, doi: 10.1162/qjec.2009.124.4.1403.
- Hsieh, C.T. and Klenow, P.J. (2010). 'Development accounting', *American Economic Journal: Macroeconomics*, vol. 2(1), pp. 207–23, doi:10.1257/mac.2.1.207.
- ILO (2012). 'Global wage report - africa brief', International Labour Organization.
- La Porta, R. and Shleifer, A. (2014). 'Informality and development', *Journal of Economic Perspectives*, vol. 28(3), pp. 109–26, doi:10.1257/jep.28.3.109.
- Lagakos, D. and Waugh, M.E. (2013). 'Selection, agriculture, and cross-country productivity differences', *American Economic Review*, vol. 103(2), pp. 948–80, doi: 10.1257/aer.103.2.948.
- Lewis, W.A. (1954). 'Economic development with unlimited supplies of labour', *The Manchester School*, vol. 22(2), pp. 139–191, doi:10.1111/j.1467-9957.1954.tb00021.x.
- Matsuyama, K. (1992). 'Agricultural productivity, comparative advantage, and economic growth', *Journal of Economic Theory*, vol. 58(2), pp. 317–334, doi: [http://dx.doi.org/10.1016/0022-0531\(92\)90057-O](http://dx.doi.org/10.1016/0022-0531(92)90057-O).
- McKenzie, D. (2017). 'Identifying and spurring high-growth entrepreneurship: Experimental evidence from a business plan competition', *American Economic Review*, vol. 107(8), pp. 2278–2307, doi:10.1257/aer.20151404.

- McKenzie, D. and Paffhausen, A.L. (2019). 'Small firm death in developing countries', *The Review of Economics and Statistics*, vol. 101(4), pp. 645–657, doi: 10.1162/rest.a_00798.
- Meghir, C., Narita, R. and Robin, J.M. (2015). 'Wages and informality in developing countries', *American Economic Review*, vol. 105(4), pp. 1509–46, doi: 10.1257/aer.20121110.
- Melitz, M.J. (2003). 'The impact of trade on intra-industry reallocations and aggregate industry productivity', *Econometrica*, vol. 71(6), pp. 1695–1725, doi: 10.2307/1555536.
- Mortensen, D.T. and Pissarides, C.A. (1994). 'Job creation and job destruction in the theory of unemployment', *The Review of Economic Studies*, vol. 61(3), pp. 397–415, doi:10.2307/2297896.
- Pallais, A. (2014). 'Inefficient hiring in entry-level labor markets', *American Economic Review*, vol. 104(11), pp. 3565–99, doi:10.1257/aer.104.11.3565.
- Petrongolo, B. and Pissarides, C.A. (2001). 'Looking into the black box: A survey of the matching function', *Journal of Economic Literature*, vol. 39(2), pp. 390–431, doi:10.1257/jel.39.2.390.
- Reinikka, R. and Svensson, J. (2002). 'Coping with poor public capital', *Journal of Development Economics*, vol. 69(1), pp. 51 – 69, ISSN 0304-3878, doi: [http://dx.doi.org/10.1016/S0304-3878\(02\)00052-4](http://dx.doi.org/10.1016/S0304-3878(02)00052-4).
- Robinson, S. (1976). 'A note on the u hypothesis relating income inequality and economic development', *The American Economic Review*, vol. 66(3), pp. 437–440, ISSN 00028282.
- Rud, J.P. (2012a). 'Electricity provision and industrial development: Evidence from india', *Journal of Development Economics*, vol. 97(2), pp. 352–367, doi: <http://dx.doi.org/10.1016/j.jdeveco.2011.06.010>.
- Rud, J.P. (2012b). 'Infrastructure regulation and reallocations within industry: Theory and evidence from indian firms', *Journal of Development Economics*, vol. 99(1), pp. 116–127, doi:<http://dx.doi.org/10.1016/j.jdeveco.2011.10.001>.

- Sahin, A., Song, J., Topa, G. and Violante, G.L. (2014). 'Mismatch unemployment', *American Economic Review*, vol. 104(11), pp. 3529–64, doi:10.1257/aer.104.11.3529.
- Silva, J.I. and Toledo, M.E. (2009). 'Labor turnover costs and the cyclical behavior or vacancies and unemployment', *Macroeconomic Dynamics*, vol. 13, pp. 76–96, ISSN 1469-8056, doi:10.1017/S1365100509080122.
- Song, J., Price, D.J., Guvenen, F., Bloom, N. and von Wachter, T. (2018). 'Firming Up Inequality*', *The Quarterly Journal of Economics*, vol. 134(1), pp. 1–50, ISSN 0033-5533, doi:10.1093/qje/qjy025.
- Ulyssea, G. (2010). 'Regulation of entry, labor market institutions and the informal sector', *Journal of Development Economics*, vol. 91(1), pp. 87–99, doi: <http://dx.doi.org/10.1016/j.jdeveco.2009.07.001>.
- Vollrath, D. (2014). 'The efficiency of human capital allocations in developing countries', *Journal of Development Economics*, vol. 108(0), pp. 106–118, doi: <http://dx.doi.org/10.1016/j.jdeveco.2014.01.009>.
- World Bank (2013). 'World development report: Jobs', (1).
- Zenou, Y. (2008). 'Job search and mobility in developing countries. theory and policy implications', *Journal of Development Economics*, vol. 86(2), pp. 336–355, doi: <http://dx.doi.org/10.1016/j.jdeveco.2007.06.009>.

ONLINE APPENDIX

Job Creation and Wages in Least Developed Countries: Evidence from

Sub-Saharan Africa

Juan Pablo Rud and Ija Trapeznikova

A Data Description

For the empirical part of this paper we use the following individual-level datasets: Burkina Faso Enquête Multisectorielle Continue 2014, Cameroon Household Survey 2007; Ethiopia Socioeconomic Survey ESS2 2013 and ESS3 2015; Ghana Living Standards Survey 2013; Kenya Integrated Household Budget Survey 2005; Mali Enquête Agricole de Conjoncture Intégrée 2014; Niger National Survey on Household Living Conditions and Agriculture (ECVMA) 2011 and 2014; Nigeria General Household (Post-Planting and Post-Harvest) Surveys 2010, 2012 and 2015; South Africa Labour Force Survey 2007 for wages and South Africa General Household Survey 2007 for expenditure data; Tanzania National Panel Survey 2008, 2010 and 2012; Uganda National Panel Survey 2010, 2011, 2012, and 2014; Zambia Living Condition Monitoring Survey 2010; and Survey of Income Program Participation (SIPP) 2008 panel for the US.

When measuring wages and employment status of workers, we restrict the sample to individuals aged between 15 and 65. In addition, the public sector employees (government administration, state enterprises and parastatals, NGOs, and diplomatic missions) are excluded from our analysis.

(a) Wages

Most of the wage earnings data are given at a monthly frequency. However, when wages refer to a payment period other than a month, we use 20 working days per month, 5 days per week, and 3 months per quarter, to convert them into

monthly series. We compute real wages using CPI index with 2010=100 and convert them into international dollars using private consumption based PPP conversion rate. We trim off top and bottom 1% of wages. Residual wage dispersion is obtained from a wage regression that controls for demographics (gender, age, age squared, marital status, education), regions, urban status, and industry.

(b) Self-employment

Self-employed individuals in our analysis include unpaid family members, which represent about one third in Ghana and Cameroon to about a half of all self-employed in Ethiopia and Uganda. Employers (self-employed workers with non-household members as employees) and self-employed workers that are managers, professionals and technicians are excluded from the sample, comprising about 1%-3% and less than 1% of all self-employed, respectively (one exception is Nigeria where about 3% of self-employed are high-skilled workers). In South Africa, self-employed and unemployed individuals are treated together. On average, about 70% of self-employed individuals live in rural areas, the vast majority of them (60%-70%) work in agriculture and about 20%-30% work in sales or personal services.

(c) Transitions

The transition rates between self-employment and wage employment are calculated using the following datasets: Niger ECVMAI (2011) and ECVMAII (2014); Ethiopia Socioeconomic Survey Waves 2 (2006) and 3 (2008); Uganda National Panel Survey 2010, 2011, 2012 and 2014; Tanzania National Panel Survey Waves 1 (2008), 2 (2010) and 3 (2012); Nigeria General Household Survey Waves 1 (2010), 2 (2012) and 3 (2015); and South Africa Labour Force Survey March and September 2007. These datasets have a panel structure. The transitions are calculated based on individuals that are surveyed in both periods, i.e. the exit rates from the survey are assumed to be random. The time period between survey dates varies across coun-

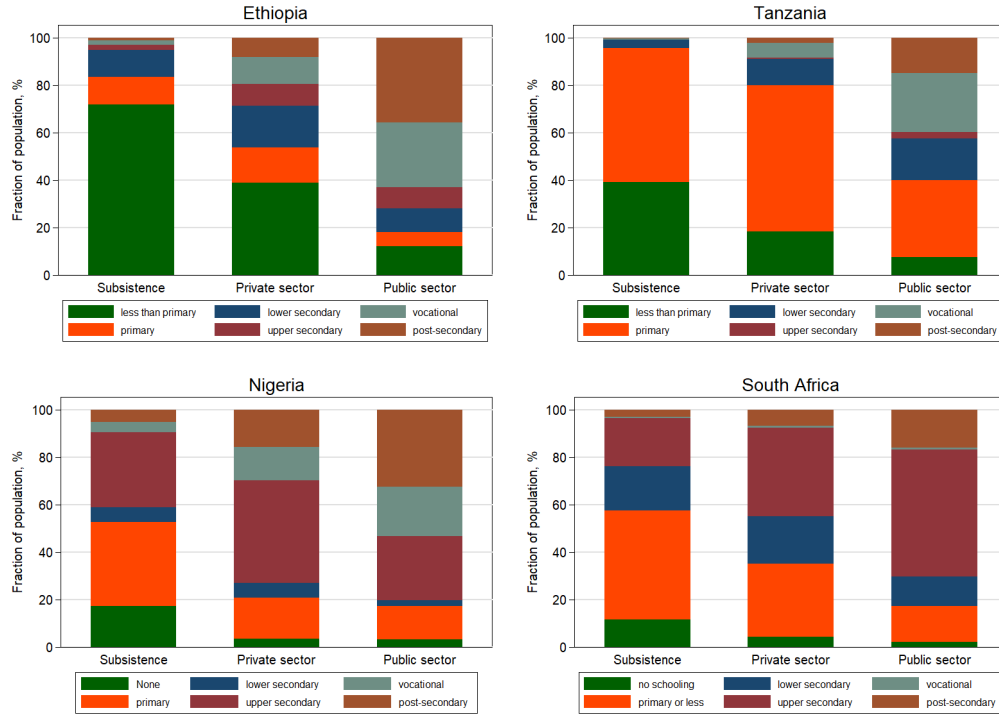
tries (e.g. 6 months in South Africa versus 3 years in Niger), hence all transition rates were converted into monthly rates. That is, under the assumption of a Poisson arrival rate the average share of workers losing a job within the time period t is equal to $1 - e^{-\delta t}$. In this way, we can recover monthly rate δ to be used in the estimation. Moreover, the transition rates are weighted by the inverse of the length of the time interval between two interviews if they differ within one dataset.

(d) Selection into the wage sector

It is difficult to ensure that different educational groups are comparable across countries due to (a) the existing differences between national education systems, and (b) the level of detail in educational attainment data in different surveys. However, we tried to classify workers into six broad categories and keep them broadly consistent across countries: (i) no schooling or less than primary education, (ii) completed primary schooling, (iii) completed lower secondary schooling (e.g. O-Levels in Tanzania, Secondary schooling first cycle in Niger, Junior Secondary School Certificate in Nigeria) or incomplete secondary (in South Africa), (iv) upper secondary education, (v) technical or vocational education (including teacher training certificate in Ethiopia, post-primary specialized education in Uganda, nursing or teaching certificate in Nigeria, National Technical Certificate in South Africa), and (vi) post-secondary education that includes Bachelor, Masters, and Doctorate degrees, as well as 4-year/level 4 Diploma in Nigeria and Ethiopia, post-secondary specialized education in Uganda.

Figure 2 in the main text and Figure A.1 below show the composition of workers by education across the three sectors of the economy. The fact that the public sector hires predominantly more educated workers is true for all countries in our sample. Since workers employed by private firms are more similar to the self-employed in terms of their educational attainment, we consider only the private sector in this

Figure A.1: Education distribution by sector



Source: Authors' computations using the following datasets: Ethiopia ESS 2013 and 2015; Tanzania NPS 2008, 2010, 2012; Nigeria GHS 2010, 2012, 2015; South Africa LFS 2007.

paper.

The probability to get a job in the wage sector might differ depending on worker characteristics. We can look at how transition rates between self-employment and wage employment depend on workers' demographics using the panel structure of the data. However, the number of transitions in the data is very low; thus we chose to run a simple selection regression of being employed in the private wage sector (as opposed to engaging in own account work) using a linear probability model. The results are shown in Table A.1.

We find that education is a key determinant of worker's employment status. After controlling for regions, rural/urban areas, gender and age, workers with

technical/vocational and post-secondary education are more likely to be in paid employment. In some cases also secondary education increases workers' chances to get a job.

Table A.2 shows that the education groups that have a statistically significant effect on employment selection (highlighted in bold) comprise a very small fraction of the population. For most countries in our sample, less than 10% of workers have a higher chance of getting a paid job. We think that having two types of workers - less educated and highly educated - in the model is sufficient to capture differential selection into the wage sector. Note that even though the coefficients on lower secondary education in South Africa and primary education in Niger are statistically significant, we assign these groups to the low type since the magnitude of the coefficients is relatively small.

Table A.1: Selection regression - the effect of educational attainment

Y = working in the wage sector	Ethiopia	Niger	Nigeria	Tanzania	Uganda	South Africa
Education level (Reference group: less than primary)						
Primary	0.013 (1.02)	0.022 (2.39)**	-0.025 (3.98)***	-0.000 (0.06)	-0.018 (2.29)**	0.024 (1.07)
Lower secondary	-0.005 (0.32)	0.045 (3.29)***	0.006 (0.42)	0.042 (3.20)***	0.020 (1.18)	0.044 (1.87)*
Upper secondary	-0.008 (0.27)	0.253 (2.75)***	0.019 (2.47)**	0.119 (1.24)	0.064 (1.34)	0.164 (7.00)***
Vocational	0.141 (4.29)***	0.460 (6.72)***	0.169 (8.29)***	0.217 (6.65)***	0.145 (5.48)***	0.135 (3.13)***
Post-secondary	0.117 (3.18)***	0.529 (7.30)***	0.181 (10.97)***	0.393 (7.21)***	0.260 (10.30)***	0.332 (12.55)***
Age	0.005 (2.79)***	0.001 (1.19)	0.002 (1.50)	0.006 (6.95)***	0.007 (4.31)***	0.037 (20.60)***
Age Squared/100	-0.007 (3.19)***	-0.001 (0.88)	-0.005 (2.43)**	-0.008 (7.93)***	-0.012 (6.45)***	-0.038 (16.79)***
Female	-0.056 (7.73)***	-0.035 (6.85)***	-0.062 (9.33)***	-0.077 (20.86)***	-0.142 (20.29)***	-0.103 (16.53)***
Married	-0.020 (2.09)**	0.010 (1.58)	-0.059 (6.56)***	-0.050 (11.07)***	-0.100 (11.33)***	0.059 (8.42)***
Urban	0.231 (17.17)***	0.100 (10.59)***	0.119 (16.82)***	0.103 (14.92)***	0.155 (14.26)***	n/a n/a
\bar{Y}	0.103	0.060	0.114	0.072	0.169	0.617
R^2	0.31	0.11	0.15	0.17	0.17	0.12
N	7,107	14,390	17,142	23,630	19,404	62,910

Note: * $p < 0.10$; ** $p < 0.05$, *** $p \leq 0.01$, absolute t-values are given in parentheses. The outcome variable is working in the wage sector as opposed to being self-employed (or also unemployed in South Africa). Controlling for year and region fixed effects. The sample is limited to 15-65 year old individuals and excludes the public sector. There is no urban/rural variable for South Africa in 2007. Source: Authors' computations using the following datasets: Ethiopia ESS 2013 and 2015; Niger ECVMAI (2011) and ECVMAII (2014), Nigeria GHS 2011, 2012, 2015, Uganda NPS 2010, 2011, 2012, and 2014; Tanzania NPS 2008, 2010, 2012; South Africa LFS 2007.

Table A.2: Percentage distribution of workers by educational attainment

Education level	less than primary	primary	lower sec.	upper sec.	techn./vocational	post-sec.	Share of H-type
<i>Ethiopia</i>							
Wage sector	39.0	14.8	17.6	9.2	11.3	8.1	19.4
Home sector	72.0	11.6	11.4	2.2	1.9	0.9	2.8
Total	68.6	12.0	12.0	2.9	2.9	1.7	4.5
<i>Niger</i>							
Wage sector	62.6	19.8	10.6	1.0	3.0	2.9	17.6
Home sector	84.6	10.7	4.4	0.1	0.1	0.1	4.7
Total	83.3	11.3	4.8	0.1	0.3	0.2	5.4
<i>Nigeria</i>							
Wage sector	3.5	17.4	6.3	43.0	13.9	16.0	29.9
Home sector	18.0	35.3	6.1	31.7	4.1	5.0	9.0
Total	16.3	33.2	6.1	33.0	5.2	6.2	11.4
<i>Tanzania</i>							
Wage sector	18.3	61.7	11.2	0.5	6.1	2.1	20.0
Home sector	38.8	56.7	3.6	0.1	0.7	0.1	4.5
Total	37.3	57.1	4.2	0.1	1.1	0.2	5.6
<i>Uganda</i>							
Wage sector	43.7	25.7	10.1	3.7	5.2	11.7	16.9
Home sector	62.3	26.4	5.8	1.2	2.1	2.3	4.3
Total	59.1	26.3	6.5	1.6	2.6	3.8	6.4
<i>South Africa</i>							
Wage sector	4.3	30.9	19.9	37.5	0.8	6.6	44.9
Home sector	5.2	38.6	26.0	28.8	0.5	0.8	30.2
Total	4.6	33.8	22.3	34.2	0.7	4.4	39.3

Note: The sample is limited to 15-65 year old individuals, being self-employed or working in the private wage sector. The education groups highlighted in bold are considered to be high type. Source: Authors' computations using the following datasets: Ethiopia ESS 2013 and 2015; Niger ECVMAI (2011) and ECVMAII (2014), Uganda NPS 2010, 2011, 2012, and 2014; Tanzania NPS 2008, 2010, 2012; Nigeria GHS 2011, 2012, 2015, South Africa LFS 2007.

B Model

B.1 Equilibrium conditions

Denote by p_{Rj} the reservation productivity at which the worker will be indifferent between accepting the offer and continuing searching, $w_j(p_{Rj}) = w_{Rj}$, i.e.

$$w_{Rj} = (\beta p_{Rj} + (1 - \beta)\tilde{h})a_j. \quad (\text{B.1})$$

Then, using the equation for a worker's reservation wage, we can derive her reservation productivity p_{Rj} as follows

$$p_{Rj} = \tilde{h} - \frac{z_j}{\beta a_j} + \frac{s_j \lambda(\theta)}{r + \delta} \int_{p_{Rj}} (p - p_{Rj}) \frac{d\Gamma(p)}{\bar{\Gamma}(\hat{p})}. \quad (\text{B.2})$$

Using integration by part, we can show that $\int_{p_{Rj}} (p - p_{Rj}) d\Gamma(p) = \int_{p_{Rj}} \bar{\Gamma}(p) dp$.

We are looking for an equilibrium, in which the ex-post productivity distribution, and hence the wage distribution, are determined by the labor demand side, i.e. $\hat{p} \geq p_{Rj}$ for both worker types, which means that workers accept all job offers.

$$\begin{aligned} p_{Rj} &= \tilde{h} - \frac{z_j}{\beta a_j} + \frac{s_j \lambda(\theta)}{r + \delta} \int_{p_{Rj}} (p - p_{Rj}) \frac{d\Gamma(p)}{\bar{\Gamma}(\hat{p})} \\ &\leq \tilde{h} - \frac{z_j}{\beta a_j} + \frac{s_j \lambda(\theta)}{r + \delta} \int_{\hat{p}} (p - \hat{p}) \frac{d\Gamma(p)}{\bar{\Gamma}(\hat{p})} \\ &= \hat{p} - \frac{z_j}{\beta a_j} + \frac{1}{(1 - \beta)\bar{a}(\theta)} \left(s_j \left(\theta + \frac{\lambda(\theta)}{r + \delta} \right) \frac{rk}{\bar{\Gamma}(\hat{p})} - \frac{c(r + \delta)}{q(\theta)} \right), \end{aligned}$$

where the first step uses the fact that the surplus function $\varphi(\cdot)$ is a decreasing function so that $\int_{p_{Rj}} (p - p_{Rj}) d\Gamma(p) \geq \int_{\hat{p}} (p - \hat{p}) d\Gamma(p)$ for $p_{Rj} \leq \hat{p}$, while the second step uses equations (ZP) and (FE) to substitute for \tilde{h} and the surplus function. For the reservation productivity p_{Rj} to be less or equal than \hat{p} , the searching costs z_j

need to satisfy the following condition:

$$z_j \geq \frac{\beta a_j}{(1-\beta)\bar{a}(\theta)} \left(\frac{rk}{\bar{\Gamma}(\hat{p})} \left(s_j \theta + \frac{s_j \lambda(\theta)}{r+\delta} \right) - \frac{c(r+\delta)}{q(\theta)} \right). \quad (\text{B.3})$$

In our simulations, we assume that the reservation wage is not binding and that workers accept all wage offers. That is, the searching costs z are such that the lowest wages in the market are determined by productivity threshold \hat{p} and not by the reservation wage. We find this assumption to be justified given very low labour mobility and low levels of self-employment income in least developed countries.³⁷

The second condition that has to be satisfied in equilibrium is the participation constraint. That is, the value of search for a j -type worker, $U_j = w_{Rj}/r$, has to be higher than the value of dropping out of the labour market and producing at home (i.e. getting self-employment income forever), which is equal to h_j/r . In the latter case the worker does not incur searching costs, but at the same time she is giving up the opportunity to find a job in the wage sector. This might happen if the searching costs are too high relative to the benefits of the job search, which in turn depend on the rate of finding a job, market wages, and the destruction rate that determines how long jobs last on average. Hence, in order for workers to be willing to participate in the labour market the reservation wage w_{Rj} has to be at least as high as home production income h_j , or equivalently, p_{Rj} needs to be higher

³⁷Based on our parameter estimates, we find that inequality (B.3) is not binding for our model in any country, except for South Africa. To match the observed magnitudes of wage dispersion, the implied searching costs in SA need to be higher than \hat{h} , resulting in a negative consumption flow while searching. Although this is an unrealistic assumption, we choose to make it nevertheless to keep the modelling framework exactly the same between all countries in our analysis. If one were interested in South Africa only, a more advanced model is needed that would account separately for the large informal sector and allow for on-the-job search to fit the data (see, for example, Meghir *et al.*, 2015 and Ulyssea, 2010).

than \tilde{h} for each type j . That is,

$$p_{Rj} = \frac{r + \delta}{s_j \lambda(\theta) + r + \delta} \left(\tilde{h} - \frac{z_j}{\beta a_j} \right) + \frac{s_j \lambda(\theta)}{s_j \lambda(\theta) + r + \delta} \int_{\hat{p}} p \frac{d\Gamma(p)}{\bar{\Gamma}(\hat{p})} \geq \tilde{h}, \quad (\text{B.4})$$

where we have used the fact that all market wages are accepted by the workers, so that the lower limit of the integral is determined by \hat{p} .

This implies that the searching costs need to satisfy the following inequality:

$$\begin{aligned} \frac{z_j}{\beta a_j} &\leq \frac{s_j \lambda(\theta)}{r + \delta} \int_{\hat{p}} (p - \tilde{h}) \frac{d\Gamma(p)}{\bar{\Gamma}(\hat{p})} \\ &= \frac{s_j \lambda(\theta)}{r + \delta} \left(\int_{\hat{p}} (p - \hat{p}) \frac{d\Gamma(p)}{\bar{\Gamma}(\hat{p})} + \frac{c(r + \delta)}{q(\theta)(1 - \beta)\bar{a}(\theta)} \right) \\ &= \frac{1}{(1 - \beta)\bar{a}(\theta)} \left(\frac{rk}{\bar{\Gamma}(\hat{p})} \left(s_j \theta + \frac{s_j \lambda(\theta)}{r + \delta} \right) + cs_j \theta \right) \end{aligned} \quad (\text{B.5})$$

where we used equation (ZP) to substitute for \tilde{h} and equation (FE) to substitute for the surplus function.

Combining the two conditions above, we get the following interval for z

$$\left(\frac{rk}{\bar{\Gamma}(\hat{p})} \left(s_j \theta + \frac{s_j \lambda(\theta)}{r + \delta} \right) + cs_j \theta \right) \geq \frac{(1 - \beta)\bar{a}(\theta)}{\beta a_j} z_j \geq \left(\frac{rk}{\bar{\Gamma}(\hat{p})} \left(s_j \theta + \frac{s_j \lambda(\theta)}{r + \delta} \right) - \frac{c(r + \delta)}{q(\theta)} \right), \quad (\text{B.6})$$

which is non-empty as long as $c > 0$. The interval is wider if the vacancy costs are higher or matching efficiency is lower.

Discussion of the Participation constraint

Now suppose that inequality (B.5) does not hold for some workers, so that their participation constraint binds. This might happen if z_j is too high so that some workers will prefer to quit searching and to produce at home instead. In that case, as the number of low skilled workers searching in the market falls, the market tightness increases and so does the job finding rate, which in turn pushes the reservation wage up. Hence, the number of job-seekers will adjust until the reservation

wage is level with home production income, so that workers are indifferent between searching in the market or not, that is $p_{Rj} = \tilde{h}$. In our model, it is reasonable to assume that the participation constraint will bind for low skilled workers (they have less to gain in paid employment due to lower wages and lower job finding probability), while all high skilled workers continue to search for a wage job, i.e. $p_{RH} > p_{RL} = \tilde{h}$.

Let $\kappa \in (0, \alpha)$ to be a mass of self-employed a_L -type workers who search in the market. The market tightness is still given by $\theta = \frac{v}{u_L + s_H u_H}$, while the steady state mass of low ability job seekers is equal to

$$u_L = \frac{\delta \kappa}{\delta + \lambda(\theta)}. \quad (\text{B.7})$$

The average ability of a randomly met worker is given by

$$\bar{a}(\theta) = a_L + (a_H - a_L) \frac{s_H(1 - \alpha)(\delta + \lambda(\theta))}{(\delta + s_H \lambda(\theta)\kappa + s_H(1 - \alpha)(\delta + \lambda(\theta)))} \quad (\text{B.8})$$

which is decreasing in κ , as it shifts the composition of the job seekers towards low ability workers. The efficiency units of labour engaged in home production is equal to

$$L_H(\theta) = a_L(u_L(\theta) + \alpha - \kappa) + a_H u_H(\theta) = \left(\alpha - \frac{\lambda(\theta)\kappa}{\delta + \lambda(\theta)} \right) a_L + \frac{\delta(1 - \alpha)}{\delta + s_H \lambda(\theta)} a_H, \quad (\text{B.9})$$

Now, in addition to equilibrium equations (ZP) and (FE), we have one additional condition $h_L = w_{RL}$, or equivalently $\tilde{h} = p_{RL}$, and one additional unknown κ . We can express the participation constraint as

$$\frac{z_L}{\beta a_L} = \frac{\lambda(\theta)}{r + \delta} \int_{\hat{p}} (p - \hat{p}) \frac{d\Gamma(p)}{\bar{\Gamma}(\hat{p})} + \frac{c\theta}{(1 - \beta)\bar{a}(\theta)}, \quad (\text{PC})$$

which is an increasing schedule of \hat{p} and θ , since $\frac{\partial}{\partial \hat{p}} E(p|p > \hat{p}) < 1$ for log-concave density functions.

While we can solve for an equilibrium of this model, it is difficult to identify from the data what share of the self-employed are actually looking for a job, or alternatively what the searching costs are. Therefore, for simplicity of exposition we choose to impose the condition that the searching costs are such that all workers engage in active job search.

B.2 Comparative statics

Using equations (FE) and (ZP), we can show that a reduction in entry costs k increases \hat{p} and θ . Using graphical analysis similar to the one performed in Figure 3, we can unambiguously show the effect of changes in underlying productivity distribution Γ and self-employment productivity A on equilibrium variables. The impact of the other parameters in the model cannot be proven analytically and in general will depend on the parameter values.

(a) Underlying productivity distribution

Suppose that Γ_1 distribution first-order stochastically dominates Γ_2 , i.e. $\Gamma_1(p) \leq \Gamma_2(p)$ for all p . One example is a location shift when the mean is higher under Γ_1 than under Γ_2 , while the variance is the same for both. In this case, a greater mean productivity implies a greater expected gain from entry and shifts the Free Entry condition upwards. As a consequence, the equilibrium values of the reservation productivity and the market tightness are higher under Γ_1 than under Γ_2 . Similarly, both \hat{p} and θ are higher under Γ_1 if it is a mean-preserving spread of Γ_2 . A higher productivity threshold under a riskier distribution (for the same mean) reflects the option value of risk. That is, firms benefit from increased chances of very high realizations of productivity p , while the costs of having a very low realization is

always bounded by the value of exiting from the market.

(b) *Home sector productivity*

Now consider an increase in productivity of the self-employment sector. Larger values of A increase workers' outside option and thus their wages. Now, the marginal firm needs to be more productive, shifting the Zero Profit condition upwards. As a result, \hat{p} increases and θ falls, leading to more workers moving to the home production sector.

(c) *Matching efficiency*

To show the effect of a change in matching efficiency m it is useful to rewrite the free entry and zero profit conditions in terms of the job finding rate, λ , instead of market tightness θ . In that case, home production income \tilde{h} and the average ability among job seekers \bar{a} are constant for a given λ . That is,

$$\hat{p} = \tilde{h}(\lambda) + \frac{c(r + \delta)}{m^{\frac{1}{\eta}} \lambda^{\frac{\eta-1}{\eta}} (1 - \beta) \bar{a}(\lambda)}, \quad (\text{B.10})$$

$$rk = \frac{m^{\frac{1}{\eta}} \lambda^{\frac{\eta-1}{\eta}}}{r + \delta + m^{\frac{1}{\eta}} \lambda^{\frac{\eta-1}{\eta}}} (1 - \beta) \bar{a}(\lambda) \varphi(\hat{p}), \quad (\text{B.11})$$

where we have expressed the vacancy filling rate as a function of λ , i.e. $q = m^{\frac{1}{\eta}} \lambda^{\frac{\eta-1}{\eta}}$. Similarly to equation (ZP), equation (B.10) represents an increasing relationship between \hat{p} and λ , since \tilde{h} is increasing in λ , while the vacancy filing rate q and the average ability among job seekers \bar{a} are decreasing in λ . Equation (B.11), similarly to (FE), is a downward-sloping schedule in (\hat{p}, λ) dimensions.

An increase in m shifts the ZP curve downwards, corresponding to a lower value of \hat{p} for each value of λ . An increase in matching efficiency makes it easier to fill a vacancy for firms and leads to a higher value of a firm, $V(p)$, for every productivity level p . At the same time, the FE curve shifts upwards, leading to a

higher job finding rate λ in equilibrium and, as a result, to a larger wage sector.

In order to see what happens to productivity threshold \hat{p} , we combine equations (B.10) and (B.11) to obtain

$$\hat{p} = \tilde{h}(\lambda) - \frac{c}{(1-\beta)\bar{a}(\lambda)} + \frac{c\varphi(\hat{p})}{rk} \quad (\text{B.12})$$

Then, differentiating \hat{p} with respect to m we get:

$$\frac{\partial \hat{p}}{\partial m} = \left(1 - \frac{c\varphi'(\hat{p})}{rk}\right)^{-1} \left[\frac{\partial \tilde{h}}{\partial \lambda} + \frac{c}{(1-\beta)\bar{a}(\lambda)^2} \frac{\partial \bar{a}}{\partial \lambda} \right] \frac{\partial \lambda}{\partial m}.$$

We know that in equilibrium $\frac{\partial \lambda}{\partial m} > 0$ and $\varphi'(\hat{p}) = \Gamma(\hat{p}) - 1 < 0$, thus the effect on $\frac{\partial \hat{p}}{\partial m}$ is determined by the sign of the term in the square brackets. The first term $\frac{\partial \tilde{h}}{\partial \lambda}$ is positive since higher λ implies a smaller home sector, which in turn leads to a higher self-employment income due to decreasing returns to scale in home production. However, the second term $\frac{\partial \bar{a}}{\partial \lambda}$ is negative as the composition of the unemployed shifts towards the low type as the job finding rate increases. Thus, the overall effect of an increase in m on \hat{p} will depend on the parameters. Consider, for example, the limit case of $\gamma = 1$, i.e. the constant returns to scale in the home sector. Then self-employment income is independent of the home sector size and $\frac{\partial \tilde{h}}{\partial \lambda} = 0$, implying $\frac{\partial \hat{p}}{\partial m} < 0$. Another extreme is $s_H = 1$, i.e. there is no selection into the wage sector and both worker types find jobs at the same rate. In that case, $\frac{\partial \bar{a}}{\partial \lambda} = 0$ and the overall effect on an increase in m on \hat{p} is positive. Overall, the increase in m leads to a larger wage sector, but the effect on \hat{p} is ambiguous.

(d) Worker's bargaining power

Similarly to an increase in matching efficiency, a decrease in workers' bargaining power parameter β shifts the ZP curve downwards and the FE curve upwards leading to a rise in the job finding rate λ (and in market tightness θ). Then, we can

use equation (B.12) again to differentiate \hat{p} with respect to β :

$$\frac{\partial \hat{p}}{\partial \beta} = \left(1 - \frac{c\varphi'(\hat{p})}{rk}\right)^{-1} \left[\left(\frac{\partial \tilde{h}}{\partial \lambda} + \frac{c}{(1-\beta)\bar{a}(\lambda)^2} \frac{\partial \bar{a}}{\partial \lambda} \right) \frac{\partial \lambda}{\partial \beta} - \frac{c}{(1-\beta)^2 \bar{a}(\lambda)} \right].$$

Since λ decreases with β , the term in the square brackets is negative when there is no selection into the wage sector ($s_H = 1$ and $\frac{\partial \bar{a}}{\partial \lambda} = 0$). If this is not the case, the overall effect will again depend on the model's parameters.

(e) *The share of high type workers in the population*

A higher share of more productive workers in the population, i.e. a decrease in α , would increase the average ability of a randomly met worker, $\bar{a}(\lambda)$, and thus tend to move the FE condition upwards and the ZP condition downwards. However, there is an additional effect that a fall in α has on the ZP condition by changing the total efficiency units of labour in the self-employment sector, L_H , that in turn affects income in that sector, \tilde{h} . That is,

$$\frac{\partial L_H}{\partial \alpha} = \frac{\delta}{\delta + \lambda} \left(-(a_H - a_L) + a_H \frac{(s_H - 1)\lambda}{\delta + s_H \lambda} \right).$$

In our simulations, this derivative is typically negative, which implies that a drop in α increases L_H and reduces the marginal product of labour in the home sector, pushing the ZP curve further down. However, for some parameter values it can be positive meaning that L_H might actually fall when the share of high-type workers increases in the economy. This can happen if the difference in productivity $a_H - a_L$ is relatively small while s_H is relatively large, so that the direct effect from an increase in efficiency units in L_H is dominated by the indirect effect of high type workers exiting the home sector at a much higher rate. In the latter case, the outside option \tilde{h} will increase pushing the ZP condition upwards and the overall effect on the market tightness will be indeterminate.

For more reasonable parameter values, however, $\frac{\partial L_H}{\partial \alpha} < 0$, and the ZP condition moves downwards as the share of high type workers rises. As a result, the job finding rate will increase, leading to a larger wage sector, analogously to the drop in worker's bargaining power β or an increase in matching efficiency m . The effect on the productivity threshold \hat{p} will depend on the model's parameters as

$$\frac{\partial \hat{p}}{\partial \beta} = \left(1 - \frac{c\varphi'(\hat{p})}{rk}\right)^{-1} \left[\frac{\partial \tilde{h}}{\partial L_H} \left(\frac{\partial L_H}{\partial \alpha} + \frac{\partial L_H}{\partial \lambda} \frac{\partial \lambda}{\partial \alpha} \right) + \frac{c}{(1-\beta)\bar{a}(\lambda)^2} \left(\frac{\partial \bar{a}}{\partial \alpha} + \frac{\partial \bar{a}}{\partial \lambda} \frac{\partial \lambda}{\partial \alpha} \right) \right].$$

Given that $\frac{\partial L_H}{\partial \alpha} < 0$, $\frac{\partial L_H}{\partial \lambda} < 0$ and $\frac{\partial \lambda}{\partial \alpha} < 0$, we cannot unambiguously tell if the sum in the first round brackets is positive or negative. Similarly, both $\frac{\partial \bar{a}}{\partial \alpha}$ and $\frac{\partial \bar{a}}{\partial \lambda}$ are negative, which means that we cannot sign the sum in the second round brackets either. Therefore, it is not possible to determine the effect of α on the productivity threshold in general. In fact, in our simulations \hat{p} increases for some countries and decreases for others, which is reflected in Table D.2 in that the elasticity of the wage rate with respect α is positive, for example, in Niger but negative in Ethiopia.

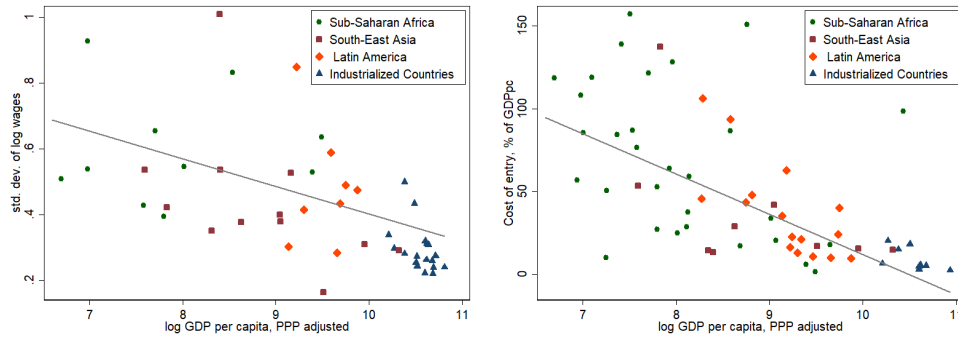
C Empirical Evidence

C.1 Additional empirical evidence

(a) Wage dispersion and development

It has been well documented in the literature that there is a large productivity gap between agriculture and manufacturing in developing countries. However, even within the manufacturing sector the wage dispersion in poor countries is much higher than in the developed world. To show this, we look at the variance of log wages across 4-digit industry groups using the INDSTAT database. These data are not limited to Sub-Saharan Africa and include both developing and industrialized countries. The left panel of Figure C.1 confirms a strong negative relationship

Figure C.1: *Wage dispersion, entry costs and development*



Note: The left panel plots the standard deviation of log wages in the manufacturing sector, derived at 4-digit industry level. Source: INDSTAT4 2004-2013 (Industrial Statistics Database), the United Nations Industrial Development Organization. The right panel shows the correlation between legal entry costs and the level of development. Source: The World Bank's Doing Business survey for 2010.

between country's wage inequality in the manufacturing sector and the level of GDP per capita (or the average wage).

(b) Entry costs and development

In line with our model, the right panel of Figure C.1 shows a negative relationship between the entry costs (measured as the legal entry costs as a share of GDP per capita) and a country's level of GDP per capita. The entry costs are derived from the World Bank's Doing Business survey that records the costs of all procedures officially required an entrepreneur to start up a commercial business (such as obtaining all necessary approvals, licenses and permits from the relevant authorities).

(c) Frequency of payment

In this section we provide information on frequency of payment and the extent to which it can explain differences in residual wage distributions. Table C.1 shows frequency of payment by country. At the bottom of the table, we compare the standard deviation of residual log wages with and without controlling for frequency of payment.

In Figure C.2 we show log wage residuals, after controlling for workers' demographics, industry, region, urban/rural areas by frequency of payment for Niger and Uganda to illustrate the point that wage distributions do not seem to be driven by how workers report their payments.

Table C.1: *Frequency of payment*

	Niger	Uganda	Tanzania	Ethiopia	Nigeria	South Africa
Percentage of wage earners by frequency of payment						
Daily	20.3	32.6	22.0	14.0	11.5	–
Weekly or fortnightly	7.4	15.5	11.8	11.8	7.0	22.2
Monthly	42.0	51.8	65.8	72.3	80.4	77.6
Quarterly	–	–	0.06	0.47	0.71	–
Semi-annually	–	–	0.03	0.83	0.03	–
Annually	30.3	–	0.32	0.54	0.33	0.23
Std. deviation of log wage residual, controlling for payment frequency						
No	0.962	0.799	0.701	0.762	0.695	0.579
Yes	0.842	0.714	0.690	0.711	0.679	0.578

Note: The sample is limited to 15-65 year old private sector employees. Log wage residuals are obtained from wage regressions, controlling for workers' demographics, industry, region, and urban/rural areas.

C.2 Case study: Uganda

Our model predicts that search frictions and entry costs can generate a negative correlation of wage inequality with the wage sector and the average wage. We show that this prediction can hold also across regions within a single country when there is sufficient interregional variation in market frictions. For this purpose, we need a sufficient variation across regions (or over time) and the regions should be large enough (in terms of wage employees) to have a meaningful measure of wage dispersion. We focus on Uganda, for which we have four waves of data and four regions (Central, Northern, Western and Eastern) that we split into rural and urban areas. Hence, we end up with eight regions with the sample of wage

workers ranging from about 30 in Northern urban region to about 250 in Central urban areas (which includes the capital city Kampala).

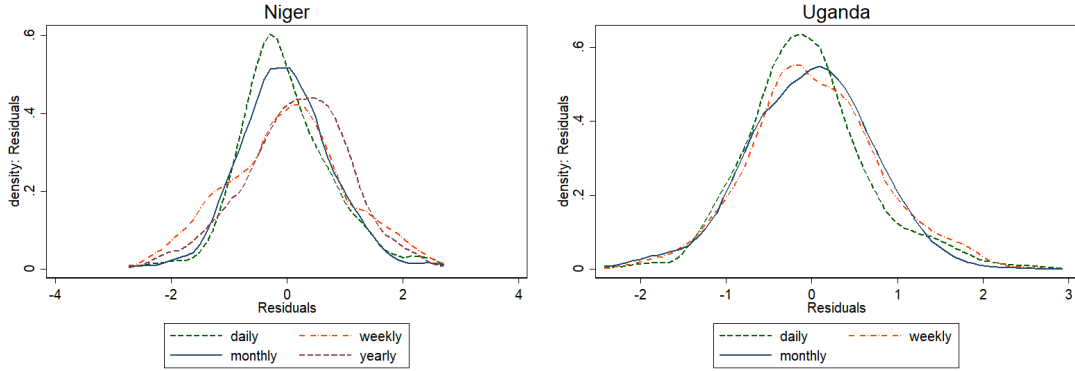
First, there is a clear negative relationship between the residual wage dispersion (after controlling for observed worker's characteristics) and the mean wage, in line with the model's predictions (see the upper left panel of Figure C.3). Second, we show that regions with a larger wage sector have higher wages, on average (the upper right panel).

It is more difficult to get firm entry costs or search frictions data at the region level. As we argue in Section 4, electricity connection costs seem to be a reasonable metric for the entry costs. In Uganda, more than 80% of electricity is generated by hydro-power and there are substantial regional differences in access to electricity, especially between rural and urban areas.³⁸ Since the Doing Business dataset only provides country-wide data for entry costs in Uganda, we use the share of households in a given region that have electricity connection as a proxy for entry costs. This measure is not ideal (as it shows the overall stock instead of new connections) but it is likely to be inversely related to the costs of connecting to the grid. The lower left panel of Figure C.3 shows that regions with a higher propensity of households with power (hence possibly lower entry costs for firms) are associated with a lower wage dispersion, supporting our model.

Finally, the bottom right panel of Figure C.3 plots a negative relationship (although less strong) between the wage dispersion and the share of communities with tarmac roads. Based on the evidence presented in Abebe *et al.* (Forthcoming), transportation costs act as a significant barrier in the labour market preventing workers from searching for stable jobs in Ethiopia. Therefore, we expect that better connected regions are less segmented and thus have higher matching efficiency

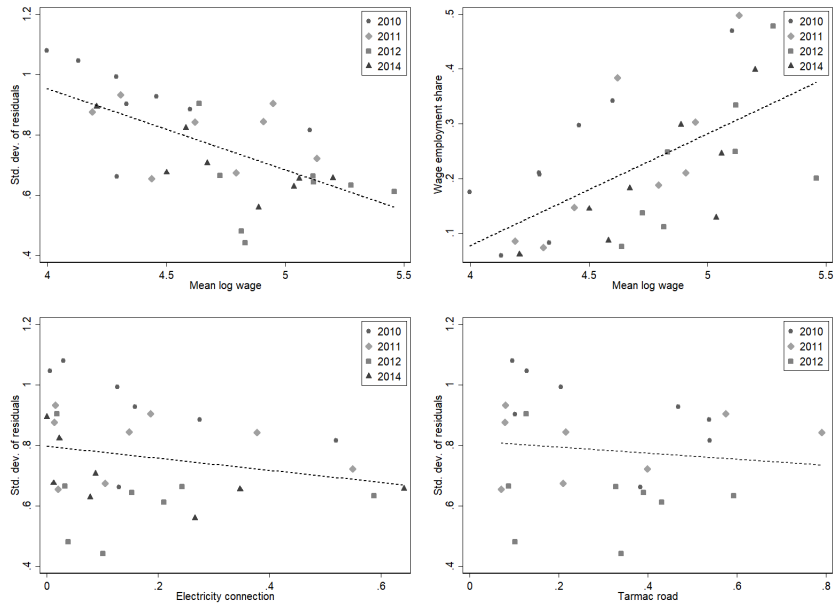
³⁸See for example <https://www.usaid.gov/powerafrica/uganda>.

Figure C.2: Residual wage dispersion by frequency of payment



Note: Log wage residuals are obtained after controlling for workers' demographics, industry, region, urban/rural areas and the frequency of payment.

Figure C.3: Wage dispersion, wages and the size of the wage sector in Uganda



Source: Authors' computations based on Uganda National Panel Survey Waves 1-4. The sample is limited to 15-65 year old individuals, excluding public sector employees. Monthly wages are deflated using CPI and expressed in constant 2010 PPP dollars. Residual wage dispersion is obtained from the residuals of a wage regression that controls for demographics (gender, age, age squared, marital status, education), regions, urban status, and industry. Electricity connection is the propensity of households within the region that have electricity. The tarmac road variable shows the average share of communities within the region that have a tarmac road.

(corresponding to a higher value of m in the model). The data on roads is only available for the first three waves. Overall, the data supports our hypothesis that the regions with a higher fraction of tarmac roads exhibit more compressed wage distributions.

D Estimation

D.1 Empirical moments

The estimation methodology we use is based on the indirect inference approach, as described in Gourieroux *et al.* (1993). In that approach the estimated parameters minimize the distance between the structural model and the auxiliary model, used to summarize the key elements of the data. In this paper, the auxiliary model is a set of empirical moments that we describe in detail below.

We use an exponential distribution for underlying firm types $\Gamma(p)$ with the mean and standard deviation of σ , a parameter that we estimate.³⁹ We exogenously set the monthly interest rate r to 1.25% and the elasticity of the matching function with respect to vacancies to $\eta = 0.5$.

Empirical targets

We use the following eleven moments to recover eleven parameters in the model (i.e. the model is just identified and fits the moments perfectly): the share of low type workers in the population, the mean log wages for the two worker types, the standard deviation of log wage residuals, the self-employment share for the two worker types, the transition rate from employment to the home sector, the labour share in value added, the labour elasticity in agricultural production, the entry

³⁹We have tried several distributions of the family with a log-concave density (including normal, logistic, Weibull, etc.) and have found that a Gamma distribution with a shape parameter of 1 (which is equivalent to an exponential distribution) performed the best.

costs per worker, and the average hiring costs.

- (a) *The share of low-type workers:* We rely on educational attainment to determine worker's type based on the selection equation in Table A.1. Table A.2 shows the share of highly educated workers for each country, which is equal to $1 - \alpha$ in the model.
- (b) *Mean wages:* The mean log wage is derived from the household surveys for all private sector employees aged 15-65. We use the average log wages for less educated (a_L -type) and more educated (a_H -type) workers as estimation targets.
- (c) *The standard deviation of wages:* We run a wage regression for each country controlling for workers' education, age and age squared, rural/urban dummy, regions and industries. The remaining residual dispersion corresponds to a firm-generated variance of log wages, $Var(\ln \tilde{w}(p))$.
- (d) *Self-employment share and transition rates:* We estimate the job destruction rate, δ , from the transitions between the wage sector and self-employment using the panel structure of the data. Using the self-employment rate for less educated workers, we can then recover the job finding rate and hence the matching efficiency, m . The difference in employment rates among the two types of workers identifies the relative search intensity. In particular, using equations (14) and (15), we can show that $\frac{e_H}{u_H} / \frac{e_L}{u_L} = \frac{s_H \lambda}{\delta} / \frac{\lambda}{\delta} = s_H$.
- (e) *Labor share:* We use the average labor share in the model to back out workers' bargaining power, β . In particular,

$$E \left[\frac{\sum_{i \in j} w_i}{y_j} \right] = E \left[\frac{\tilde{w}_j \sum_{i \in j} a_i}{p_j \sum_{i \in j} a_i} \right] = E \left[\frac{\tilde{w}_j}{p_j} \right] = (1 - \beta) E \left[\frac{\tilde{h}(\theta)}{p} \right] + \beta,$$

where $\sum_{i \in j} w_i$ is the total wage bill of firm j and $y_j = p_j \sum_{i \in j} a_i$ is the total

production. Given the proportionality assumption, the average labor share is determined by the productivity cutoff \hat{p} and the outside option \tilde{h} and is independent of workers ability composition within a firm.

To obtain the empirical counterpart of the labour share, we use a standardized international dataset of firm-level information drawn from the Enterprise Survey data. The Enterprise Survey collects firm-level data from business owners and top managers and covers a broad range of topics including firm's costs, employment, and performance measures.⁴⁰ First, we construct the value added series for each firm as the value of sales less purchases of raw materials and intermediate goods, as well as the costs of fuel, electricity and telecommunication. We then compute the labour share at the firm level as the ratio of the labour costs to the value added and take the average over the sample of firms. The labor share ranges between 0.14 in Niger to 0.42 in South Africa. Note that these values are likely to overestimate the true share of the production surplus paid to workers since (i) the labour share is derived from the firm's total labour costs that include payroll taxes, pension contributions, etc., and (ii) the Earnings Survey is limited to the formal sector firms that are likely to be larger and to employ better-qualified workers.

- (f) *Returns to scale in home production:* Instead of measuring self-employment income directly, we choose to use agricultural production data to obtain the returns to scale parameter γ . We utilize the aggregate data on the value added per worker in agriculture for the time period of 1990-2012. In particular, we run the following regression across SSA countries:

$$\ln Y_{Hjt} = b_0 + b_1 \ln L_{Hjt} + b_2 \ln T_{jt} + v_{jt}, \quad (\text{D.1})$$

⁴⁰See <http://www.enterprisesurveys.org> for details. The data are available for Ethiopia in 2015, Niger in 2009, Nigeria 2014, South Africa in 2007, Tanzania and Uganda in 2013.

where Y_{Hjt} is the value of agricultural production in country j at year t , L_{Hjt} is the number of workers employed in agriculture, T_{jt} is land in hectares, b' s are the coefficients to be estimated, and v_{jt} is the error term. We also control for year and country fixed effects. The parameter of interest can be recovered from $\gamma = \hat{b}_1$ and is equal to 0.246. We assume that the returns to scale parameter γ is the same across the six countries we analyse, while the overall home sector productivity A is allowed to vary and is recovered from the model directly. In section D.4. below, we report the model estimates based on an alternative value of γ , estimated from the cereal yield, and show that the results are virtually the same.

Note that in the model labour in the home sector is measured in efficiency units, i.e. $L_H = a_L * u_L + a_H * u_H$. Given that the share of highly educated workers among the self-employed is less than 10% in our sample (with the exception of South Africa, where it is equal to 30%), we consider the total number of workers in agriculture to be a good proxy for L_H . Moreover, as long as the share of high type workers in the home sector and their relative productivity does not change significantly over time, the differential quality weights across countries will be absorbed by the country fixed effects.

- (g) *Entry costs*: In the model, high entry costs and labour market frictions (i.e. low matching efficiency) have the same qualitative implications for the size of the paid employment sector, mean wages, and wage dispersion and therefore cannot be identified separately. We use the costs of getting electricity connection to proxy the entry costs in the model and use other data moments to pin down the matching efficiency parameter. We consider the costs of electricity connection to be a more tangible estimate than the legal fees of starting a business; however, it should be thought of as a lower bound on the actual entry costs as

Table D.1: *Entry costs*

Country	Electricity connection cost, % GDP per capita	Average firm size	Cost per worker, % of GDP per capita
Niger	7,424	16.5	449.0
Uganda	11,259	18.5	609.2
Tanzania	2,496	23.9	104.6
Ethiopia	2,040	38.2	53.4
Nigeria	1,040	20.8	49.9
South Africa	875	57.5	15.2

The costs of electricity connection is based on the World Bank's Doing Business Surveys available after 2010. We computed the average cost across the following years: Ethiopia 2013-2015; Niger 2011 and 2014; Nigeria 2010-2016; South Africa 2010; Tanzania 2010-2012; Uganda 2010-2014. The average firm size is based on the Enterprise Survey data.

we abstract from other costs (e.g. credit constraints) that we cannot measure well.

The entry barrier k is expressed as per-worker costs in the model. Hence, to get its equivalent in the data we divide the electricity connection costs by the average firm size derived from the Enterprise Surveys. As the Enterprise Surveys data covers only formal firms, the firm size is likely to be greater than the number of workers at an average firm in the economy, which again understates the magnitude of the entry costs. The data used for the estimation are summarized in Table D.1.

- (h) *Hiring costs*: Given that there is no consensus in the literature on the magnitude of hiring costs even for industrialized countries, we choose our target ad hoc to be equal to one month of wages on average and run a sensitivity analysis with three months of wages.⁴¹

⁴¹The estimates of hiring costs - including recruiting, training and monitoring costs - vary across industrialized countries. Silva and Toledo (2009), for example, find that recruiting costs are 14 percent of quarterly pay per hire in the US, or about half of monthly wages, based on data collected by PriceWaterhouseCoopers. Abowd and Kramarz (2003) estimate firing and hiring costs directly based on survey data for a representative sample of French firms. They find that the average hiring costs (including the direct training costs) per hire are approximately equal to the median of monthly wages. In developing countries, the training costs might be lower if the jobs are less skill-intensive, on the other hand the recruiting and monitoring costs might be higher given the labour market inef-

Table 3 in Section 5 summarizes the empirical moments that we are matching in the estimation.

D.2 Estimation methodology

Denote by Ψ^* a set of the empirical moments that comprise our auxiliary model. To estimate the structural parameters, we solve for a steady state equilibrium and obtain the corresponding simulated moments generated by the model, $\Psi_s(\vartheta)$, where $\vartheta = (m, k, c, \sigma, A, \gamma, \beta, \delta, s_H, a_H, \alpha)$ is a given vector of parameters. Essentially, indirect inference approach can be viewed as an extension of the GMM, where the estimator is the choice of structural parameters that minimizes the weighted distance between the data moments and the simulated moments:

$$\hat{\vartheta} = \arg \min_{\vartheta \in \Theta} (\Psi_s(\vartheta) - \Psi^*)' \Omega (\Psi_s(\vartheta) - \Psi^*),$$

where Ω represents the weighting matrix. The optimal weighting matrix is the inverse of the variance-covariance matrix of the data moments. However, since the model is exact-identified the choice of the weighting matrix is irrelevant.

iciencies.

In our case, the set of auxiliary (de-meaned) moment conditions are given by

$$\Psi = E(g) = E \begin{pmatrix} 1[E_{it} = 1]1[a_i = a_H](\ln w_{it} - \mu_1) \\ 1[E_{it} = 1]1[a_i = a_L](\ln w_{it} - \mu_2) \\ 1[E_{it} = 1](\varepsilon_{it}^2 - \mu_3) \\ 1[SE_{it} = 1]1[a_i = a_H] - \mu_4 \\ 1[SE_{it} = 1]1[a_i = a_L] - \mu_5 \\ 1[a_i = a_L] - \mu_6 \\ 1[E_{it-1} = 1](1[SE_{it} = 1] - \mu_7) \\ 1[E_{it} = 1]\left(\frac{\bar{w}_{it}}{p_{jt}} - \mu_8\right) \\ \gamma - \mu_9 \end{pmatrix} = 0,$$

where $1[\cdot]$ is an indicator function, w_{it} is monthly wage of individual i at time t , ε_{it} is an error term from the log wage regression, E_{it} is employment state and SE_{it} is self-employment, p_{jt} is the value added of the firm j at time t , γ is the labour elasticity in the home production sector, and μ 's are the corresponding population means. There are two additional moments that we use for the estimation: (i) the average hiring costs in the model are equal to one month of wages, i.e. $\frac{c}{q} = E(w_{it})$ and (ii) the entry costs per worker as a fraction of total output k/Y are given by the electricity connection costs in the data. These two moments are taken as deterministic since we cannot measure their variance in the data.

The corresponding empirical moments Ψ^* can be obtained as a sample average from the micro data. For example, the first moment is $\frac{N_E^H}{N} \frac{1}{N_E^H} \sum (\ln w_{it} - \hat{\mu}_1)$, where $\hat{\mu}_1$ is the sample average of log wages among high-type wage employees and N_E^H is their corresponding number. The last moment condition is based on estimating γ coefficient from the OLS regression described in equation (D.1).

Using the optimal weighting matrix $\Omega = S^{-1}$, the asymptotic variance of ϑ is

given by

$$\frac{1}{N}(D'S^{-1}D)^{-1}, \quad (\text{D.2})$$

where $S = E(gg')$ is the variance-covariance matrix of auxiliary moments and $D = E(\partial g(\vartheta)/\partial \vartheta')$. We obtain D using numerical derivatives from the model and evaluate them at $\hat{\vartheta}$.

The variance-covariance matrix S is given by:

$$\left(\begin{array}{cccc} e_H \text{Var}(\ln \tilde{w}_{it}) & & & \\ 0 & e_L \text{Var}(\ln \tilde{w}_{it}) & & \\ e_H \text{Cov}(\ln \tilde{w}_{it}, \varepsilon_{it}^2) & e_L \text{Cov}(\ln \tilde{w}_{it}, \varepsilon_{it}^2) & (e_H + e_L) \text{Var}(\varepsilon_{it}^2) & \\ 0 & 0 & 0 & u_H(1 - u_H) \\ 0 & 0 & 0 & -u_H u_L \\ 0 & 0 & 0 & -\alpha u_H \\ 0 & 0 & 0 & e_H(1 - \delta)\delta \\ e_H \text{Cov}(\ln \tilde{w}_{it}, \frac{\tilde{w}_{it}}{p_{it}}) & e_L \text{Cov}(\ln \tilde{w}_{it}, \frac{\tilde{w}_{it}}{p_{it}}) & (e_H + e_L) \text{cov}(\varepsilon_{it}^2, \frac{\tilde{w}_{it}}{p_{it}}) & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ u_L(1 - u_L) & & & \\ (1 - \alpha)u_L & \alpha(1 - \alpha) & & \\ \dots & e_L(1 - \delta)\delta & 0 & (e_H + e_L)\delta(1 - \delta) \\ 0 & 0 & 0 & (e_H + e_L) \text{Var}(\frac{\tilde{w}_{it}}{p_{it}}) \\ 0 & 0 & 0 & 0 & \text{Var}(\hat{\gamma}) \end{array} \right).$$

Note that we need to find the covariance between log wages and squared residuals and the labor share; however, latter is obtained from the firm-level data that lacks the information on workers' characteristics. Thus, both the level (due to the

wage bill including other labour costs) and the variance of firm-level wages will be different between the two datasets. To be consistent within the model, we use the variance of residuals from individual wage regression and compute the covariance as

$$\text{cov} \left(\frac{\tilde{w}}{p}, \ln \tilde{w} \right) = \text{corr} \left(\frac{\tilde{w}}{p}, \ln \tilde{w} \right) \sqrt{\text{Var} \left(\frac{\tilde{w}}{p} \right)} \sqrt{\text{Var}(\ln \tilde{w})}.$$

The correlation coefficient is estimated using the average log wage at the firm level, i.e. $\text{corr} \left(\ln \frac{\sum_{i \in j} w_i}{n_j}, \frac{\tilde{w}_j}{p_j} \right) = \text{corr} \left(\ln \tilde{w}_j + \ln \frac{\sum_{i \in j} a_i}{n_j}, \frac{\tilde{w}}{p} \right) = \text{corr} \left(\ln \tilde{w}_j, \frac{\tilde{w}}{p} \right)$, since there is no sorting into different firms based on ability. Similarly, we use the squared mean deviations of the firm-level log wages to get $\text{corr} \left(\frac{\tilde{w}}{p}, \varepsilon^2 \right)$.

Finally, we use the sample analogue of S as

$$S_N = \frac{1}{N} \sum g^* g^{*'},$$

where the last entry of S is the asymptotic variance of $\hat{\gamma}$ from the OLS regression. In that way, we account for extra statistical uncertainty coming from pre-estimation of γ .

D.3 Outcome elasticities

We can simulate changes in model's parameters to illustrate what happens to the size of wage sector, the average wage rate and its dispersion, as well as the overall income and income inequality. Table D.2 summarizes our estimates for all six countries.

D.4 Sensitivity analysis

Here, we check the sensitivity of our analysis to the choice of pre-determined parameters. For illustration purposes, Table D.3 shows how the estimates for Uganda

Table D.2: Elasticities of outcome variables with respect to changes in parameters

Country	Elasticity w.r.t	Wage employment	Average wage rate	St. dev of log wage rate	Average income	St. dev of log income
Niger	<i>k</i>	-0.91	-0.01	0.02	-0.50	-0.33
	<i>m</i>	1.77	0.01	-0.02	0.97	0.62
	<i>A</i>	-0.01	0.06	-0.15	0.45	-0.25
	σ	0.93	0.94	0.14	1.06	0.57
	α	-7.97	0.09	-0.22	-7.90	-6.00
Uganda	<i>k</i>	-0.94	-0.03	0.06	-0.55	-0.30
	<i>m</i>	1.50	0.04	-0.06	0.87	0.46
	<i>A</i>	-0.03	0.14	-0.24	0.45	-0.36
	σ	0.98	0.88	0.20	1.11	0.66
	α	-4.43	-0.01	0.01	-4.34	-2.94
Tanzania	<i>k</i>	-0.88	-0.02	0.03	-0.24	-0.34
	<i>m</i>	1.76	0.02	-0.04	0.48	0.67
	<i>A</i>	-0.04	0.22	-0.30	0.76	-0.40
	σ	0.93	0.80	0.29	0.49	0.74
	α	-4.76	0.05	-0.07	-2.06	-3.74
Ethiopia	<i>k</i>	-0.82	-0.02	0.03	-0.33	-0.33
	<i>m</i>	1.65	0.03	-0.04	0.65	0.65
	<i>A</i>	-0.02	0.17	-0.27	0.65	-0.41
	σ	0.85	0.84	0.25	0.68	0.74
	α	-4.89	-0.03	0.05	-2.71	-3.01
Nigeria	<i>k</i>	-0.83	-0.02	0.03	-0.31	-0.31
	<i>m</i>	1.66	0.04	-0.05	0.61	0.61
	<i>A</i>	-0.03	0.22	-0.31	0.68	-0.42
	σ	0.87	0.80	0.28	0.63	0.74
	α	-2.59	0.01	-0.01	-1.35	-1.62
South Africa	<i>k</i>	-0.35	-0.15	0.18	-0.38	0.14
	<i>m</i>	0.68	0.27	-0.33	0.73	-0.28
	<i>A</i>	-0.04	0.30	-0.35	0.37	-0.32
	σ	0.39	0.84	0.18	1.01	0.18
	α	-0.55	-0.02	0.03	-1.19	-0.16

The elasticity with respect to α shows the percentage change in the outcome variable if α increases by one percentage point.

Table D.3: *Estimates for Uganda under alternative specifications*

	σ	A	β	m	k	c
Baseline	734.0 (19.1)	71.8 (21.5)	0.155 (0.004)	0.0082 (0.0007)	12912	2.03
Returns to scale in subsistence, $\gamma = 0.14$	734.0 (19.1)	123.7 (90.1)	0.155 (0.004)	0.0082 (0.0007)	12912	2.03
Average hiring costs, $\frac{c}{q} = 3E(w)$	678.2 (18.6)	62.5 (18.9)	0.168 (0.005)	0.0084 (0.0007)	11942	6.40
Matching elasticity, $\eta = 0.3$	734.0 (19.1)	71.8 (21.5)	0.155 (0.004)	0.0070 (0.0006)	12912	2.03
Yearly interest rate, $r = 5\%$	740.3 (19.2)	72.5 (21.7)	0.154 (0.004)	0.0039 (0.0004)	13005	0.46

Note: Asymptotic standard errors are given in parentheses. Other parameters estimates, δ , s_H , α and a_H , remain unchanged in all of these specifications and hence are not reported here. The parameter estimates that change significantly for each sensitivity check are highlighted in bold.

change under four alternative specifications. In addition, we reproduce Table 5 below that reports the variance decomposition of log wage rates for Niger and Uganda relative to South Africa for each of the four cases.

(a) *Returns to scale in home production, γ* . The first robustness check that we perform is using the cereal yield in kilos per hectare as a dependent variable in regression (D.1) (cereals are the main crop in these countries). This alternative specification results in $\hat{\gamma} = 0.14$, which is lower than our baseline of 0.25. Using $\hat{\gamma} = 0.14$ increases our estimate of the home sector productivity parameter A (so that the implied home production income $\tilde{h} = \gamma A L_H^{\gamma-1}$ stays the same), while leaving all other parameter values unchanged. The results of the policy experiments and the obtained elasticities of employment, wages and income remain virtually the same.

(b) *Hiring costs*. Next, we assume that the average hiring costs are equal to three months of wages, as opposed to the baseline of one month of wages, which implies that the value for c has now tripled. In addition, from equation (ZP), we know that higher hiring costs $\frac{c}{q}$ imply a higher productivity threshold, \hat{p} . As a result, in order to fit the average wages in the data, the model now produces lower estimates of A (home productivity) and σ (the average firm productivity) than in the baseline case

(see Table D.3). Other parameter values adjust as well, however, the difference in estimates is not substantial (less than 10% for most parameters) and the obtained outcome elasticities are very close to the baseline case.

(c) *Matching function elasticity with respect to vacancies, η* . We run a robustness check with $\eta = 0.3$, which is at the lower spectrum of the estimates found in the literature (see Petrongolo and Pissarides, 2001). The only parameter that changes in this case is the estimate of the matching efficiency parameter, m . Table D.4 for Uganda shows that wage employment elasticity falls when we assume $\eta = 0.3$, compared to the baseline case of 0.5. This result is intuitive, as a lower elasticity of matches with respect to vacancies means that a reduction in market frictions or an increase in skills are translated into job creation (and higher wages and incomes) to a lesser extent.

Table D.4: *Wage employment elasticity for Uganda for different values of the matching function elasticity, η*

Elasticity with respect to	$\eta = 0.5$	$\eta = 0.3$
k	-0.94	-0.42
m	1.50	1.11
A	-0.03	-0.01
σ	0.98	0.43
α	-4.43	-2.94

The elasticity with respect to α shows the percentage change in wage employment if α increases by one percentage point.

In terms of the role of frictions for determining the variance of wages, the largest difference compared to the baseline is observed when we assume that $\eta = 0.3$ (see Table D.5). Note that the two empirical targets that we require the model to match - the destruction rate δ and the self-employment rate among the less educated workers u_L/α - determine the job finding rate λ in steady state, according to equation (14). Then, for $\lambda = m\theta^\eta$ to remain the same when we reduce η to 0.3, the matching efficiency estimate m falls in the countries where the implied value of θ is less than

Table D.5: *Sensitivity analysis: Percentage of the variance gap in log wage rates relative to South Africa that is explained by frictions*

		k	m	$k\&m$	Other pars
Baseline	Niger	6%	10%	21%	79%
	Uganda	65%	-8%	45%	55%
Returns to scale in subsistence, $\gamma = 0.14$	Niger	7%	11%	23%	77%
	Uganda	70%	-9%	48%	52%
Average hiring costs, $\frac{c}{q} = 3E(w)$	Niger	14%	8%	25%	75%
	Uganda	86%	-6%	75%	25%
Matching elasticity, $\eta = 0.3$	Niger	5%	17%	25%	75%
	Uganda	40%	11%	54%	46%
Yearly interest rate, $r = 5\%$	Niger	6%	9%	20%	80%
	Uganda	56%	-7%	38%	62%

one (Niger and Uganda) and increases in the economies with $\theta > 1$ (Ethiopia, Tanzania, Nigeria and South Africa). This means that the gap in matching efficiency between Uganda (or Niger) and South Africa is now larger. As a result, a larger share of the relative variance gap between the two countries can be attributed to matching frictions.

(d) *Interest rate, r .* In the baseline model, we use the annual interest rate r of 15% to reflect high borrowing costs in developing countries. To check the sensitivity of our results to this assumption, we run our estimation under a more common assumption of 5% annual interest rate. From equation (FE) we know that a reduction in r lowers the flow cost of entering rk , which would require a fall in the vacancy filling rate q . Recall that the filling rate can be written as $q(\lambda) = m^{\frac{1}{\eta}} \lambda^{\frac{\eta-1}{\eta}}$. Given that the job finding rate λ is fixed in steady state as we have argued above, the resulting matching efficiency parameter has to be lower. Our estimates suggest that m falls by about a half when $r = 5\%$ annually; however, most outcome elasticities remain virtually the same.

In general, our results are robust to different specifications. Table D.5 suggests

that about a half of the relative variance gap in wage rates in Uganda and one quarter in Niger can be explained by differences in market frictions relative to South Africa across all specifications. In terms of wage employment elasticity, we consistently find that a reduction in the entry costs is about half as effective for job creation as an increase in matching efficiency. Moreover, the lack of skills remains a significant barrier to job creation in the poorest countries in our sample.