Wage Dispersion, Job Creation and Development *

Juan Pablo Rud † Ija Trapeznikova‡

October 25, 2016

Abstract

Least developed economies are characterized by poorly functioning labor markets: only a small fraction of the labor force is in paid employment, and both productivity and wages are very low. In this paper, we provide a new empirical fact: wage dispersion in these countries is high and negatively correlated with both the size of the wage sector and mean wages. While the small size of the wage sector points towards the presence of entry barriers, a high wage dispersion suggests the presence of labor market frictions. We build an integrated framework that incorporates the tools of standard search literature into a traditional two-sector model of development. We estimate the model using micro data for a number of countries in Sub-Saharan Africa. Our results highlight the empirical relevance of labor market frictions and their interactions with firm entry costs for job creation, productivity, wages, and wage dispersion. We also show how our estimates can be used to assess cost-effectiveness of labor market policies.

Keywords: wage dispersion, job creation, inequality, job search

JEL: J21, J31, J64, O11, O15

---

*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, Jon Temple, Gianluca Violante, Yves Zenou and seminar participants at Alicante, Chicago Fed, Edinburgh, Essex, Federal Reserve Board, RHUL, St. Gallen, St. Louis Fed, Sussex, Universidad Di Tella, 2nd African Search and Matching Workshop, CSAFE, EEA-ESEM, IFS-EdePo, Novafrica, Sandbjerg and SED Conference for useful comments and suggestions. We are grateful to Taye Mengistae at World Bank for helping us procure the data.

†Department of Economics, Royal Holloway, University of London. Contact: juan.rud@rhul.ac.uk

‡Department of Economics, Royal Holloway, University of London. Contact: ija.trapeznikova@rhul.ac.uk
1 Introduction

Least developed countries are characterized by an extremely small wage sector. In many Sub-Saharan African (SSA) countries, up to nine out of ten workers are engaged in own-account work or helping family activities for no pay. Moreover, most of these activities are limited to low-earning occupations, such as subsistence farming and petty trade. While the wage sector in these countries is relatively more productive and more desirable for workers than self-employment, there are not enough jobs available for people willing to work for a wage\footnote{For example, Fields\citeyear{fields2011} 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\citeyear{banerjee2007} write: “If they [petty entrepreneurs] could only find the right salaried job, they might be quite content to shut their business down.”}. It is not surprising then that a lot of the effort by international organisations and donors has focused on job creation through private sector development (see, for example, World Bank\citeyear{worldbank2014}).

While the expansion of the private sector seems to be a natural policy recommendation, there are many indications that it does not function well. First, the wage sector in poor economies has low productivity. For example, labor productivity in Sub-Saharan Africa is, on average, fourteen times lower than in advanced economies and four times lower than in Latin America\footnote{See ILO\citeyear{ilo2012}.} Second, we find that the wage sector in these countries is characterized by high levels of wage dispersion - a fact that has not been documented before. This evidence suggests that not only there are barriers to firm entry that prevent job creation, but that there are other obstacles, such as labor market frictions, that prevent workers from moving to the highest-paying firms within the wage sector. Hence, standard policies aiming solely at expanding the wage sector may be insufficient. In this paper, we propose a unifying framework that allows us to analyze job creation, productivity and wage dispersion simultaneously. We then estimate the model using micro data from a number of countries in SSA to examine the empirical relevance of various constraints to job creation and wage growth.

Our first contribution is to provide new empirical evidence that seems to be intrinsically linked to labor market underperformance in poor economies. Using detailed household level data, we document that wage dispersion is greater in countries with smaller wage sectors and lower average wages (and that it holds even after accounting for differences in workers’ characteristics). This relationship seems to be at odds with the well-established fact that there is a pos-
itive cross-country correlation between average income and inequality among the poorest countries.\textsuperscript{3} We show that high \textit{wage} inequality and low \textit{income} inequality can coexist in a country, when a large fraction of the workforce is engaged in low-income non-wage activities.\textsuperscript{4}

The second innovation of this paper is to propose an integrated model of home production and the wage sector that allows for both entry barriers and labor market frictions. In particular, we incorporate the tools of a standard search and matching framework (Mortensen and Pissarides 1994) 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 labor for production and are subject to entry costs. Labor 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 home production, or a self-employment sector. Finally, a wage bargaining process links wages to firms’ productivity and the worker’s outside option (i.e. self-employment income). Our model predicts that both entry barriers and labor market frictions 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. At the same time, we 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 big role in shaping labor market outcomes.

We proceed to estimate the model for five SSA countries: low-income Ghana, Nigeria, Tanzania, and Uganda, and middle-income South Africa, which is regarded here as a reference economy. For the estimation, we use individual-level panel data that not only have detailed information on workers’ demographic and employment characteristics, but also allow us to construct transitions between the private sector and self-employment activities. Our main set of results suggests that a reduction in labor market frictions has the largest impact on job creation: a one percent increase in labor market efficiency leads to a 0.8 percent increase in wage employment in South Africa and almost two percent increase in the other four countries in our sample; whereas the reduction in entry costs is only half as effective. Moreover, the effect of labor market frictions on wage inequality is amplified in the presence of higher barriers to entry.

\textsuperscript{3}Namely, the left section of the Kuznets curve.
\textsuperscript{4}The argument for the positive correlation between income and income inequality is similar to Robinson 1976, where inequality increases when workers move from a large, low-productivity sector to the small modern sector. Our innovation is to enrich the model of the wage sector to account for high wage dispersion at the same time.
Finally, we use our estimated parameters to analyze three types of interventions: a reduction in the entry costs, an increase in labor market efficiency, and an increase in self-employment productivity. Our counterfactual experiments show the enormous task that countries face if trying to catch up with richer economies. For example, in order to achieve the same size of the wage sector as in South Africa, Ghana needs to decrease its entry costs about nine times or, alternatively, increase its labor market efficiency three times. A much larger change is required if the policy aim is the average level of income: Ghana’s home sector productivity has to increase more than 8 times, the labor market needs to become roughly 9 times more efficient, or, alternatively, the entry costs have to be reduced 75 times. We also use our estimates to assess how countries could best use their resources (e.g. aid money) to improve income levels. For example, while $1 in firm entry subsidies in Ghana is more effective at increasing income levels than subsidizing self-employed income, the opposite is true in South Africa. More generally, our results suggest that a unifying model is a valuable analytical tool to inform debates about policy effectiveness and their complementarities. Moreover, it can be used to assess general equilibrium effects of existing randomized field experiments that focus on improving labor market outcomes.

Our framework builds on recent progress in the development and labor economics literature. The importance of productivity and wage dispersion as measures of labor market performance has been recognized and documented for industrialized economies yet, it remains under-explored in developing countries. Recent studies in development economics have established a number of factors that determine the size of the wage sector and the level of wages. How-

---

5 See, for example, recent studies on reducing search costs (Abebe, Caria, Fafchamps, Falco, Franklin and Quinn 2016 and Franklin 2015), improving firm entry (de Mel, McKenzie and Woodruff 2012 and de Mel, McKenzie and Woodruff 2013) or enhancing productivity in the home sector through asset transfers (Banerjee, Duflo, Goldberg, Karlan, Osei, Pariente, Shapiro, Thuysbaert and Udry 2015 and Blattman, Fiala and Martinez 2013).

6 See Lentz and Mortensen (2010) and Rogerson, Shimer and Wright (2005) for a survey of the literature, see Bontemps, Robin and van den Berg (2000) for France, Jolivet, Postel-Vinay and Robin (2000) for a set of European countries and the U.S.

7 A few notable exceptions are Amaral and Quintin (2006), Meghir, Narita and Robin (2015), Satchi and Temple (2009), Ulyssea (2010) and Zenou (2008). However, these papers focus primarily on the size and composition of the informal sector, i.e. non-registered wage workers, the distinction that only becomes relevant for middle-income countries.

8 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, La Porta, Lopez-de Silanes and Shleifer 2002 and Herrendorf and Teixeira 2011 among others). Misallocation of resources have also been identified as constraints to job creation and as important determinants of wages (see Hsieh and Klenow 2009 and Vollrath 2014 and Hsieh and Klenow 2010).
ever, looking at wage levels is not always sufficient to characterize labor markets: conditional on the same mean, a greater wage dispersion reflects a higher degree of market inefficiencies and hence is more informative about constraints to job creation and wage growth. Our paper aims at highlighting their importance for understanding labor markets in least developed countries, as well as for designing a structural framework for policy analysis.

The paper is organized as follows. In Section 2, we discuss a set of stylized facts that characterize labor markets in developing countries. In Section 3, we develop the model and derive its main predictions. In Section 4, we discuss our empirical strategy and the estimation procedure. Section 5 presents the estimated parameters for a set of countries in Sub-Saharan Africa, and Section 6 shows some counterfactual experiments. In Section 7 we conclude. Further information on the data sources used in the paper, key model assumptions, as well as a numerical simulation exercise that demonstrates the main drivers of the model, can be found in Appendix.

2 Labor markets in developing countries

2.1 Wage employment and average wages

It has been well documented in the literature that least developed countries are characterized by very low levels of wage employment (see Fields, 2011 for a review). In the left panel of Figure 1, we show that the share of wage employment increases as countries get richer and that in many Sub-Saharan African countries it is below 20%. However, it does not discourage labor force participation, which is greater in SSA (around 70%), than in OECD countries (around 60%). This evidence, coupled with the fact that income from self-employment activities tends to be lower than wages, suggests that in poorer countries there are not enough jobs for the number of people willing to work for a wage. As a consequence, most workers end up in less desirable self-employment occupations, own-account work or helping family activities for no pay.

Furthermore, the wage sector is characterized by relatively low levels of productivity and low earnings. For example, GDP per person employed in Sub-Saharan Africa is, on average, fourteen times lower than in advanced economies.
and four times lower than in Latin America. Even when focusing on the relatively high-productivity manufacturing sector, labor 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 labor share in total value added is almost twice as large in developed economies than in the SSA countries.

2.2 Wage dispersion and labor market performance

In this subsection, we use household survey data for a number of countries in Sub-Saharan Africa to analyze the wage sector in detail (see Appendix A for data description). Figure 2 plots log wage densities for a number of SSA countries and the US. This evidence suggests that the wage sector in developing countries exhibits high levels of wage dispersion and that dispersion decreases with mean wages and with levels of GDP per capita.

Table 1 presents different measures of wage dispersion. Countries are ordered from lowest to highest average log wages (column (1)) and are grouped in roughly four categories: (i) the poorest countries in the sample with GDP per capita less than $2000 in 2011 PPP dollars (Ethiopia, Tanzania, Uganda), (ii) less

10See ILO (2012). Note that these differences in productivity and pay are not explained by the composition of the labor 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 and find that the same person would earn on average more than 7 times when relocating from Ghana to the US and less than 3 times if coming from South Africa.
Figure 2: Log wage densities in Uganda, Ghana, South Africa and the US.

Source: Authors’ computations based on Uganda National Household Survey 2010, Ghana Living Standards Survey 2005, and Survey of Income Program Participation (SIPP) 2004 for the US. These calculations are based on individuals aged between 15 and 65, excluding public sector employees. Note that Uganda’s GDP per capita in 2005 PPP dollars is $1268, while Ghana’s is $2612 and South Africa’s, $10413.

poor ones with GDP per capita up to $3000 (Cameroon, Ghana, Kenya), (iii) more developed countries with average income between $3000-$5000 (Zambia and Nigeria), and (iv) South Africa and the US as a reference, with GDP per capita of more than $11,000 and $51,000, respectively.

Column (2) shows the unconditional standard deviation of log wages and, with some exceptions such as resource-rich Zambia and Nigeria, the general pattern shows a negative correlation between mean log wages and wage dispersion. This observation holds in column (3), when we use a Gini Index of wages. The next two columns explore whether the dispersion is coming from the top or the bottom of the distribution. Namely, column (4) reports the ratio between the median wage and the 10th percentile and column (5) looks at the ratio between the 90th percentile and the median. In general, there seems to be substantial wage differences for SSA countries at both tails of the wage distribution. In some cases both indicators are particularly high but relatively similar between them, such as Ethiopia, Tanzania and Uganda. In some others, such as Kenya or Zambia, differences are more important for high earners. Column (6) reports the dispersion of residuals from a log wage regression that controls for demographics (gender, age, education, marital status), location (rural/urban, region) and industry. Wage dispersion remains substantial after controlling for observables and, more importantly, the negative correlation between mean wages and residual wage dispersion holds. Finally, the highest wage dispersion seems to
Table 1: Wage and income distributions for Sub-Saharan African Countries and the US.

<table>
<thead>
<tr>
<th>Country</th>
<th>Mean log Wages (1)</th>
<th>Std. dev of log Wages (2)</th>
<th>Gini index (3)</th>
<th>Wage Ratio 50:10 (4)</th>
<th>Std. dev of Residuals (5)</th>
<th>Wage Ratio 90:50 (6)</th>
<th>Std. dev of Wage Size (7)</th>
<th>Gini index Income (8)</th>
<th>Std. dev of log Income (9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethiopia</td>
<td>3.93</td>
<td>1.06</td>
<td>0.48</td>
<td>5</td>
<td>0.70</td>
<td>0.40</td>
<td>0.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tanzania</td>
<td>4.14</td>
<td>1.11</td>
<td>0.57</td>
<td>3.6</td>
<td>0.87</td>
<td>0.41</td>
<td>0.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uganda</td>
<td>4.45</td>
<td>0.99</td>
<td>0.53</td>
<td>3.86</td>
<td>0.76</td>
<td>0.43</td>
<td>0.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kenya</td>
<td>4.63</td>
<td>0.98</td>
<td>0.53</td>
<td>3.1</td>
<td>0.74</td>
<td>0.30</td>
<td>0.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ghana</td>
<td>4.79</td>
<td>0.88</td>
<td>0.47</td>
<td>2.75</td>
<td>0.70</td>
<td>0.20</td>
<td>0.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cameroon</td>
<td>4.88</td>
<td>0.80</td>
<td>0.44</td>
<td>2.66</td>
<td>0.64</td>
<td>0.19</td>
<td>0.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zambia</td>
<td>4.94</td>
<td>1.00</td>
<td>0.57</td>
<td>2.75</td>
<td>0.62</td>
<td>0.25</td>
<td>0.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nigeria</td>
<td>4.96</td>
<td>1.02</td>
<td>0.65</td>
<td>3.2</td>
<td>0.75</td>
<td>0.07</td>
<td>0.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Africa</td>
<td>6.02</td>
<td>0.83</td>
<td>0.48</td>
<td>2.50</td>
<td>0.58</td>
<td>0.51</td>
<td>1.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>USA</td>
<td>7.76</td>
<td>0.74</td>
<td>0.38</td>
<td>2.67</td>
<td>0.60</td>
<td>0.87</td>
<td>0.55</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: For wage and wage sector size measures, the sample is limited to 16-65 year old private sector employees (as opposed to self-employed and unpaid family members). In South Africa and the US, the reference group also includes unemployed individuals. Mean log wages are expressed in constant 2005 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 dispersion of log income is calculated based on household consumption diaries. The Gini index for income (column 8) is drawn from the World Bank’s World Development Indicators database. The remaining measures are obtained from own calculations using the following datasets: Ethiopia Labour Force Survey 2005, Uganda National Household Survey 2010, Tanzania National Panel Survey 2010, Kenya Integrated Household Budget Survey 2005, Cameroon Household Survey 2007, Ghana Living Standards Survey 2005, Zambia Living Condition Monitoring Survey 2010, South Africa Labour Force Survey 2007.

A negative correlation between wage levels and wage dispersion seems to be coming from countries with smaller wage sectors, as shown in column (7).

A negative correlation between wage levels and wage dispersion seems to be at odds with a well-established stylized fact that economists have long associated with the Kuznets curve (see Figure 3). Namely, that there is a positive correlation between average income and inequality for countries at low levels of income (i.e. the SSA countries are mostly placed on the left part of the graph).

The last two columns of Table 1 present the magnitudes of income inequality for the same set of countries. We compute the standard deviation of log income using monthly household consumption data from the same household surveys (where that information was available). Measures of income inequality, whether Gini index in column (8) or the standard deviation of log income in column (9), seem to suggest a negative cross-country correlation between wage dispersion and income dispersion (the latter includes workers in both paid and self-employment). Below, we will show that high wage inequality and low income inequality can coexist when a large fraction of the workforce is engaged in low-income non-wage activities.

---

11 Since 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.
The evidence presented above suggests a general pattern of labor markets in least developed countries: the wage sector in poor economies, despite being very small in size, is characterized by very high levels of wage dispersion. Therefore, a model of the labor market that attempts to rationalize its poor performance needs to be consistent with a greater wage dispersion. It also suggests that understanding the performance of the wage sector, including wage dispersion, should be an integral part of the understanding of income inequality in these countries.

3 Model

In this section, we set up a simple model of heterogenous firms that generates a link between low levels of job creation, low mean wages, and high levels of residual wage dispersion that we observe in the data. Our theoretical contribution consists of using a frictional labor markets’ search and matching model (e.g. Mortensen and Pissarides [1994]) in a framework that captures the main features of standard development economic models, such as Harris and Todaro (1970) model of migration and Lewis (1954) two-sector model of structural change.

While firms are heterogeneous, workers are assumed to be homogeneous. We focus on the labor demand side as a driver of wage dispersion for two reasons. First, observed worker characteristics cannot explain fully the differences in wages as the data show residual dispersion to matter. Second, consider

Moreover, recent empirical evidence shows that changes of earnings inequality in many
differences in unobserved worker characteristics - such as quality of education or innate differences in ability - that might explain a higher wage dispersion in poorer countries. Since the number of jobs is very limited, more able workers are also more likely to be employed. 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 findings that a larger wage sector is associated with a lower wage dispersion.

3.1 Environment

This is a continuous time model of two labor markets - the wage sector and the self-employment sector, or home production. The wage sector is populated by heterogenous 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 labor as input.\(^{13}\)

There is a continuum of infinitely lived workers, with a mass normalized to one, that supply labor to firms. 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. Reflecting search frictions, the offer arrival rate and the vacancy filling rate 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 \( u \). As is standard in the literature, we assume a Cobb-Douglas form, i.e.

\[
M(v, u) = mv^\eta u^{1-\eta}, \quad 0 < \eta < 1, \tag{1}
\]

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 \( \lambda = \frac{M(v,u)}{u} = m\theta^\eta \), while the vacancy filling rate is

\(^{13}\) Given 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.
given by $q = \frac{M(v,u)}{v} = m\theta^{\eta-1}$.

Jobs are subject to the exogenous destruction shock that arrives at rate $\delta$. Competition and entry costs endogenously determine the number of firms in the market. 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 $r$.

Workers without a job end up in home production, or the self-employment sector. Unlike in industrialized countries, the unemployment rate in developing countries is very low or virtually non-existent; therefore, self-employment income is a more relevant outside option for workers. The aggregate production $Y_H$ in the home sector is assumed to be an increasing concave function of home sector labor, 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) and $0 < \gamma < 1$ is a returns to scale parameter.

The home production sector is assumed to be competitive so that wages in this sector are determined by the marginal product of labor. Then, a worker’s income in home production is equal to

$$w_H = \gamma A L_H^{\gamma-1}.$$ (2)

Given that $\gamma < 1$, this setup implies that a larger self-employment sector is associated with lower incomes. Moreover, all self-employed workers are assumed to be looking for a wage job and hence the number of jobseekers is equal to $u = L_H$ (we discuss the significance of this assumption in Appendix B.1).

### 3.2 Worker’s problem

Workers are either in self-employment and searching for a job or working. If the latter, the value of employment at a firm that pays wage $w$, $W(w)$, satisfies

---

14 We 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 labor product. While it leads to higher estimates of $A$ in our numerical exercises, none of the policy experiments or other results are affected.

15 For 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 labor quality. Lagakos and Waugh (2013), for example, propose a Roy model where a small non-agricultural sector implies a larger agriculture sector populated with relatively unproductive workers.

---
the following Bellman equation:

\[ rW(w) = w + \delta(U - W(w)), \quad (3) \]

where \( r \) is the common firms’ and workers’ discount rate and \( U \) is the value of search. The right-hand side of the equation is the sum of income flow from working, \( w \), and the expected capital loss if the job is destroyed and the worker becomes self-employed. The latter event happens at constant Poisson rate \( \delta \). The value of working can be re-written as

\[ W(w) = \frac{w + \delta U}{r + \delta}. \quad (4) \]

The job search is a costly process that involves direct search costs and time away from home production. Hence, we postulate that a self-employed worker obtains consumption flow \( w_H - z \) while looking for a job by means of home production less search costs \( z \), and she has an option of finding a job in the wage sector. The value of search, \( U \), then solves the following Bellman equation:

\[ rU = w_H(L_H) - z + \lambda(\theta) \int (\max\{W(w), U\} - U) dG(w), \quad (5) \]

where \( \lambda(\theta) \) is the job offer arrival rate that depends on market tightness \( \theta \), and \( G(w) \) is the cumulative distribution function of firms that pay wage \( w \).

Using equation (3), we can solve for the worker’s reservation wage \( w_R \) that equates the value of search with that of working:

\[ w_R = w_H(L_H) - z + \frac{\lambda(\theta)}{r + \delta} \int w_R (w - w_R) dG(w), \quad (6) \]

so that only wage offers above \( w_R \) are accepted.

### 3.3 Firm’s problem

Firms operate a CRS technology in labor and differ in their productivity level \( p \). The value of a job in a firm with productivity \( p \), \( J(p) \), solves the following Bellman equation:

\[ rJ(p) = p - w + \delta(V(p) - J(p)). \quad (7) \]

The first term on the right-hand side of equation (7) is the firm’s profit flow, \( p - w \). 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. The value of an open vacancy can be found as

\[
rV(p) = -c + q(\theta)(J(p) - V(p)),
\]

where \( c \) is a vacancy posting cost. The value of a vacant job, \( V(p) \), and a filled job, \( J(p) \), 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 as in [Hall and Milgrom (2008)](Hall and Milgrom). During a potential delay, the worker engages in home production and receives the flow value of \( w_H \), 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:

\[
w = \arg \max_w (p - w)^{1-\beta}(w - w_H)^\beta,
\]

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

\[
w(p) = \beta p + (1 - \beta)w_H,
\]

conditional on \( p > w_H \).

### 3.5 Labor market clearing

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

---

\[^{16}\text{The results are qualitatively similar, but mathematically cumbersome, with an alternative bargaining setup where the threat point of a worker is the value of search and that of a firm is the value of an open vacancy.}\]
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. \(w(\hat{p}) \geq w_R\).

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
\[rV(p) = -\frac{r + \delta}{r + \delta + q(\theta)} c + \frac{q(\theta)}{r + \delta + q(\theta)} (1 - \beta) \left(p - w_H(L_H(\theta))\right). \quad (12)\]

Below, we show that the wage in the home production sector is determined endogenously in equilibrium and it depends positively on \(\theta\).

It is useful to rewrite the free market condition from equation (11) as the expected gain relative to the outside option of exiting the market, \(V(\hat{p})\):
\[rk = \int_{\hat{p}} (rV(p) - rV(\hat{p})) d\Gamma(p) = \frac{q(\theta)}{r + \delta + q(\theta)} (1 - \beta) \int_{\hat{p}} (p - \hat{p}) d\Gamma(p), \quad (FE)\]
where we have used the fact that \(V(\hat{p})\) is equal to zero. It is convenient to define 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 \(\varphi''(\hat{p}) = \Gamma''(\hat{p}) > 0\). Therefore, the free entry condition implies a decreasing relationship between \(\hat{p}\) and \(\theta\).

The reservation productivity level \(\hat{p}\) is derived from setting \(V(\hat{p}) = 0\), i.e.
\[\frac{c}{q(\theta)} = (1 - \beta) \frac{\hat{p} - w_H(L_H(\theta))}{r + \delta}. \quad (13)\]
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 \((1 - \beta)\) share of the present discounted value of the match surplus (output flow \(p\) less home production \(w_H\)). Rearranging this equation,
we get the following zero profit condition

$$\hat{p} = w_H(L_H(\theta)) + \frac{c(r + \delta)}{q(\theta)(1 - \beta)}. \quad (ZP)$$

The link between $\hat{p}$ and $\theta$ is given by the vacancy filling rate $q$ and the wage in the home sector $w_H$, which depends on the size of the home sector $L_H$. 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,

$$\lambda(\theta) L_H = \delta L_F = \delta (1 - L_H) \Rightarrow L_H = \frac{\delta}{\delta + \lambda(\theta)}, \quad (14)$$

where $L_F$ is the mass of workers employed by firms. Note that the wage in the home sector is increasing in $\theta$, as the job finding rate $\lambda$ is increasing in $\theta$, and $w_H$ is decreasing in $L_H$.

In sum, we have two equations - the zero profit (ZP) condition and the free entry (FE) condition - and two unknowns: $\hat{p}$ and $\theta$. 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. labor demand) as opposed to the workers’ reservation wage, and (ii) all self-employed workers search for a job. In Appendix B.1, we discuss these assumptions in detail and show that they are not restrictive.

### 3.6 Comparative statics

There are a number of parameters that affect equilibrium variables and that can potentially generate the relations we observe in the data between the size of the wage sector, wage mean, and wage dispersion. Some of the key parameters are entry costs $k$, labor market efficiency $m$, home sector productivity $A$, workers’ bargaining power $\beta$, and ex-ante productivity distribution, $\Gamma(p)$.

First, consider an increase in the 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 4). As it becomes more difficult to enter the market, the number of firms falls and so does the vacancy-to-unemployment ratio.
Figure 4: The effect of an increase in entry costs $k$.

decrease in market tightness $\theta$ reduces the job finding rate $\lambda$ and, as a consequence, the size of the wage sector. Perhaps counterintuitively, 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 the threshold, $\hat{p}$, is now lower.

An increase in matching efficiency $m$ has an effect on both equilibrium conditions. On one hand, as the vacancy filling rate $q$ increases with $m$, the value of a vacancy rises. As a consequence, there is more entry for a given value of $k$ and the Free Entry condition shifts to the right. On the other hand, the increase in $q$ implies that the marginal firm can be less productive, thus shifting the Zero Profit condition downwards. A similar situation happens when workers’ bargaining power $\beta$ decreases. Intuitively, higher matching efficiency, $m$, and higher firms’ bargaining power, $1 - \beta$, have a positive effect on the value of a firm, which induces entry and increases market tightness $\theta$. The impact on productivity threshold $\hat{p}$ is less obvious from the graphs; however, below we show that also $\hat{p}$ unambiguously increases.

Now consider an increase in the self-employment sector productivity. Larger values of $A$ increase workers’ outside option and thus their wages. Hence, the marginal firm needs to be more productive, shifting the Zero Profit condition upwards. As a result, $\hat{p}$ increases and $\theta$ falls, leading to more workers moving to the home production sector. Finally, we inspect changes in the underlying productivity distribution. Consider, for example, a location shift, so that distribution $\Gamma_1$ has a higher mean than $\Gamma_2$, 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 $\Gamma_1$ than under $\Gamma_2$. 

16
The following lemma summarizes our results.

**Lemma 1** The productivity threshold \( \hat{p} \) and market tightness \( \theta \) are greater when one of the following is true, everything else equal: (i) entry cost \( k \) is lower; (ii) matching efficiency \( m \) is greater; (iii) workers’ bargaining power \( \beta \) is lower and (iv) a distribution \( \Gamma \) first-order stochastically dominates another (e.g. a location shift) or is a mean-preserving spread of another distribution. A greater level of self-employment productivity \( A \), however, is associated with a greater \( \hat{p} \) but a lower market tightness \( \theta \).

**Proof.** See Appendix B.2. ■

### 3.7 Wage employment, wages, and income dispersion

In this subsection, we explore to what extent differences in the parameters discussed above can generate our stylized facts, namely a negative relationship between wage inequality and mean wages, and between wage inequality and the size of the wage sector.

First, because of the direct link between \( \theta \) and the size of the wage sector, it is clear from Lemma 1 how the model can deliver different levels of labor allocations. Second, the results of Lemma 1 have a direct implication in terms of mean wages. Recall that wages are a linear combination of productivity and self-employment income, so that the average wage is equal to \( E(w) = \beta E(p | p \geq \hat{p}) + (1 - \beta)w_H \). Therefore, the parameter differences that raise the productivity threshold and market tightness simultaneously will increase mean wages through both channels: average productivity and a greater outside option.

The variance of wages can be written as \( \text{Var}(w) = \beta^2 \text{Var}(p | p \geq \hat{p}) \). [Heckman and Honore (1990)] show that the variance is decreasing in \( \hat{p} \) as shown in Figure 4 if the productivity distribution belongs to the family of log-concave density functions. It follows that if productivity is distributed log-concave, our model delivers the stylized facts, unequivocally, for differences in \( k \) and \( m \) and productivity location shifts. Differences in workers’ bargaining position (either due to the variation in self-employment productivity \( A \) or workers’ bargaining power \( \beta \)) cannot deliver these stylized facts as the size of the wage sector falls when mean wages increase, contrary to the data.

The impact of an increase in the underlying productivity dispersion on wage inequality is more ambiguous, as it has a direct positive effect on wage dispersion through a higher unconditional variance of productivity and an indirect

---

\[^{17}\text{Log-concave distributions include normal, exponential, logistic, gamma (for shape parameter greater or equal than 1), beta, extreme value, among others.} \]
negative effect through a higher truncation point \( \hat{p} \). Although the model does not allow for a full analytical characterization in this case, we can use a numerical simulation to show that differences in underlying productivity dispersion alone are not sufficient to generate the observed empirical regularities. The numerical simulation also illustrates complementarities between different variables. For example, high entry costs amplify the existing differences in the underlying productivity distributions across countries. Moreover, entry barriers are more detrimental for the wage sector in countries with a higher degree of labor market frictions. All these results are shown graphically in Figures 6 to 10 in Appendix B.3.

The following proposition summarizes how the model can deliver our main stylized facts.

**Proposition 1** Assume \( \Gamma(p) \) has a log-concave density. An economy will have a smaller dispersion in (log) wages, higher mean wages and a larger wage sector than another economy when, everything else equal, one of the following is true: (i) \( k \) is lower, (ii) \( m \) is greater, or (iii) the underlying productivity distribution exhibits a higher mean for a given variance.

**Proof.** See Appendix B.2.

The determination of the size of the wage sector and the variance of (log) wages is 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

\[
\text{Var}(\ln I) = L_H (1 - L_H)(\ln w_H - E \ln w)^2 + (1 - L_H) \text{Var}(\ln w),
\]

which depends on log wage dispersion, the percentage gap between the self-employment wage \( w_H \) and the average wage, and the shares of workers in the wage and home sectors. To understand the mechanism behind it, 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. The variance of income is zero in this case. The opposite case is when the frictions are so high that no firms enter the market and everyone is self-employed, receiving wage \( w_H \). Also in this case, there is no income dispersion. Departing from either extreme will result in an increase in income inequality, generating an inverse U-shape relationship between the levels of income and inequality. We show this formally in Appendix B.2.
4 Estimation Strategy

Our model relies on estimating transitions into and from wage employment and hence we need to observe the same individuals for at least two time periods. For this reason we restrict our analysis to five countries that fulfil these data requirements - Ghana, Nigeria, Uganda, Tanzania, and South Africa. Below we outline the estimation procedure, paying a particular attention to the measurement of entry costs in the data and the decreasing returns to scale parameter in self-employment.

The model is solved under the assumption that the economy is in steady state. We use an exponential distribution for underlying firm types \( \Gamma(p) \) with the mean and standard deviation of \( \sigma \), a parameter that we estimate.\(^{18}\) The full set of parameters then is the vector of 10 parameters \( (r, \delta, \eta, c, k, m, \sigma, \beta, A, \gamma) \)' that fully characterize the equilibrium of the model. We partition the parameter space into two groups. In the first group, parameters are set exogenously for the lack of necessary data or are estimated from other data sources outside the main optimization procedure. This group of variables includes the interest rate, \( r \), the elasticity of the matching function with respect to vacancies, \( \eta \), the returns to scale parameter in the home sector, \( \gamma \), and the job destruction rate, \( \delta \). The remaining vector of parameters is estimated by the indirect inference approach by minimizing a distance criterion between key moments obtained from the model and the data (see Gourieroux, Monfort and Renault, 1993).

4.1 Pre-determined parameters

We set the interest rate \( r \) at 1.25% (where a unit of time is a 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. Without data on vacancies, \( \eta \) cannot be identified separately. The elasticity of the matching function with respect to vacancies, \( \eta \), is usually estimated in the range of 0.3 – 0.5 (see Petrongolo and Pissarides, 2001). Here, we set it to 0.5, as is common in the literature.

*Returns to scale in the home production sector*

In order to obtain the returns to scale parameter \( \gamma \), we use the value added per worker in agriculture for the time period of 1990-2012. Measuring self-employment income in the data is challenging since it includes unreported pay-

\(^{18}\)We have tried several distributions in 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.
ments, payments in kind or working for a family farm without pay. Instead, we choose to recover it from the aggregate agricultural sector’s production, given that the majority of self-employed activities in developing countries are related to subsistence farming. In particular, we run the following regression across SSA countries:

\[ \ln Y_{Hjt} = b_0 + b_1 \ln L_{Hjt} + b_2 \ln T_{jt} + \epsilon_{jt}, \]  

(15)

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 \( \epsilon_{jt} \) is the error term. We also control for year and country fixed effects. Then, the parameter of interest can be recovered from \( \gamma = \hat{b}_1 \) and is equal to 0.24\(^{19}\). This value of \( \gamma \) 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).\(^{20}\) We assume that the returns to scale parameter \( \gamma \) is the same across the five countries we analyze, while the overall home sector productivity \( A \) is allowed to vary.

**Job destruction rates**

We construct transition rates between self-employment and formal employment using the panel structure of the datasets in Nigeria, Uganda, Tanzania, and South Africa. In addition, we construct transitions based on Ghanaian household survey using retrospective information on economic activity (see Appendix A for details). Table 2 shows yearly transition rates for five countries. The most striking observation is how low these rates are - even for South Africa, where the transition rates are much higher than in the other four countries, the job finding rate (out of self-employment and unemployment together) is about 4% a month, 10 times lower than in the US\(^{21}\). In the other four countries, the job finding rate is lower than 0.4% a month. This suggests that labor market inefficiencies are prevalent in this region and labor mobility is extremely

---

\(^{19}\)In addition, we run the same regression with the cereal yield in kilos per hectare as a dependent variable (since cereals are the main crop in these countries). This alternative specification results in \( \gamma = 0.15 \), which we then use to run a robustness check for our counterfactual experiments. Using \( \gamma = 0.15 \) increases our estimate of the home productivity parameter \( A \), while leaving other parameter values unchanged. The results of the policy experiments and the obtained elasticities of job creation, wages and income remain virtually the same.

\(^{20}\)As Lewis (1954, p. 402) states about the poorest regions: “The marginal product of labor is negligible, zero or even negative. (...) This phenomenon is not, however, by any means confined to the countryside (...) These occupations have a multiple of the number they need, each of them earning very small sums from occasional employment.”

\(^{21}\)Note 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 with an annual transition rate into wage employment of 48.8% and 41.5%, respectively.
Table 2: Average yearly transition rates between self-employment and the wage employment.

<table>
<thead>
<tr>
<th>Country</th>
<th>SE to E rate, $\lambda$</th>
<th>E to SE rate, $\delta$</th>
<th>Steady state SE share, $L_H$</th>
<th>Actual SE share</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tanzania</td>
<td>1.6%</td>
<td>13.0%</td>
<td>89.1%</td>
<td>91.4%</td>
</tr>
<tr>
<td>Uganda</td>
<td>4.4%</td>
<td>31.5%</td>
<td>87.7%</td>
<td>87.7%</td>
</tr>
<tr>
<td>Ghana</td>
<td>0.9%</td>
<td>4.1%</td>
<td>82.0%</td>
<td>86.0%</td>
</tr>
<tr>
<td>Nigeria</td>
<td>1.8%</td>
<td>37.2%</td>
<td>95.4%</td>
<td>93.6%</td>
</tr>
<tr>
<td>South Africa</td>
<td>45.7%</td>
<td>29.6%</td>
<td>39.3%</td>
<td>40.6%</td>
</tr>
</tbody>
</table>


Using the flows into and out of self-employment, we can compute a steady-state home production share as predicted by the model (based on equation (14)) and compare it to the actual share of workers in the self-employment sector. We find that the two numbers are aligned together very closely, suggesting that these economies are not too far from the steady state. Given the steady state assumption, one of three variables $\lambda$, $\delta$, and $L_H$ is redundant, as it can always be derived from the other two. We choose to use the empirical destruction rate $\delta$ and the self-employment share $L_H$ to inform our estimation.

4.2 Empirical targets

Using a set of the pre-defined parameters described above, we estimate the remaining vector of parameters $\vartheta = (c, k, m, A, \sigma, \beta)'$ by the simulated method of moments (see Appendix C.1 for details). We are matching the following moments: the mean log wage, the standard deviation of log wage residuals, the self-employment share, the labor share in value added, the entry costs per worker, and the average 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.\footnote{The 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...}
then run a robustness check with the hiring costs of three months of wages to check the sensitivity of our analysis and find virtually the same results. Below we provide more details on how we obtain labor shares and entry costs.

**Labor share**

We use the average labor share in the model to back out workers’ bargaining power, $\beta$. In particular,

$$E \left( \frac{w}{p} \right) = E \left( \frac{(1 - \beta)w_H(\theta) + \beta p}{p} \right) = (1 - \beta)E \left( \frac{w_H(\theta)}{p} \right) + \beta,$$

which depends on the equilibrium productivity cutoff $\hat{p}$ and market tightness $\theta$. To obtain the empirical counterpart of the labor 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 cover a broad range of topics including firm’s costs, employment, and performance measures.

First, we construct the value added series for each firm as the value of sales less the purchases of raw materials and intermediate goods, as well as the costs of fuel, electricity and telecommunication. We then compute the labor share at the firm level as the ratio of the labor costs to the value added and match the median labor share $w/p$ for each country.

The labor share ranges between 0.16 (in Tanzania) to 0.43 (in South Africa). Note that these values are likely to overestimate the true share of the production surplus paid to workers since (i) the labor share is derived from the firm’s total labor 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.

**Determination of entry costs**

Entry costs $k$ in the model can be interpreted broadly as a regulatory variable (e.g. government red tape), a borrowing constraint (e.g. the collateral required in order to get a credit), or access to advance technology. Credit constraints have been shown to be an important barrier to development.

Similarly, regulatory...
Table 3: Entry costs

<table>
<thead>
<tr>
<th>Country</th>
<th>Electricity connection cost, % GDP per capita</th>
<th>Average firm size, % GDP per capita</th>
<th>Cost per worker, % of GDP per capita</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tanzania</td>
<td>2,861</td>
<td>18.1</td>
<td>158</td>
</tr>
<tr>
<td>Uganda</td>
<td>7,022</td>
<td>14.7</td>
<td>477</td>
</tr>
<tr>
<td>Ghana</td>
<td>5,902</td>
<td>23.5</td>
<td>251</td>
</tr>
<tr>
<td>Nigeria</td>
<td>1,433</td>
<td>15.8</td>
<td>91</td>
</tr>
<tr>
<td>South Africa</td>
<td>875</td>
<td>48.9</td>
<td>18</td>
</tr>
</tbody>
</table>

The costs of electricity connection is based on the World Bank’s Doing Business survey in 2010. The average firm size is based on the Enterprise Survey data and refers to the number of permanent full-time workers.

barriers to firm entry have been associated with higher employment in non-wage activities.

In order to quantify the entry costs in the data, one alternative is to use legal costs of starting a business based on the World Bank’s Doing Business indicators. While we find that the legal fees are negatively correlated with the level of development as our model predicts, they might not be the best indicator for the firms’ entry costs. On one hand, the legal fees may overstate the actual costs as many of the enterprises in the wage sector in the SSA countries 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. Instead, we try to get a more tangible estimate of the entry costs in these countries by looking at major barriers to firms’ operations. In particular, we use the World Bank’s Enterprise Survey data that collects information on various factors that firms perceive to be major obstacles to operations, including poor electricity supply, access to finance, crime and disorder, corruption, land access, and informal sector competition among others. We find that the top three commonly named barriers to operations is electricity and power outages (e.g. more than 80% of firms in Ghana find it to be a major or a very severe obstacle), access to finance (67% ), and tax rates (31%). Therefore, we use the costs of getting electricity connection to proxy the entry costs in the model and we consider it to be a lower bound on the actual costs as we abstract from other costs (e.g. credit constraints) that we cannot measure well.

See Djankov et al. (2002) and Herrendorf and Teixeira (2011), among others.

According to World Bank’s Doing Business, a firm in Sub-Saharan Africa has to wait, on average, around 140 days to get a connection to the network and pay more than 4000% of the average income per capita. In an OECD country, a firm has to wait around 77 days and it costs around 70% of the average income. Furthermore, according to Enterprise Survey, a SSA firm can expect around 8 outages per month. To cope with costs, cuts and other shocks, under a half of the firms adopt a stand-alone power generator, implying that accessing the market involves a
Table 4: Empirical targets

<table>
<thead>
<tr>
<th>Country</th>
<th>$E(\ln w)$</th>
<th>$sd(\ln w)$</th>
<th>$L_H$</th>
<th>$E\left(\frac{w}{Y}\right)$</th>
<th>$k/Y$</th>
<th>Number of observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tanzania</td>
<td>4.14</td>
<td>0.87</td>
<td>0.91</td>
<td>0.16</td>
<td>19.0</td>
<td>1186</td>
</tr>
<tr>
<td>Uganda</td>
<td>4.45</td>
<td>0.79</td>
<td>0.88</td>
<td>0.28</td>
<td>76.6</td>
<td>1311</td>
</tr>
<tr>
<td>Ghana</td>
<td>4.79</td>
<td>0.70</td>
<td>0.86</td>
<td>0.39</td>
<td>30.1</td>
<td>1513</td>
</tr>
<tr>
<td>Nigeria</td>
<td>4.96</td>
<td>0.75</td>
<td>0.94</td>
<td>0.30</td>
<td>10.9</td>
<td>403</td>
</tr>
<tr>
<td>South Africa</td>
<td>6.02</td>
<td>0.58</td>
<td>0.41</td>
<td>0.43</td>
<td>2.2</td>
<td>21703</td>
</tr>
</tbody>
</table>

Note: The costs of electricity connection per worker (see Table 3) is converted into monthly income. Mean log wages are expressed in constant 2005 international dollars. The standard deviation $sd(\ln w)$ refers to log wage residuals. The average labor share $\frac{w_H}{Y}$ is obtained as a median ratio of labor costs to value added based on the Enterprise Survey data, World Bank. The number of observations is the number of private sector wage employees drawn from the labor force surveys. The sixth target (not included in the table) is the average hiring costs, $\xi$, that is chosen to be equal to one month of wages. The empirical moments are calculated based on the following datasets: Uganda National Household Survey 2010-2011, Tanzania National Panel Survey 2010, Ghana Living Standards Survey 2005, Nigeria General Household (Post-Planting and Post-Harvest) Survey 2010-2011, South Africa labor Force Survey March and September 2007.

In the data, the entry costs are recorded as a percentage of the economy’s annual GDP per capita. First, we transform the costs in terms of monthly income as a period in the model is one month. The model’s analog of GDP is monthly output, expressed as total production in the wage sector and the home sector less vacancy posting costs. Secondly, the entry barriers are 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 3. With the exception of Tanzania, the electricity connection costs are monotonically decreasing with the level of economic development.

Table 4 summarizes the empirical moments that we are matching in our estimation. While it is not possible to associate individual parameters with individual moments as they are determined together in the model, we provide some intuition for identification. The variance of log wages is determined primarily by the variance of underlying productivity $\sigma$. However, a higher variance implies a higher mean wage (due to a greater productivity threshold $\hat{p}$) in contrast to what we observe in the data. Effectively, we use the worker’s outside option $w_H$ (by adjusting $A$) to reduce the average wage, while keeping the variance the same. The matching efficiency is then used to fit the self-employment share.
The worker’s bargaining power is related to the labor share in the model. Finally, we back out the entry costs, $k$, from electricity connection costs, and the vacancy cost parameter, $c$, from the hiring costs.

5 Results

5.1 Estimated parameters

Table 5 shows the estimated parameters and their standard errors.

A natural way to think about differences in wages and productivity across countries is that they stem primarily from differences in underlying productivity. This, however, is not supported by our findings. Despite there being a substantial variation in the estimated values of $\sigma$ across countries, the observed differences in mean wages and wage inequality cannot be explained solely by differences in the underlying productivity distributions. For example, countries that appear to be similar in terms of their ex-ante productivity distribution, such as Tanzania and Nigeria, look very different in terms of their ex-post productivity and wage distributions due to the presence of frictions in the market. This result is in line with other studies that show that misallocation of resources due to frictions lowers aggregate productivity and growth.

While the relationship between underlying firm productivity and mean wages (or GDP per capita) in our sample is not monotone, the estimated values of home sector productivity, $A$, are increasing in the level of development. In order to check whether the model generates realistic values of self-employment income, we compare them with the poverty data for each country (see Table 10 in Appendix C.2). For example, our estimates for Tanzania generate self-employment income of less than $10 a month, which is consistent with the fact that close to 70 percent of the country’s population live below $1.25 a day.

Given low transition rates between self-employment and wage employment in the data, we find very low estimates of the matching efficiency parameter for all five countries. Our parameter values imply that South Africa has eight, six, and five times more efficient labor market than Tanzania, Ghana, and Nigeria, respectively. For comparison, a recent study by Sahin, Song, Topa and Violante

---

27 Note that the standard errors here understate statistical uncertainty as they do not account for the fact that $\gamma$ and $\delta$ are pre-estimated.

28 See, for example, Hsieh and Klenow (2009). Bartelsman, Haltiwanger and Scarpetta (2013a) 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 producing in each market.
Table 5: Estimated parameters in the model

<table>
<thead>
<tr>
<th>Country</th>
<th>σ</th>
<th>A</th>
<th>m</th>
<th>β</th>
<th>k</th>
<th>c</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tanzania</td>
<td>713.3</td>
<td>37.9</td>
<td>0.0008</td>
<td>0.112</td>
<td>1350.6</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>(25.108)</td>
<td>(3.163)</td>
<td>(4.2 × 10^{-5})</td>
<td>(0.003)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Uganda</td>
<td>477.1</td>
<td>65.0</td>
<td>0.0050</td>
<td>0.205</td>
<td>4357.1</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>(13.691)</td>
<td>(4.189)</td>
<td>(2.1 × 10^{-4})</td>
<td>(0.004)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Ghana</td>
<td>446.0</td>
<td>124.3</td>
<td>0.0011</td>
<td>0.269</td>
<td>2901.2</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>(10.613)</td>
<td>(6.631)</td>
<td>(4.2 × 10^{-5})</td>
<td>(0.005)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Nigeria</td>
<td>717.5</td>
<td>129.7</td>
<td>0.0014</td>
<td>0.215</td>
<td>839.5</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>(38.684)</td>
<td>(15.952)</td>
<td>(1.4 × 10^{-4})</td>
<td>(0.010)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>South Africa</td>
<td>1227.9</td>
<td>324.5</td>
<td>0.0066</td>
<td>0.270</td>
<td>1969.1</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>(7.469)</td>
<td>(3.220)</td>
<td>(4.6 × 10^{-5})</td>
<td>(0.001)</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

Note: Asymptotic standard errors are given in parentheses. Given that there is no sampling variation for the entry costs nor the hiring costs in the data, the standard errors for k and c, cannot be obtained.

(2014) estimates the aggregate matching efficiency parameter in the US to be 0.94, while Albrecht, Robayo-Abril and Vroman (2015) use 0.25 for Colombia. This evidence suggests the presence of high mobility costs in SSA that leads to labor misallocation across sectors, regions and productive opportunities in general.29

The estimates of workers’ bargaining power parameter β range from 0.11 in Tanzania 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 unionization30 and relatively low levels of workers’ human capital31.

5.2 Productivity and wage gains

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 and lowering wage inequality. 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

29It has been shown in the literature that labor market frictions lead to misallocation of resources, lower firm productivity and a fall in output (see for example Lagos, 2006, Restuccia and Rogerson, 2013, Hsieh and Klenow, 2009, Hsieh and Klenow, 2010 and Vollrath, 2014).

30The Global Wage Report 2010/2011 (ILO) shows that the share of unionized workers as a fraction of the workforce is 1.1 percent in Uganda, 2.2 percent in Tanzania, 14 percent in Ghana and 29 percent in South Africa.

31Cahuc, Postel-Vinay and Robin (2006) estimate bargaining power parameters for different groups of workers by their skill levels using French matched employer-employee data. They find that manual workers have a very low (close to zero) bargaining power parameter, while high-skilled workers had a much higher parameter.
Table 6: Ex-ante and ex-post distributions of firm productivity and wages.

<table>
<thead>
<tr>
<th>Country</th>
<th>Average productivity</th>
<th>Average wage</th>
<th>St.dev. of ln w</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ex-ante</td>
<td>ex-post</td>
<td>%Δ</td>
</tr>
<tr>
<td>Tanzania</td>
<td>713</td>
<td>726</td>
<td>1.8</td>
</tr>
<tr>
<td>Uganda</td>
<td>477</td>
<td>501</td>
<td>5.0</td>
</tr>
<tr>
<td>Ghana</td>
<td>446</td>
<td>484</td>
<td>8.5</td>
</tr>
<tr>
<td>Nigeria</td>
<td>718</td>
<td>762</td>
<td>6.2</td>
</tr>
<tr>
<td>South Africa</td>
<td>1228</td>
<td>1409</td>
<td>14.7</td>
</tr>
</tbody>
</table>

Note: For ex-ante wage distributions we assume that in the limit when the degree of frictions is the highest, almost all of the workforce is employed in home production, i.e. \( L_H \rightarrow 1 \), and the outside option is close to \( w_H = \gamma A \). adopt new technologies and replace less productive older firms and that restricting entry leads to lower average productivity and sluggish growth. Using the estimated model parameters, we can compare ex-ante and ex-post productivity and wage distributions to quantify the gains that arise due firm entry and competition. Table 6 presents the percentage difference in average productivity, wages and wage dispersion between each country’s ex-ante and ex-post distributions.

Again, our results indicate that frictions play a large role in shaping labor market outcomes. Average productivities and wage gains under current labor market conditions are very modest in the poorest four countries in our sample, ranging from about 2 percent in Tanzania to 9 percent in Ghana. With the market frictions being significantly lower in South Africa, it is not surprising to find a 15 percent increase in average productivity and a 27 percent gain in average wages. Similarly, wage inequality gets reduced by more than a quarter in South Africa, compared to only 4 percent in Tanzania.

5.3 Outcome elasticities

In this subsection, we examine empirical relevance of various determinants of job creation, wages and inequality for each country. We focus on four channels: the entry costs, \( k \), the labor market efficiency, \( m \), self-employment productivity, \( A \), and underlying productivity dispersion, \( \sigma \). We simulate changes in these variables to illustrate what happens to the size of wage sector, wage levels and dispersion, as well as the overall income. The results are expressed in terms of elasticities and are presented in Table 7.

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 wage dispersion. Productivity in self-employment activities is generally lower in poorer
Table 7: Elasticities of outcome variables with respect to changes in parameters.

<table>
<thead>
<tr>
<th></th>
<th>Elasticities with respect to</th>
<th>Wage employment</th>
<th>Average wage</th>
<th>St. dev of log wages</th>
<th>Average income</th>
<th>St. dev of log income</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tanzania</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k$</td>
<td>-0.93</td>
<td>-0.01</td>
<td>0.02</td>
<td>-0.42</td>
<td>-0.41</td>
<td></td>
</tr>
<tr>
<td>$m$</td>
<td>1.83</td>
<td>0.01</td>
<td>-0.03</td>
<td>0.84</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td>$A$</td>
<td>-0.01</td>
<td>0.11</td>
<td>-0.21</td>
<td>0.58</td>
<td>-0.38</td>
<td></td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.95</td>
<td>0.90</td>
<td>0.20</td>
<td>0.85</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td>Uganda</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k$</td>
<td>-1.01</td>
<td>-0.03</td>
<td>0.05</td>
<td>-0.48</td>
<td>-0.41</td>
<td></td>
</tr>
<tr>
<td>$m$</td>
<td>1.75</td>
<td>0.03</td>
<td>-0.05</td>
<td>0.84</td>
<td>0.70</td>
<td></td>
</tr>
<tr>
<td>$A$</td>
<td>-0.04</td>
<td>0.15</td>
<td>-0.24</td>
<td>0.57</td>
<td>-0.42</td>
<td></td>
</tr>
<tr>
<td>$\sigma$</td>
<td>1.06</td>
<td>0.87</td>
<td>0.21</td>
<td>0.93</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>Ghana</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k$</td>
<td>-0.95</td>
<td>-0.03</td>
<td>0.05</td>
<td>-0.40</td>
<td>-0.37</td>
<td></td>
</tr>
<tr>
<td>$m$</td>
<td>1.71</td>
<td>0.05</td>
<td>-0.06</td>
<td>0.72</td>
<td>0.65</td>
<td></td>
</tr>
<tr>
<td>$A$</td>
<td>-0.07</td>
<td>0.21</td>
<td>-0.30</td>
<td>0.63</td>
<td>-0.48</td>
<td></td>
</tr>
<tr>
<td>$\sigma$</td>
<td>1.04</td>
<td>0.81</td>
<td>0.26</td>
<td>0.78</td>
<td>0.85</td>
<td></td>
</tr>
<tr>
<td>Nigeria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k$</td>
<td>-0.93</td>
<td>-0.02</td>
<td>0.03</td>
<td>-0.26</td>
<td>-0.43</td>
<td></td>
</tr>
<tr>
<td>$m$</td>
<td>1.87</td>
<td>0.02</td>
<td>-0.03</td>
<td>0.51</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>$A$</td>
<td>-0.04</td>
<td>0.17</td>
<td>-0.26</td>
<td>0.75</td>
<td>-0.45</td>
<td></td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.99</td>
<td>0.84</td>
<td>0.25</td>
<td>0.51</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td>South Africa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k$</td>
<td>-0.39</td>
<td>-0.15</td>
<td>0.18</td>
<td>-0.42</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>$m$</td>
<td>0.76</td>
<td>0.27</td>
<td>-0.32</td>
<td>0.80</td>
<td>-0.33</td>
<td></td>
</tr>
<tr>
<td>$A$</td>
<td>-0.05</td>
<td>0.30</td>
<td>-0.35</td>
<td>0.38</td>
<td>-0.46</td>
<td></td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.44</td>
<td>0.84</td>
<td>0.19</td>
<td>1.03</td>
<td>0.29</td>
<td></td>
</tr>
</tbody>
</table>

countries and exogenous positive shocks in that sector have been used to explain structural change (see for example Lewis, 1954 and Matsuyama, 1992). We find that an increase in home productivity $A$, while increasing the wage level and decreasing the wage dispersion, has a negative impact on the size of the wage sector. Finally, consider variation in the underlying productivity dispersion that might be driven by changes in capital intensity, technology adoption, or opening to trade. An increase in $\sigma$ in the model leads to a rise in the size of the wage sector, average wages and average income. At the same time, it generates higher wage and income inequality, with the latter effect being especially large for poorer countries.

While many of our results confirm the existing wisdom in development literature, below we focus on what we consider to be new insights that come directly from our use of a unifying modeling framework.

---

32 This is more in line with 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.
1. **A reduction in labor market frictions has the largest impact on wage employment.** A one percent rise in the matching efficiency parameter leads to a 0.8 percent increase in the size of the wage sector in South Africa and about a 1.7-1.9 percent increase in the other four countries. A reduction in the entry costs is about half as effective.

2. **Complementarities between $m$ and $k$.** The effect of an increase in labor market efficiency $m$ on mean wages and wage dispersion is larger when the existing entry costs $k$ are low. Similarly, a reduction in entry barriers is more prominent when matching efficiency is high. This can be seen when comparing the effects of frictions on mean wages and wage dispersion in South Africa to the remaining four countries in our sample. (These interactions are also shown in the model simulations in Appendix B.3.) For example, a one percent increase in the matching efficiency leads to a 0.27 percent increase in average wages in South Africa and only to a 0.01 percent increase in Tanzania. Therefore, improving both variables at the same time amplifies their effects on labor market outcomes.

3. **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 (above 85 percent) are self-employed in the poorer countries in our sample, this leads to an increase in income inequality in Ghana, Nigeria, Uganda, and Tanzania, while the opposite is true for South Africa.

4. **A substantial impact of self-employment productivity on mean wages, income, and inequality.** 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 (0.38 percent increase) and income dispersion (0.46 percent drop). In addition, it also has a large effect on wages: for example, a one percent increase in $A$ causes about a 0.1-0.3 percent rise in the mean wage and about a 0.2-0.35 percent drop in the standard deviation of log wages. Although higher self-employment productivity has a negative impact on job creation, these effects are close to zero, especially for the poorest countries in our sample.
6 Policy discussion and counterfactual experiments

Wage employment in developing countries has been identified by international organizations as key in generating economic growth and reducing poverty. For example, the first of the United Nations’ eight Millennium Development goals is to eradicate extreme poverty and hunger and it includes “achieving full and productive employment and decent work for all, including women and young people”. We can use our model to find out how a given country can achieve its development goals, be it the size of the wage sector or the overall income, through a number of policies.

An obvious candidate policy is improving firms’ productivity through, for example, international trade or FDI inflows. We can represent it in our model through an increase in the underlying productivity dispersion parameter, $\sigma$. 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.33

In what follows, however, we want to focus on three alternative strategies that might be considered as potential substitutes for technological improvement: (i) a reduction in entry barriers, (ii) an increase in labor market efficiency, and (iii) an increase in home sector productivity.

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.34 We propose that a good way of looking at entry

---

33Interestingly, 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 the home and wage sector by the same amount, while raising the level of wages and income, will fail to induce the reallocation of workers from self-employment to the wage sector.

34For example, Bartelsman, Haltiwanger and Scarpetta (2013b) talk about a combination of factors, including entrepreneur’s effort and administrative fees; Aw, Chung and Roberts (2003) simply mentions regulatory and technological differences in entry costs in two countries (Taiwan and South Korea) and Ulysse (2010) refers to technological determinants of entry for firms in Brazil. Direct microevidence 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.
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, 2015 and Allcott, Collard-Wexler and O’Connell, 2016). There is also evidence that, in the presence of shortages, some firms in developing countries cope with electricity shocks by buying captive power generators (Reinikka and Svensson, 2002 and Rud, 2012b). As a consequence, infrastructure policies reducing the cost of access or improving its quality seem to be a good proxy for a reduction in $k$.

Similarly, understanding policies that increase efficiency in labor markets has not been easy, in particular in developing countries. A thorough analysis of job creation in the developing world concludes that “there is no consensus on what the content of labor policies should be” (World Bank, 2013). A review of active labor 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. The evidence in favor of active labor market policies is stronger in developed countries. However, these type of interventions may fall short in developing countries. For example, Groh, McKenzie, Shammout and Vishwanath (2014) show that addressing search and matching frictions for both employers and job-seekers may fail to improve hiring rates. More promising results are shown in Franklin (2015) and Abebe et al. (2016). The authors used a series of randomized experiments aimed at improving search effectiveness in spatially dislocated areas of Ethiopia. For example, Franklin (2015) 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. This type of innovative labor market policies aimed at reducing search frictions for workers or firms seem to be the most promising way of strengthening labor markets in least developed countries, as poor infrastructure and costs of search are high enough to fragment labor markets.

Card, Kluve and Weber (2010) provide a literature review on the microeconometric evaluation of active labor 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 favorable impacts, especially in the short run. Moreover, a follow-up study by Card, Kluve and Weber (2015) show that the effectiveness of active labor 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 providing information about workers’ abilities increases employment outcomes.
Finally, improvements in the productivity of self-employed 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. These policies can be interpreted as an increase in $A$ in our model.

In the analysis below, we run two policy experiments. In the first subsection, we use the model to find out how a given country can achieve its development goals by reducing frictions or increasing self-employment productivity. In the second subsection, we show which alternatives are more effective in terms of the cost-benefit analysis given a fixed policy budget. Note that in both of these experiments we do not consider transition dynamics after the changes are introduced, instead they should be thought of as a comparison between two steady state economies - the baseline versus a counterfactual economy with alternative parameter values.

6.1 Reaching development goals

Here, we focus on two development goals - the size of the wage sector and the average income. For illustration purposes, we choose South Africa’s empirical moments as targets as we think of it as a reasonable benchmark for the least developed countries in our sample. We then simulate the model and compute what values of a given variable $k$, $m$ or $A$ are required to get to these targets.

Table 8 presents the baseline parameter values and the new values that are required if Ghana were to achieve a given target by using one policy instrument at a time. (Table 11 in Appendix C.2 show the results for other countries, including for two additional goals - the mean and standard deviation of log wages). For example, in order to achieve the same size of the wage sector as in South Africa, Ghana needs to decrease its entry costs nine times or, alternatively, increase its labor market efficiency three times. An increase in self-employment productivity cannot be used as a policy instrument for the first target as it reduces the size of wage employment. A much larger change is required if the
policy aim is the average level of income: Ghana’s home sector productivity has to increase more than 8 times, the labor market needs to become roughly 10 times more efficient, or, alternatively, the entry costs have to be reduced 75 times.

Table 8 shows the magnitude of the required change in each policy when only one parameter is allowed to vary at a time. However, as discussed in the previous section, there exist complementarities between these policies and hence it might be easier to reach the desired target by using multiple changes at the same time. For example, if matching efficiency \( m \) is kept constant then \( k \) needs to be lowered from 2,901 to 321 international dollars to reach South Africa’s level of wage employment, hence the total gap in \( k \) is 2,580 international dollars. However, if \( m \) is increased at the same time the required change in \( k \) will be smaller.

Figure 5 illustrates these complementarities for Ghana by plotting all possible combinations of the changes in the entry costs \( k \) and the matching efficiency \( m \) (or home sector productivity \( A \)) that lead to the same outcome. The axis show the required changes in the parameter values expressed as a percentage of the total gap that needs to be filled to achieve the desired target. The green dotted line is the line of perfect substitutability between two policies where one percent of the gap in one parameter can be traded off for one percent of the gap in another. Our simulations show that desirable combinations of \( k \) and \( m \) (or \( A \)) policies lie below the dotted line, suggesting that using two instruments at the same time is more efficient. These graphs also show that there are more complementarities between \( k \) and \( A \) than between \( k \) and \( m \). For example, Ghana can catch up with South Africa’s average income level by fully eliminating the gap in the entry costs or, alternatively, by reducing the gap in \( k \) by half and simultaneously reducing the gap in \( A \) only by 12 percent.

### 6.2 Fixed policy budget

While Table 7 gives the estimates of the benefits of each policy, it is silent on its costs. We can use our model directly to estimate the costs of two policies -
subsidies given to firms upon the entry and income subsidies given to the self-employed workers - and compare their effectiveness for every dollar spent. We consider a fixed policy budget that can be spent on either of these two policies. As an example, we consider the amount of foreign aid that the United States has sent to Ghana, which totalled $215 million in 2015. After converting it into 2005 international dollars and dividing by working age population in Ghana, it amounts to $B = $1.38 per person per month. We use our parameter estimates for Ghana to examine the effect of both policies on aggregate income and choose the one with the highest return. Again, we assume that the change is permanent and the subsidy is being paid from now on forever, or equivalently, we analyze Ghana’s economy in steady state with new parameter values.

Table 9 shows the estimated change in average income for each policy. Given the total foreign aid budget, if the government were to share this amount equally across all individuals, the average income would increase by $1.38 per person, or by 2.74 percent, in our baseline model. Alternatively, the government can pay an income subsidy to the self-employed, a policy that is equivalent to an increase in home productivity $A$. We re-run the model and find the corresponding value of $A$ such that the total amount of the subsidy paid to the self-employed workers is equal to the total budget, i.e.

$$w_1^H L_1^H - w_0^H L_0^H = B,$$

where superscript 1 denotes the economy under a new policy, while 0 denotes the baseline model. As a result of the subsidy, workers’ outside option increases, followed by a rise in wages, and so does the size of the home pro-

---

36 Total population in Ghana in 2015 is 27.4 million and 57.8 percent of individuals belong to the 15-64 age group.
duction sector, $L_H$. This has to be taken into account when computing the total subsidy budget. We find that the subsidy is equal to $1.59 per self-employed, which is equivalent to 4.8 percent increase in $A$. Average income increases by 3 percent, or by $1.11 for every dollar of the foreign aid budget.

In the next step, we re-estimate the model with an entry costs subsidy. Since the entry costs are fixed one-time costs, to make them comparable to income subsidies we need to express them in terms of monthly flow values, $rk$. Again, we re-run the model to find the corresponding value of $k$ such that the total amount of the subsidy paid to all the firms is equal to the total foreign aid budget, i.e.

$$\frac{rk^1(1 - u^1 + v^1)}{1 - \Gamma(\hat{p}^1)} - \frac{rk^0(1 - u^0 + v^0)}{1 - \Gamma(\hat{p}^0)} = B,$$

where the total mass of existing firms if equal to the sum of the filled jobs, $1 - u$, and unfilled vacancies, $v$, and $1 - \Gamma(\hat{p})$ is the fraction of firms that survived after drawing their productivity $p$. The resulting policy is equivalent to a 13 percent reduction in entry costs $k$, which leads to an 6 percent increase in average income. Here, every dollar of the entry costs subsidy generates a much higher return: with a $2.3 increase in income for $1 spent.

What is the economic intuition behind this result? Even though a one percent increase in $A$ has a larger impact on income than a one percent reduction in $k$, as shown in Table 7, it is more expensive for poor countries to increase $A$ through self-employment income than to subsidize firm entry. The relative size of self-employment as opposed to the firm sector is key. To confirm it, we repeat these experiments for South Africa using the same (as a proportion of its average income) budget amount. Given the self-employment rate in South Africa is only about 40 percent, the same amount of foreign aid budget generates more than 3 times higher effect on home sector productivity than in Ghana - 15.2 percent versus 4.8 percent, respectively, which leads to a much higher benefits-to-costs ratio (2.11 versus 1.11). The opposite is true for entry subsidies: the corresponding reduction in the entry costs is smaller in South Africa than in Ghana and so is the resulting bang-to-buck ratio.
Table 9: Policy experiments: Bang for buck

<table>
<thead>
<tr>
<th>Policy</th>
<th>Percentage change in average income</th>
<th>$ increase in income for every $ spent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ghana</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uniform income subsidy for each person</td>
<td>–</td>
<td>2.74</td>
</tr>
<tr>
<td>Income subsidy for self-employed</td>
<td>4.78</td>
<td>3.03</td>
</tr>
<tr>
<td>Entry subsidy for firms</td>
<td>–</td>
<td>-13.33</td>
</tr>
<tr>
<td><strong>South Africa</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uniform income subsidy for each person</td>
<td>–</td>
<td>2.74</td>
</tr>
<tr>
<td>Income subsidy for self-employed</td>
<td>15.2</td>
<td>5.77</td>
</tr>
<tr>
<td>Entry subsidy for firms</td>
<td>–</td>
<td>-7.06</td>
</tr>
</tbody>
</table>

7 Conclusion

Labor 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 labor force end up in less desirable and even less productive self-employment occupations (e.g. subsistence farming) or helping family activities for no pay.

In this paper, 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 propose a unifying framework that endogenously generates the link between the size of the wage sector, mean productivity and wages, and wage inequality. In particular, we incorporate channels identified by both the development and the labor 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,
ferences in workers’ bargaining power and outside options (e.g. subsistence level farming), and labor market inefficiencies - that can interact to generate these outcomes.

We subsequently estimate the model using micro data for Ghana, Nigeria, South Africa, Tanzania, and Uganda. We then run counterfactual experiments aimed at illustrating the main mechanisms affecting wage inequality across countries, as well as their interactions. The numerical exercise shows that the variation in the entry costs and labor market frictions can qualitatively and quantitatively reproduce our main stylized facts. Differences in the underlying productivity dispersion, on the other hand, are not sufficient to explain differences in wage distributions across countries. Our estimation results also reveal that there are significant complementarities between policy variables: for example, the effect of a change in labor 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 labor 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 analyze 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.

References


—, *Raising Productivity in Africa’s Modern Wage Enterprises to Foster Job Growth for Youth*, The World Bank, 2014/06/04


Appendix

Appendix A: Data


The sample is restricted to individuals aged between 15 and 65. We exclude public sector (government administration, state enterprises and parastatals, NGOs, and diplomatic missions) employees.

(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 and the number of periods worked is not reported, 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 2005=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 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: Uganda National Household Survey 2010-2011, Tanzania National Panel Survey 2008 and 2010, Nigeria General Household (Post-Planting and Post-Harvest) Survey 2010-2011, and South Africa Labour Force Survey March and September 2007. These datasets have a panel structure. In addition, we construct transitions using Ghanaian household survey information on economic activity in the past 7 days and contrast it to the respondents’ main activity in the last 12 months. We define flows from the wage sector to the home sector as those workers who are self-employed in the past 7 days and whose main occupation in last 12 months was wage employment, and the opposite for the flow from the home sector to the wage sector. This method is likely to underestimate the transitions since the survey has information only for the main occupation of a worker during the last 12 months and hence might miss workers that had more than one occupation during this period. For the other countries in our sample we expect the transitions to be estimated more precisely.

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. In some countries the time period between survey dates was different (e.g. 6 months in South Africa versus 2 years in Tanzania), hence all transition rates were converted into yearly rates. Moreover, the transition rates were weighted by the inverse of the length of the time interval between two interviews if they differed within one dataset.
Appendix B: Model

B.1 Equilibrium conditions

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. \( w(\hat{p}) \geq w_R \). Using the equation for workers’ reservation wage, we can derive their reservation productivity as follows

\[
p_R = \frac{w_H(L_H(\theta)) - z \beta}{r + \delta} + \frac{\lambda(\theta)}{r + \delta} \int_{p_R} (p - p_R) \frac{d\Gamma(p)}{1 - \Gamma(\hat{p})}
\]

\[
\leq \frac{w_H(L_H(\theta)) - z \beta}{r + \delta} + \frac{\lambda(\theta)}{r + \delta} \int_{\hat{p}} (p - \hat{p}) \frac{d\Gamma(p)}{1 - \Gamma(\hat{p})}
\]

\[
= \frac{\beta}{1 - \beta} \left( \frac{r k}{1 - \Gamma(\hat{p})} \left( \theta + \frac{\lambda(\theta)}{r + \delta} \right) - \frac{c(r + \delta)}{q(\theta)(1 - \beta)} \right) \frac{r k}{1 - \Gamma(\hat{p})},
\]

where the first step uses the fact that the surplus function \( \phi(\cdot) \) is a decreasing function so that \( \int_{p_R} (p - p_R) d\Gamma(p) \geq \int_{\hat{p}} (p - \hat{p}) d\Gamma(p) \) for \( p_R \leq \hat{p} \), and while the second step uses equations (ZP) and (FE) to substitute for \( w_H \) and the surplus function. For the reservation productivity \( p_R \) to be less or equal than \( \hat{p} \), the searching costs \( z \) need to satisfy the following condition:

\[
z \geq \frac{\beta}{1 - \beta} \left( \frac{r k}{1 - \Gamma(\hat{p})} \left( \theta + \frac{\lambda(\theta)}{r + \delta} \right) - \frac{c(r + \delta)}{q(\theta)(1 - \beta)} \right).
\] (16)

The second condition that has to be satisfied in equilibrium is the participation constraint. That is, the value of search in equilibrium, \( U = \frac{w_R}{r} \), has to be higher than the value of dropping out of the labor market and producing at home (i.e. getting self-employment income forever), which is equal to \( \frac{w_H}{r} \). 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 on average jobs last. Hence, in order for workers to be willing to participate in the labor market the reservation wage \( w_R \) has to be at least as high as home production income \( w_H \), which implies that the searching costs need to satisfy the following inequality:

\[
z \leq \frac{\lambda(\theta) \beta}{r + \delta} \int_{\hat{p}} (p - w_H(L_H(\theta))) \frac{d\Gamma(p)}{1 - \Gamma(\hat{p})}
\]

\[
= \frac{\beta}{1 - \beta} \left( \frac{r k}{1 - \Gamma(\hat{p})} \left( \theta + \frac{\lambda(\theta)}{r + \delta} \right) + c\theta \right),
\] (17)
where we used equation (ZP) to substitute for $w_H$ and equation (FE) to substitute for the surplus function. Combining the two conditions, we get the following interval for $z$

$$
\frac{\beta}{1 - \beta} \left( \frac{rk}{1 - \Gamma(\hat{p})} \left( \theta + \frac{\lambda(\theta)}{r + \delta} \right) + c \theta \right) \geq z \geq \frac{\beta}{1 - \beta} \left( \frac{rk}{1 - \Gamma(\hat{p})} \left( \theta + \frac{\lambda(\theta)}{r + \delta} \right) - \frac{c(r + \delta)}{\rho(\theta)} \right),
$$

(18)

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

(a) Discussion of Condition 1: Reservation wage

Consider the case when the reservation wage is higher than the minimum wage in the market, $w(\hat{p})$. In that case, not all offers are accepted by workers and the value of a job for a firm with productivity $p < p_R$ becomes negative since no workers will accept that job. Then, the lowest wage and thus the degree of wage dispersion in the market will be determined by the labor supply (as opposed to demand) side.

The importance of the reservation wage in determining how much wage dispersion a simple search framework can produce has been recently argued by Hornstein, Krusell and Violante (2011). They show that it is not necessarily true that search models can generate any amount of wage differentials as long as the wage-offer distribution is sufficiently dispersed. To see this, they propose a new measure of frictional wage dispersion: the mean-min (Mm) wage ratio - the ratio of the average accepted wage to the lowest accepted, or the reservation, wage. In a simple search model, it can be derived as

$$
Mm = \frac{\lambda + 1}{\lambda + 2} + \rho
$$

(19)

where the notations for $\lambda$, $r$, and $\delta$ are the same as above, and $\rho$ is the replacement rate, equal to the ratio of consumption flow of non-employed job-seekers to the mean wage. They show that the Mm ratio is independent of the wage offer or productivity distribution but depends on labor-market flows and preferences instead. The intuition is as follows: as the variance of the wage offer distribution increases, thus raising the mean-min ratio, so does the option value of getting a higher wage while searching that translates into a higher reservation wage, which in turn tends to decrease the Mm ratio.

Hornstein et al. (2011) find that observed magnitudes for worker flows in the US imply a very small frictional wage dispersion in a basic search model. In particular, as an empirical equivalent of the mean-min ratio they use 50-10 wage percentile ratios, which range between 1.7 and 1.9 in the recent years. Given
the monthly job finding rate of 43%, the job separation rate of 3%, the monthly interest rate of 0.41% and the replacement rate of 0.4, they find the $M_m$ ratio in the model to be around 1.05. Therefore, Hornstein et al. (2011) argue that to match the observed magnitudes of wage dispersion within a simple search model, the replacement rate needs to be very low or even negative, implying an unreasonably high disutility from non-market time.

A potential solution to this problem is either to extend the model to allow for on-the-job search or to assume that most of wage dispersion in the data is driven by unobserved heterogeneity of workers. Both of these suggestions are difficult to implement empirically, as these model extensions put extra demands on data that are hard to satisfy in developing countries. Moreover, the second option shifts the attention to the labor supply side of the market, making entry costs and other firm factors irrelevant. We choose not to follow any of these paths, but instead show that in our data the critique outlined in Hornstein et al. (2011) does not apply. The reason for that is that in the SSA countries the labor mobility flows are much lower than in the US and a simple search model can produce enough of wage dispersion.

To see that this is, in fact, a general observation, we can re-write the mean-min ratio in terms of the share of unemployed or self-employed workers. Using equation (14) and assuming that $r \approx 0$, we get

$$M_m \approx \frac{1-u}{u+1} \rho. \quad (20)$$

This equation implies that there is a negative relationship between the share of wage employment in the labor market and the mean-min ratio. Firstly, Table 1 shows that this relationship holds qualitatively in our data, using 50-10 wage percentile ratios. Secondly, also in terms of the magnitudes the $M_m$ ratio implied by the model is much closer to the data. In particular, given that the self-employment share of workers is above 80-90% in the SSA region, the $M_m$ ratio is about 2 if we assume the same replacement rate of 0.4 as in the US. It is conceivable, however, that the replacement rate $\rho$ in least developed countries is much lower in the absence of a well-established social security system and in the presence of high information frictions. Moreover, the interest rates are on average higher than in developed countries; hence, the exact $M_m$ ratio numbers estimated from equation (19) would be even higher. In Ghana, for example, using the yearly job finding rate of about 1%, the job separation rate of 4.1%, the replacement rate of 0.2, and the interest rate of 15%, the model delivers the

[37]In Ghana, the household survey has information on self-employment income. The replace-
Mm ratio of 4.2, which is greater than the 2.75 found in the data.

Intuitively, low labor mobility and low levels of self-employment income in least developed countries mean that the reservation wage is not binding in the labor market, that is, workers accept virtually all offers. Therefore, wage dispersion in the labor market is determined primarily by the demand side. Relying on this empirical evidence, we find it both convenient and justified to assume that the searching costs $z$ are such that the lowest wages in the market are determined by productivity threshold $\hat{p}$. An important implication of this is that, as it is apparent both in the simulation and estimation of the model, a standard search model applied to least developed economies can generate enough wage dispersion.

Also note that in our policy experiments we assume that the searching costs are such that the reservation wage is not binding. For example, consider an increase in matching efficiency, $m$, that raises the productivity threshold, $\hat{p}$, and market tightness, $\theta$. It means that the searching costs, $z$, will have to increase as well for inequality (16) to hold. This assumption is not very restrictive due to two reasons. First, we can think of search costs as being proportionate to home production income (e.g. due to lost output when workers actively look for a job instead of producing at home) and hence we can expect an increase in $z$ associated with a rise in $w_H$. Second, even if $z$ were not to increase and the reservation wage became binding, the wage dispersion would decrease even further as a result of a rise in $m$ (or a fall in $k$). Hence, we consider our results to give the lower bound of the effect of policy variables on wage levels and dispersion.

(b) Discussion of Condition 2: Participation constraint

Now suppose that inequality (17) does not hold so that the participation constraint binds, which might happen if $z$ is too high. In this case, some workers will prefer to quit searching and to produce at home instead. As the number of workers searching in the market falls, 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. Note that it ignores the searching costs $z$, so in reality it might be even lower.

38Based on our parameter estimates, we find that inequality (16) 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 $w_H$, resulting in a negative consumption flow while searching. Although this is an unrealistic assumption, we choose to make it nevertheless to keep the modeling framework exactly the same between all countries in our analysis. If one were interested in South Africa only, a more advanced model is needed, which would allow for either workers’ heterogeneity or on-the-job search and that would account separately for the large informal sector to fit the data (see, for example, Meghir et al. [2015] and Ulyssea [2010]).
production income, so that workers are indifferent between searching in the market or not.

Let \( \alpha \in (0, 1) \) to be a fraction of all self-employed workers who search in the market. Then market tightness is \( \theta = \frac{\delta}{\delta + \alpha \lambda H} \) and the steady state self-employment is determined by \( L_H = \frac{\delta}{\delta + \alpha \lambda H} \). Now, in addition to equilibrium equations (ZP) and (FE), we have one additional condition \( w_H = w_R \) and one additional unknown \( \alpha \). We can express the participation constraint as

\[
z = \frac{\lambda(\theta)}{r + \delta} \left( \beta \int_p (p - \hat{p}) \frac{d\Gamma(p)}{1 - \Gamma(\hat{p})} + \frac{c(r + \delta)}{q(\theta)(1 - \beta)} \right), \quad \text{(PC)}
\]

which is an increasing schedule of \( \hat{p} \) and \( \theta \), since \( \frac{\partial}{\partial \hat{p}} \mathbb{E}(p|p > \hat{p}) < 1 \) for log-concave density functions. Then the intersection of two curves - equations (FE) and (PC) - comprises an equilibrium.

Given that 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, for simplicity of exposition we choose to impose the condition that the searching costs are such that all workers in the home sector engage in active job search.

**B.2 Proofs**

**Proof of Lemma 1**

From the graphical analysis above, we showed that a greater \( k \) reduces \( \hat{p} \) and \( \theta \) and that an increase in \( A \) also raises \( \hat{p} \) but reduces \( \theta \). An increase in \( m \) unambiguously leads to higher equilibrium market tightness. In order to see what happens to productivity threshold \( \hat{p} \), we combine equations (FE) and (ZP) to obtain \( \hat{p} = w_H(L_H(\theta)) - \frac{c}{r + \delta} + \frac{c\varphi(\hat{p})}{rk} \). Differentiating \( \hat{p} \) with respect to \( m \), we get:

\[
\frac{\partial \hat{p}}{\partial m} = \left( 1 - \frac{c\varphi'(\hat{p})}{rk} \right)^{-1} \frac{\partial w_H}{\partial L_H} \left( \frac{\partial L_H}{\partial m} + \frac{\partial L_H}{\partial \theta} \frac{\partial \theta}{\partial m} \right).
\]

We know that in equilibrium \( \frac{\partial \theta}{\partial m} > 0 \). Then, given \( \frac{\partial w_H}{\partial L_H} < 0 \) due to decreasing returns to scale in the home production sector, the negative derivatives of \( L_H \) with respect to both \( \theta \) and \( m \), and \( \varphi'(\hat{p}) = \Gamma(\hat{p}) - 1 < 0 \), we can conclude that \( \frac{\partial \hat{p}}{\partial m} > 0 \). Therefore, higher matching efficiency leads to an increase in the reservation productivity threshold. Similarly, we can differentiate \( \hat{p} \) with respect to \( \beta \) to get:

\[
\frac{\partial \hat{p}}{\partial \beta} = \left( 1 - \frac{c\varphi'(\hat{p})}{rk} \right)^{-1} \left[ \frac{\partial w_H}{\partial L_H} \frac{\partial L_H}{\partial \theta} \frac{\partial \theta}{\partial \beta} - \frac{c}{(1 - \beta)^2} \right].
\]
The term in the square brackets is negative (as $\theta$ decreases with $\beta$). It follows that an increase in worker’s bargaining power leads to a lower productivity threshold value $\hat{p}$.

Suppose that underlying productivity distribution $\Gamma_1$ first order stochastically dominates $\Gamma_2$, that is, $\Gamma_1 \leq \Gamma_2$. Then, $q_1(\hat{p}) = \int_{\hat{p}} (1 - \Gamma_1(p))dp \geq \int_{\hat{p}} (1 - \Gamma_2(p))dp = q_2(\hat{p})$. Hence, the free entry curve under $\Gamma_1$ distribution lies above the FE curve under $\Gamma_2$, which leads to higher equilibrium values of $\theta$ and $\hat{p}$. Instead, suppose that $\Gamma_1$ is a mean preserving spread of $\Gamma_2$, i.e. $\int_{\hat{p}} \Gamma_1(p)dp \geq \int_{\hat{p}} \Gamma_2(p)dp$ for all $\hat{p}$ and $\int p\Gamma_1(p) = \int p\Gamma_2(p)$. We can show that $q_1(\hat{p}) - q_2(\hat{p}) = \int p\Gamma_1(p) - \int p\Gamma_2(p) + \int_{\hat{p}} (\Gamma_1(p) - \Gamma_2(p))dp$. Given the same mean, $q_1(\hat{p}) > q_2(\hat{p})$, which implies that the FE curve given $\Gamma_1$ lies above its counterpart given $\Gamma_2$, resulting in higher equilibrium values of market tightness and productivity threshold.

**Proof of Proposition 1**

Proposition 1 in [Heckman and Honoré (1990)](1990) shows that for log-concave distributions, $\frac{\partial \text{Var}(p \mid p \geq \hat{p})}{\partial \hat{p}} \leq 0$. A simple Taylor expansion gives us the variance of log wages as $\text{Var}(\ln(w)) \approx \frac{\hat{p}^2}{\hat{w}_H^2} \text{Var}(p \mid p \geq \hat{p})$. As stated in Lemma 1, lower entry costs $k$, higher labor market efficiency $m$, or a higher mean of ex ante productivity (for the same variance) increases equilibrium productivity threshold $\hat{p}$ and market tightness $\theta$, and as a result, raises self-employment income $w_H$. Hence, the variance of log wages is now lower due to a fall in the variance of ex-post productivity and an increase in $w_H$.

**Variance of (log) income**

Using Taylor approximation of $\ln w$ around $\ln w_H$ to linearize log wages, we get:

$$\text{Var}(\ln I) \approx (1 - L_H) \frac{\hat{p}^2}{\hat{w}_H^2} \left( L_H(E(\hat{p}) - w_H)^2 + \text{Var}(\hat{p}) \right),$$

where $E(\hat{p})$ is the mean and $\text{Var}(\hat{p})$ is the variance of the truncated distribution of $p$, conditional on $p \geq \hat{p}$. We can then show what happens to the variance of log income as the entry costs increase in an economy with a negligible home production sector. That is,

$$\lim_{L_H \to 0} \frac{\partial \text{Var}(\ln I)}{\partial k} \approx \frac{\beta^2}{\hat{w}_H^2} \left( \frac{\partial L_H}{\partial k} (E(\hat{p}) - w_H)^2 + \frac{\partial \text{Var}(\hat{p})}{\partial k} - \text{Var}(\hat{p}) \frac{\partial L_H}{\partial k} \right) \left( 1 + \frac{2w_H'}{w_H} \right) > 0.$$ 

This derivative is positive as we have shown that $\frac{\partial \text{Var}(\hat{p})}{\partial k} > 0$ and $\frac{\partial L_H}{\partial k} > 0$ and, given the Cobb-Douglas functional form assumption with decreasing returns to scale, $\lim_{L_H \to 0} \frac{w'_H}{w_H} \to -\infty$. Hence, as the entry costs $k$ rise in this economy,
income inequality increases because (i) the variance of wages is higher due to a lower productivity threshold, and (ii) some workers move to the home sector.

Similarly, we can examine the other extreme, where virtually everyone in the labor market is working in the home sector. That is,

\[
\lim_{L_H \to 1} \frac{\partial \text{Var}(\ln I)}{\partial k} \approx -\beta^2 \frac{\partial L_H}{\partial k} \left( \left( E(\bar{p}) - w_H \right)^2 + \text{Var}(\bar{p}) \right) < 0,
\]

which means that a decrease in the entry costs \( k \) actually increases income inequality. Hence, a country which starts with a very low level of wage employment and high entry costs might experience an increase in income inequality (at least initially) if it were to decrease \( k \). Once a sufficient number of workers relocated to the wage sector, the variance of income would start falling.

### B.3 Model simulation

We use numerical simulation to analyze the role of a number of channels in explaining a small size of the wage sector, low average wages and greater wage dispersion in Sub-Saharan Africa. We simulate the model using the baseline parameters that are obtained from the data on Ghana and that are described in detail in Sections 4 and 5. For each set of parameters, we solve for equilibrium market tightness and productivity threshold given an underlying productivity distribution.

(a) Entry costs

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 size of the wage sector in the economy and put downward pressure on wages. On top of their effect on average wages, entry barriers lead to a rise in wage inequality driven by a higher survival rate of less productive firms.

The left panel of Figure 6 illustrates how an increase in the entry cost parameter \( k \) leads to a fall in average log wages (solid blue line) and a rise in wage inequality (dashed green line). More prohibitive entry barriers decrease job creation in the wage sector, as can be seen on the right panel Figure 6 in a solid blue line. Even though wage inequality is continuously increasing, the effect on income inequality is non-linear, following an inverse U-shape on the same graph in a dashed green line. That is, as the number of workers in wage employment keeps falling while the number of self-employed workers keeps growing, the effect on income inequality is eventually reversed.
Figure 6: The effect of changes in entry costs $k$ on wages, wage employment and income inequality

(b) Labor market frictions

A reduction in labor market frictions, captured by an increase in matching efficiency, encourages job creation and leads to a rise in wage employment. The top left panel of Figure 6 shows that an increase in matching efficiency parameter $m$ causes a reduction in wage inequality. That is, a more efficient labor market has more compressed wages. An interesting feature of this graph is the interaction effect with the entry costs that amplifies the effect of matching frictions on wage inequality: an increase in $m$ (or, equivalently, a reduction in search frictions) is more effective in reducing wage inequality in a country with low entry costs $k$ as opposed to a country with high $k$.

Similarly to changes in $k$, also changes in the degree of labor market inefficiencies can produce a non-linear relationship between the mean and the variance of log income (see the top right panel of Figure 6). As we reduce labor market frictions, wages are increasing in both the wage sector and the home production sector (due to a smaller size of self-employment), thus increasing the overall income level. Although wage dispersion is falling, more people switch from self-employment (characterized by a unique income) to the wage sector, leading to a rise in income inequality. Eventually, a sufficient number of people move to wage employment, which together with a continuously falling (log) wage variance, has a negative effect on income inequality. Again, there is evidence of interactions between policy variables as the initial increase in income inequality is more pronounced for higher values of entry costs $k$.

By changing matching efficiency, the model can reproduce a negative relationship between the average wage and wage inequality observed in the data. The bottom left panel of Figure 6 shows that an increase in the entry costs parameter $k$ does not change the relationship between the mean and the variance
of log wages - the curve stays roughly the same, affecting only the range of the values for the mean and the variance of log wages. The share of wage employment, however, is very responsive to changes in matching efficiency (the bottom right panel of Figure 7) and thus we can use the relationship between (log) wage dispersion and the size of the wage sector to identify the effect of labor market frictions on wage dispersion.

(c) Underlying productivity distribution

The third channel through which differences in wage dispersion can exist across countries is differences in underlying productivity that might be driven by lower capital intensity, inferior technology, or poor infrastructure. Here, we analyze separately the role of having an underlying productivity distribution with a higher mean or a higher variance.

Figure 8 illustrates the effect of changing the location parameter \( \mu \) and the scale parameter \( \sigma \). First, we find that a country with a higher average underlying productivity will also have a lower wage dispersion (see the left panel of Figure 8).

\[ \text{We use a logistic productivity distribution, because it allows us to analyze separately the effects of changes in the mean from changes in the variance of underlying productivity. Moreover, a logistic distribution belongs to the family of log-concave densities; hence, the results of our Proposition 1 are applicable here.}\]
Figure 8: The effect of changes in underlying productivity on wage dispersion

Figure 9: The effect of changes in underlying productivity dispersion on wage dispersion

the graph); while a higher underlying productivity dispersion will be translated into a higher (log) wage dispersion (see the right panel of the graph). Hence, if industrialized countries have on average better technology or if they adapt new technology faster (e.g. getting rid of obsolete technology is likely to reduce the variance of underlying productivity in a similar way as a reduction in the entry costs reduces variance of ex-post productivity in our model) they will exhibit lower wage inequality. Moreover, entry costs seem to exacerbate the initial differences in productivity between developing and industrialized countries.

Although changes in the variance of underlying productivity generate different wage dispersions across countries, we can rule them out as a key driving factor for observed differences in wage inequality. Figure 9 shows that changes in \( \sigma \) alone are incapable of matching the negative mean-variance relationship in log wages found in the data.

\((d)\) Workers’ outside option and bargaining power
Finally, as the fourth channel through which differences in wage dispersion can exist across countries, we consider the workers’ side of the model. First, we can think of the workers’ bargaining power parameter $\beta$ as stemming from collective bargaining (influenced by unionization in the labor market) or the individual bargaining position that is determined by the workers’ skills endowment. In both cases, we expect a country’s economic development to result in higher values of $\beta$, either through a greater degree of unionization or an investment in human capital. The second factor that determines a worker’s bargaining position is her outside option $w_H$, which is affected by productivity in home production, $A$, that may vary across countries. Both of these variables increase the wage cost to the firm, which discourages entry into the market and reduces job creation (see the top panel of Figure 10).

Our results dispute the hypothesis that the observed differences in wage distributions can be explained solely by differences in worker’s bargaining position. First, while we expect unionization to increase mean wages, and the model delivers that, it also reduces job creation. Interestingly, Magruder (2012) shows that unionization in South Africa prevents job creation also among small firms, as centralized bargaining agreements by large firms extend to non-unionized smaller firms.
cile with our stylized fact where countries like South Africa have larger wage employment than presumably less unionized countries. Second, the fact that unionization compresses wages in not supported in our model. As can be seen on the bottom left panel of Figure 10, a higher value of $\beta$ is associated with a higher variance of wages (through a lower value of $\hat{\rho}$ and a larger weight put on productivity as opposed to the outside option). Finally, although higher self-employment productivity $A$ reduces wage inequality (see bottom right panel of Figure 10), it fails to generate a negative relationship between the size of the wage sector and the variance of wages that we see in the data.

Appendix C: Estimation and results

C.1 Simulated method of moments

Denote by $\psi$ a vector of moments in the data that represent the auxiliary model. Its counterpart from the simulated data can be written as $\hat{\psi}(\vartheta)$. Given a vector of structural parameters, we solve for the steady state equilibrium and generate the corresponding moments. The estimator is the choice of the structural parameters that minimizes the weighted distance between the data moments and the simulated moments:

$$\hat{\vartheta} = \arg\min_{\vartheta} (\hat{\psi}(\vartheta) - \psi)' \Omega (\hat{\psi}(\vartheta) - \psi),$$

where $\Omega$ is a positive definite weighting matrix. Since the model is exactly identified, we can use an identity matrix as a weighting matrix. The variance-covariance matrix for parameter estimates is given by:

$$\hat{V} = (D'SD)^{-1},$$

where $D$ is the derivative of the vector of moments with respect to the parameter vector, i.e. $D = \frac{\partial \psi(\vartheta)}{\partial \vartheta}$, and $S$ is the inverse of the optimal weighting matrix. We calculate the derivatives numerically and we use the sample covariance matrix of the moments to approximate the optimal weighting matrix. Asymptotic standard errors of the parameters can then be calculated by

$$ASE(\hat{\vartheta}_j) = \sqrt{\frac{1}{N} \hat{V}_{jj}},$$

where $jj$ is the $j$-th diagonal element. We use the number of private sector wage employees for $N$, which is a more conservative number than the total sample of individuals of the working age. Note that for the entry costs and the hiring
Table 10: Self-employment income

<table>
<thead>
<tr>
<th>Country</th>
<th>Estimated SE income, $w_H$</th>
<th>Share of population below $1.25$ a day</th>
<th>Share of population below $2$ a day</th>
<th>Marginal product in agriculture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tanzania</td>
<td>9.7</td>
<td>67.9</td>
<td>87.9</td>
<td>14.6</td>
</tr>
<tr>
<td>Uganda</td>
<td>17.2</td>
<td>37.9</td>
<td>64.7</td>
<td>10.6</td>
</tr>
<tr>
<td>Ghana</td>
<td>33.5</td>
<td>28.6</td>
<td>51.8</td>
<td>31.5</td>
</tr>
<tr>
<td>Nigeria</td>
<td>32.7</td>
<td>62.0</td>
<td>82.2</td>
<td>139.1</td>
</tr>
<tr>
<td>South Africa</td>
<td>154.5</td>
<td>16.7</td>
<td>35.2</td>
<td>133.0</td>
</tr>
</tbody>
</table>

Note: The estimated self-employment income $w_H$ is imputed from the model, given the estimated parameters. Source: Share of population below poverty line is obtained from the World Bank World Development Indicators Database. The marginal product in agriculture is derived from regression (15).

costs we do not observe sampling variation in the data (that is, the entry costs as a percentage of output and the hiring costs are used to pin down $k$ and $c$, given other structural parameters). Therefore, the standard errors can only be obtained for $m$, $A$, $\sigma$, and $\beta$ estimates.

C.2 Results: Additional tables

Table 10 presents monthly home production income implied by the model, expressed in international 2005 dollars. Columns 2 and 3 show the fraction of a country’s population that live below the poverty line of $1.25$ and $2$ a day. In addition, we compare our estimates of $w_H$ to the marginal product of labor in agriculture based on the value added in agriculture.

Table 11 presents the values of a given policy variable $k$, $m$, or $A$ that are required for each country to get to South Africa’s level of employment, mean wages and wage inequality, as well as overall income.
Table 11: Policy experiments: Targeting South Africa’s values

<table>
<thead>
<tr>
<th></th>
<th>Parameter values required to fit the following targets</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline</td>
<td>Wage sector size</td>
<td>St. dev of log wages</td>
<td>Mean income</td>
<td>Mean wage</td>
</tr>
<tr>
<td>Tanzania</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>k</td>
<td>1350.6</td>
<td>84.0</td>
<td>31.5</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>m</td>
<td>0.0008</td>
<td>0.0032</td>
<td>0.0065</td>
<td>0.0235</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>37.9</td>
<td>–</td>
<td>150.6</td>
<td>1075.2</td>
</tr>
<tr>
<td>Uganda</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>k</td>
<td>4357.1</td>
<td>420.1</td>
<td>347.7</td>
<td>17.8</td>
</tr>
<tr>
<td></td>
<td>m</td>
<td>0.0050</td>
<td>0.0165</td>
<td>0.0250</td>
<td>0.0858</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>65.0</td>
<td>–</td>
<td>177.0</td>
<td>1063.6</td>
</tr>
<tr>
<td>Ghana</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>k</td>
<td>2901.2</td>
<td>321.4</td>
<td>439.2</td>
<td>38.7</td>
</tr>
<tr>
<td></td>
<td>m</td>
<td>0.0011</td>
<td>0.0034</td>
<td>0.0033</td>
<td>0.0104</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>124.3</td>
<td>–</td>
<td>215.7</td>
<td>1044.7</td>
</tr>
<tr>
<td>Nigeria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>k</td>
<td>839.5</td>
<td>28.5</td>
<td>61.0</td>
<td>6.4</td>
</tr>
<tr>
<td></td>
<td>m</td>
<td>0.0014</td>
<td>0.0064</td>
<td>0.0077</td>
<td>0.0161</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>129.7</td>
<td>–</td>
<td>289.6</td>
<td>1082.5</td>
</tr>
</tbody>
</table>