Wage Dispersion, Job Creation and Development:

Evidence from Sub-Saharan Africa

Preliminary *

Juan Pablo Rud †       Ija Trapeznikova‡

July 23, 2016

Abstract

Labor markets in least developed countries are characterized by small wage sectors and low productivity and wages. Using household level data for a number of countries in Sub-Saharan Africa we explore the distribution of wages and document a negative correlation between (i) wage dispersion and size of the wage sector and (ii) wage dispersion and wage mean. The latter observation is in stark contrast with the positive correlation between income mean and income inequality found for the same countries. We propose a labor search and matching framework with entry costs and firm heterogeneity that delivers endogenously the regularities observed in the data. We also show that this model can reconcile the differences between wage and income inequality by accounting for labor reallocations between wage and self-employment sectors. We focus on a number of channels to explain these phenomena in Sub-Saharan Africa and find, both theoretically and empirically, that policies addressing entry costs and labor market frictions have the potential to improve the most labor market outcomes, such as job creation, wages and wage dispersion.

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

JEL: J21, J31, J64, O11, O15

---

*We thank Jesper Bagger, John Kennan, Philipp Kircher, Michael Koelle, Fernando Martin, Santiago Oliveros, Elias Papaioannou, Victor Rios-Rull, Mark Roberts, Rob Shimer, Jon Temple, Gianluca Violante, Yves Zenou and seminar participants at CSAE Conference, RHUL, Sandbjerg, Edinburgh, Alicante, Essex, Universidad Di Tella, Sussex, St. Gallen, St. Louis Fed, IFS-EdePo Conference, 2nd African Search and Matching Workshop, Novafrica, Federal Reserve Board, and SED Conference for useful comments and suggestions. We are grateful to Taye Mengistae at World Bank for helping us procure the data.

†Royal Holloway, University of London

‡Royal Holloway, University of London
1 Introduction

Labor markets in developed and developing countries are strikingly different. First and foremost, the size of the wage sector among the poorest regions in the world is very small, falling below 15% in many Sub-Saharan African countries. Furthermore, the wage sector is characterized by relatively low levels of productivity, and hence low earnings. Albeit being very small and relatively unproductive, the wage sector still outperforms non-wage occupations and has been identified as desirable both by workers looking for a steady source of income and as an engine of economic growth. As there are not enough jobs for the number of people willing to work for a wage, most end up in less desirable self-employment occupations, own-account work or helping family activities for no pay.

A key question is: what is preventing labor markets to adjust? In other words, why are jobs not being created through lower wages, as competitive labor market models would predict? The development economics literature has established, both theoretically and empirically, a number of factors that constrain labor demand in developing countries, including frictions in labor markets (e.g. segmentation or search costs), entry barriers (e.g. credit constraints, regulations or poor infrastructure), or low aggregate productivity. The aim of this paper is to propose a unifying framework that combines these development channels in one model. We then estimate the model using micro data from a number of countries in Sub-Saharan Africa to examine the empirical relevance of various constraints for job creation and wage growth and quantify their relative importance.

We start our analysis by providing a new empirical evidence that seems to be intrinsically linked to labor market underperformance in poor economies. In particular, we use household level data to show that the wage sector exhibits very high levels of wage dispersion in Sub-Saharan Africa. For example, we find that the top 10% of wage earners in Uganda and Ethiopia are paid 15 and 20 times more than the bottom 10%, respectively, while in the US this ratio is only 8 times. Furthermore, we document that wage inequality is greater in countries with smaller...
wage sectors and lower average wages. This empirical relationship also provides a new insight into a well-established fact that economists have long associated with the Kuznets curve, namely that for the poorest countries (mostly in Sub-Saharan Africa) there is a positive cross-country correlation between average income and inequality. That is, high wage inequality and low income inequality can coexist in a country, when a large fraction of the workforce is engaged in low-income non-wage activities. We aim to propose a model that accounts for this set of stylized facts and delivers wage sector size, levels, and dispersion of wages endogenously.

We develop a model that incorporates features of traditional and modern models of development (e.g. two sectors as in [Harris and Todaro, 1970; Lewis, 1954; and Banerjee and Newman, 1993]) into a frictional search framework (Mortensen and Pissarides, 1994). In particular, there are two labor markets in the economy: a wage sector and a home production (or self-employment) sector. The former operates as a frictional labor market while the latter is the outside option, with a unique income. Firms are heterogeneous in their productivity level, revealed only after they paid an entry cost to operate in the wage sector. There are three important elements in the model that are standard and that endogenously generate the link between the size of the wage sector and the wage distribution. First, a wage bargaining process links wages to firms’ productivity and the worker’s outside option (i.e. income in self-employment). Second, a free entry condition guarantees that firms enter until the expected value of an open vacancy equals the entry cost. Finally, a zero profit condition truncates the ex post distribution of productivity as firms with productivity below a threshold exit the market immediately after entry. The model delivers two key equilibrium variables: the economy’s market tightness (a measure of competition in labor markets, defined as the ratio of vacancies to job seekers) and the productivity threshold (the minimum productivity level at which a firm can survive in the wage sector). Because the market tightness is positively correlated with job creation, it directly determines the size of the wage sector. At the same time, the productivity threshold determines the productivity distribution of active firms and through that affects mean wages and variance.

We proceed to estimate the model for five countries (Ghana, Nigeria, South Africa, Tanzania and Uganda), using panel household survey data that includes information on employment status of workers, their wages, and demographic characteristics. The purpose of the numerical exercise is to illustrate the main mechanisms that could potentially explain a small size of the wage sector, low average wages and a greater wage dispersion in Sub-Saharan Africa. Namely, we focus on the level of entry costs that includes various barriers to entry (e.g. regulations, fi-
financial or other constraints to starting a business, infrastructure costs), differences in countries’ underlying productivity distribution, labor market frictions (captured by a measure of matching efficiency), home sector productivity, and workers’ bargaining power (e.g. unions). First, we show that only some of these channels can qualitatively and quantitatively reproduce our main stylized facts. For example, we find 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. Secondly, our estimation points out to some interesting interactions that cannot be recovered analytically from the model. For example, the effect of labor market frictions on wage inequality is amplified in the presence of higher barriers to entry.

The advantage of a unifying model is that it provides a valuable analytical tool to inform debates about policy effectiveness and their complementarities. It can also be used to assess general equilibrium effects of a large number of randomized field experiments in development economics that focus on labor market efficiency. To that end, we use our estimated parameters to run a series of counterfactual policy experiments and analyze three types of interventions: a reduction in the entry costs, an increase in labor market efficiency, and an increase in self-employment productivity. We show that a reduction in labor market frictions has the largest impact on job creation: one percent increase in labor market efficiency leads to 0.8 percent increase in the size of the wage sector in South Africa and 1.7 to 1.9 percent increase in the other four countries in our sample; while the reduction in entry costs is only half as effective. Similarly, we find a substantial impact of self-employment productivity on average wages, income and inequality for poorer countries: one percent increase in home sector productivity brings about 0.6-0.75 percent increase in the overall income and about 0.2 percent rise in mean wages.

Our counterfactual experiments show two main results. First, for countries to catch up with South Africa, huge policy efforts are needed. 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 ten 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

---

5See, 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, Thuysebaert and Udry (2015) and Blattman, Fiala and Martinez (2013)).

6This is consistent with recent evidence of field experiments in Ethiopia where active labor market policies seeking to improve search intensity and efficacy improves job finding rates and the quality of matches, see Franklin (2015) and Abebe et al. (2016). Koelle and Quinn (2016) also highlight the importance of search costs in Ghana.
9 times, the labor market needs to become roughly 10 times more efficient, or, alternatively, the entry costs have to be reduced 80 times. Second, the results show substantial policy complementarities, which demonstrating the benefits of using an integrated model.

Our framework builds on recent progress in the development and labor economics literature. The development economics literature has established a number of factors that shape labor market outcomes in developing countries including the size of wage employment and the level of wages\footnote{See footnote 4.} However, 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. The importance of productivity and wage dispersion in industrialized economies has already been recognized and recent advances in labor economics (driven primarily by the availability of large comprehensive datasets and the use of rich structural models) have deepened our understanding of its determinants\footnote{See Lentz and Mortensen (2010) and Rogerson, Shimer and Wright (2005) for a survey of the literature, see Bon-temps, Robin and van den Berg (2000) for France, Jolivet, Postel-Vinay and Robin (2006) for a set of European countries and the U.S.}. Yet, productivity and wage dispersion remain under-explored in developing countries\footnote{There are several examples in the literature that focus primarily on the size and composition of the informal sector, i.e. non-registered wage workers (see for example Amaral and Quintin, 2006, Meghir, Narita and Robin, 2015, Satchi and Temple, 2009, Ulyssea, 2010 and Zenou, 2008). Note that the wage sector in Africa, albeit small, also includes informal salaried work. Only for middle income countries such as Brazil the distinction between formality and informality becomes relevant, as informal wage employment accounts for around 20% of the labor force.}. 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. Subsequently, 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 shows some counterfactual experiments. In Section 6 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 Figure 1, we show the share of wage employment increases as countries get richer and find that in many Sub-Saharan African countries it falls below 20%. However, it does not discourage labor force participation, which is greater in Sub-Saharan Africa (around 70%), than in OECD countries (around 60%).

Figure 1: Wage employment and development

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 of low levels of job creation and labor demand, 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

---

\[^{10}\text{For example, Ethiopia, Ghana, Tanzania and Uganda among others (ILO 2012). Note that this measure falls even more for African countries if public sector jobs are excluded, as they constitute a relatively large share of wage employment.}\]

\[^{11}\text{Note that the wage sector, albeit small, also includes informal salaried work. Only for middle income countries such as Brazil the distinction between formality and informality becomes relevant, as informal wage employment accounts for around 20% of the labor force. See Meghir et al. (2015) and Ulyssea (2010).}\]
Figure 2: Manufacturing wages and productivity and Development

![Graph showing manufacturing wages and productivity](image)

Source: INDSTAT2 (Industrial Statistics Database), the United Nations Industrial Development Organization

hence 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 only, labor productivity as measured by PPP value added per employee is significantly greater for industrialized economies, by a factor of around 4 (see Figure 2). Unsurprisingly, average PPP wages in the manufacturing follow a similar pattern; moreover, the share of wages in total value added is almost twice as large in industrialized countries than in Sub-Saharan African countries. 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.

2.2 Wage dispersion and labor market performance

In this section, we use household survey data for a number of countries in Sub-Saharan Africa to provide evidence that, additionally to being relatively small and unproductive, the wage sector in developing countries exhibits very high levels of wage dispersion. (Detailed data description

---

12See ILO (2012)
can be found in Appendix A.) For example, we find that the top 10% of wage earners in Uganda and Ethiopia are paid 15 and 20 times more than the bottom 10%, respectively, while in the US this ratio is 8 times. Furthermore, Figure 3 suggests that wage dispersion decreases with mean wages and with levels of GDP per capita.

Figure 3: Distribution of log wages for 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.

We bring this result to the forefront not just an empirical curiosity, but instead to provide a more complete characterization of labor markets that allows for a better understanding of their shortcomings. Namely, conditional on the same mean, a greater wage dispersion reflects a higher degree of market inefficiencies and, hence, it is more informative about constraints to job creation and wage growth.

An important question pertaining to what shapes wage distributions is whether they can be explained by differences in labor markets between rural and urban areas, regions, industries or a combination of all of them. A simple variance decomposition in Table 1 shows two important features: (i) the patterns of low mean wages and high dispersion are similar in both urban and rural areas and (ii) regional and industry variations explain an important share of wage dispersion, suggesting the presence of high mobility costs. However, as in the US, most of the variation

13The fact that there exists a large and persistent labor productivity gap across sectors (especially in agriculture versus manufacturing) in developing countries is well known (see, for example, Gollin, Lagakos and Waugh (2014) for recent evidence).
comes from within region-industry groups.

Table 1: Variance decomposition

<table>
<thead>
<tr>
<th></th>
<th>Log wages</th>
<th>Percentage of log wage variance explained by</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std Deviation</td>
</tr>
<tr>
<td><strong>Uganda</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4.49</td>
<td>1.03</td>
</tr>
<tr>
<td>Urban</td>
<td>4.72</td>
<td>1.03</td>
</tr>
<tr>
<td>Rural</td>
<td>4.27</td>
<td>0.99</td>
</tr>
<tr>
<td><strong>Ghana</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4.79</td>
<td>0.88</td>
</tr>
<tr>
<td>Urban</td>
<td>4.86</td>
<td>0.86</td>
</tr>
<tr>
<td>Rural</td>
<td>4.60</td>
<td>0.89</td>
</tr>
<tr>
<td><strong>South Africa</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.67</td>
<td>0.81</td>
</tr>
<tr>
<td><strong>US</strong></td>
<td>7.76</td>
<td>0.74</td>
</tr>
</tbody>
</table>

Note: The sample is limited to 16-65 year old private sector employees. Regional variance includes distinction between rural and urban areas for Uganda Total, Ghana Total, South Africa and the US. Source: Authors’ calculations based on Uganda National Household Survey 2010-2011, Ghana Living Standards Survey 2005, South Africa labor Force Survey 2000-2004, and Survey of Income Program Participation (SIPP) 2004 for the US.

We propose that this evidence highlights a more general pattern of labor markets in least developed countries. To explore this point further, in Table 2 we document different measures of wage dispersion obtained from micro-data (i.e. household or labor force surveys) for a set of poor Sub-Saharan African countries. We include measures in South Africa and the US as a benchmark. Countries are ordered from lowest to highest average log wages (column (1)). 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 using 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 wage and column (5) looks at the ratio between the 90th percentile and the median wages. In general, there seems to be substantial wage differences for Sub-Saharan African 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.

Table 1 suggested that wage dispersion cannot be fully explained by differences across regions or sectors. We look at this more carefully in column (6), by looking at the dispersion of residuals of a log wage regression that controls for demographics (gender, age, education, marital status),
Table 2: Measures of wage dispersion for Sub-Saharan African Countries and the US

<table>
<thead>
<tr>
<th>Country</th>
<th>Mean log Wage</th>
<th>Std. dev of log Wages</th>
<th>Gini Index 50:10</th>
<th>90:50 Wage Ratio</th>
<th>Wage Ratio residuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tanzania</td>
<td>4.14</td>
<td>1.11</td>
<td>0.57</td>
<td>3.6</td>
<td>3.75</td>
</tr>
<tr>
<td>Uganda</td>
<td>4.45</td>
<td>0.99</td>
<td>0.53</td>
<td>3.86</td>
<td>3.73</td>
</tr>
<tr>
<td>Kenya</td>
<td>4.63</td>
<td>0.98</td>
<td>0.53</td>
<td>3.1</td>
<td>3.87</td>
</tr>
<tr>
<td>Ghana</td>
<td>4.79</td>
<td>0.88</td>
<td>0.47</td>
<td>2.75</td>
<td>2.91</td>
</tr>
<tr>
<td>Cameroon</td>
<td>4.88</td>
<td>0.80</td>
<td>0.44</td>
<td>2.66</td>
<td>3</td>
</tr>
<tr>
<td>Zambia</td>
<td>4.94</td>
<td>1.00</td>
<td>0.57</td>
<td>2.75</td>
<td>5.45</td>
</tr>
<tr>
<td>Nigeria</td>
<td>4.96</td>
<td>1.02</td>
<td>0.65</td>
<td>3.2</td>
<td>3.75</td>
</tr>
<tr>
<td>South Africa</td>
<td>6.02</td>
<td>0.83</td>
<td>0.48</td>
<td>2.50</td>
<td>3.50</td>
</tr>
<tr>
<td>USA</td>
<td>7.76</td>
<td>0.74</td>
<td>0.38</td>
<td>2.67</td>
<td>2.43</td>
</tr>
</tbody>
</table>


location (rural/urban, region) and industry. As expected, wage dispersion remains substantial after controlling for observables and, more interesting, the negative correlation between mean wages and residual wage dispersion holds. This is a crucial observation, as it suggests that frictions in labor markets are important, and it will inform our modeling choice of labor markets in developing countries in the next section.

This evidence seems to be at odds with a well-established stylized fact that economists have long associated with the Kuznets curve (see Figure 4). Namely, that there is a positive correlation between average income and inequality for countries at low levels of income (i.e. Sub-Saharan African countries, mostly placed on the left part of Figure 4).

Table 3 presents the magnitudes of wage dispersion and income inequality for the same set of countries, ordered by PPP income per capita. Measures of income inequality, whether Gini Index in column (2) or the standard deviation of log income in column (3), seem to suggest a negative correlation between inequality in labor markets (as measured by wage dispersion) and inequality in income (that includes workers in the wage sector and in self-employment activities). That is, high wage inequality and low income inequality can coexist when a large fraction of the workforce is engaged in low-income non-wage activities. Finally, the highest wage dispersion seems to be coming from countries with smaller wage sectors, as shown in column (5).

The evidence presented in this section suggests that a model of labor markets in least devel-
developed countries that attempts to rationalize poor performance in terms of wages and size of wage employment 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 developing countries. This not necessarily so in middle or high-income countries where wage employment is the norm and wage dispersion is lower.
Table 3: Income and wage distributions for Sub-Saharan African Countries and the US

<table>
<thead>
<tr>
<th>Country</th>
<th>(1) GDP per capita</th>
<th>(2) Gini Index (Income)</th>
<th>(3) Std. dev of log Income</th>
<th>(4) Size of Wage sector</th>
<th>(5) Std. Dev of log Wage residuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uganda</td>
<td>1268</td>
<td>0.43</td>
<td>0.77</td>
<td>0.17</td>
<td>0.76</td>
</tr>
<tr>
<td>Tanzania</td>
<td>1579</td>
<td>0.41</td>
<td>0.71</td>
<td>0.11</td>
<td>0.87</td>
</tr>
<tr>
<td>Kenya</td>
<td>2239</td>
<td>0.48</td>
<td>n/a</td>
<td>0.30</td>
<td>0.74</td>
</tr>
<tr>
<td>Cameroon</td>
<td>2501</td>
<td>0.41</td>
<td>n/a</td>
<td>0.19</td>
<td>0.64</td>
</tr>
<tr>
<td>Ghana</td>
<td>2612</td>
<td>0.45</td>
<td>0.75</td>
<td>0.20</td>
<td>0.70</td>
</tr>
<tr>
<td>Zambia</td>
<td>3221</td>
<td>0.57</td>
<td>0.96</td>
<td>0.25</td>
<td>0.62</td>
</tr>
<tr>
<td>Nigeria</td>
<td>4825</td>
<td>0.44</td>
<td>0.79</td>
<td>0.07</td>
<td>0.75</td>
</tr>
<tr>
<td>South Africa</td>
<td>11138</td>
<td>0.61</td>
<td>1.11</td>
<td>0.51</td>
<td>0.58</td>
</tr>
<tr>
<td>USA</td>
<td>51004</td>
<td>0.55</td>
<td>0.98</td>
<td>0.87</td>
<td>0.60</td>
</tr>
</tbody>
</table>

GDP per capita, measured at PPP values (constant 2011 international $), and GINI Index (Income) are reported by the World Bank’s World Development Indicators. For wage and wage sector size measures the sample is limited to 16-65 year old private sector employees. Reported measures are obtained from own calculations using the following datasets: 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, and Survey of Income Program Participation (SIPP) 2005 for the US.

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 the 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. The model, in addition to capturing labor market frictions, search costs and self-employment sector, incorporates two other standard features from the firm (i.e. labor demand) side. First, we assume that firm entry is costly and, second, that a firm’s success or exit is linked to its idiosyncratic productivity.

While firms are heterogeneous, we assume that workers are homogeneous. We focus on the

14In the Harris-Todaro model migrants move until income in the traditional sector is equalized to the expected income in the modern/urban sector (i.e. probability of getting a job times income). This model uses random matching between jobs and job-seekers. The search framework incorporates this fundamental aspect of Harris-Todaro into a model of steady-state unemployment where wages and unemployment are jointly determined. It also adds the notion that firms and workers spend resources before jobs are created and production takes place and that labor market inefficiencies matter. The Lewis model is about the relative attractiveness of modern jobs relative to self-employment or traditional activities, where a large surplus of labor means low marginal returns. We capture this feature in a self-employment sector with lower income due to low productivity in traditional activities that make jobs in the wage sector more attractive, i.e. workers are willing to queue and to pay a search cost to find wage sector jobs.

15This framework, first developed by Hopenhayn (1992), has been used in a variety of settings. See Aw, Chung and Roberts (2003), Melitz (2003) and Bartelsman, Haltiwanger and Scarpetta (2013a), for example.
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 shows residual dispersion to matter. Second, consider differences in unobserved worker characteristics. One could think that variables like quality of education, where children from a privileged background profit more from the same level of education, or innate differences in ability could explain a higher wage dispersion in poorer countries. However, as the number of jobs is very limited, these same children are 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 model has two labor markets - the wage sector and the home production, or self-employment, sector. The wage sector is populated by heterogeneous firms that differ in their productivity level \( p \). There is infinitely many potential firms that may generate a new product and enter the market 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.

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 \( v \) 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 homogeneous of degree one in both arguments - aggregate vacancies \( v \) and job seekers \( u \). As it is standard in the literature, we assume a Cobb-Douglas form, i.e.

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

where \( m \) is a matching efficiency parameter. Given the constant returns to scale assumption, we can express the job finding rate and the job filling rate as a function of market tightness, \( \theta = \frac{v}{u} \).

That is, when workers search for a job they receive an offer at Poisson arrival rate \( \lambda = \frac{M(v,u)}{u} = \)

\[16\text{Moreover, recent empirical evidence shows that changes of earnings inequality in many countries, either increasing or decreasing, can be accounted for primarily by changes between and not within firms, suggesting a larger role for firm (as opposed to worker) heterogeneity. See, for example, Song, Price, Guvenen, Bloom and von Wachter (2015) for the US and Benguria (2015) for Brazil.} \]
While the vacancy filling rate is given by 

\[ q = \frac{M(v,u)}{\rho} = m\theta^{\gamma-1}. \]

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. Finally, both workers and firms are risk neutral and they discount the future at rate \( r \).

Workers without a job end up in the home production, or 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 home production sector is assumed to be competitive so that wages in this sector are determined by the marginal product of labor (see for example Zenou, 2008). The aggregate production function in the home sector can be written as

\[ Y_H = AF(L_H), \] (2)

where \( Y_H \) is aggregate output produced in the self-employment sector, \( A \) is the aggregate productivity and \( F(\cdot) \) is an increasing concave differentiable function of labor employed in the home sector, \( L_H \). We also assume the usual Inada conditions, i.e. \( F(0) = 0 \) and \( \lim_{L_H \to 0} F(L_H) = +\infty \). In particular, we assume that the production function in the home sector takes the Cobb-Douglas form, i.e. \( Y_H = AL_H^\gamma \), with \( 0 < \gamma < 1 \). Then, under the assumption that self-employment labor markets are competitive, a worker’s income in home production is the marginal product of labor, or

\[ w_H = \gamma AL_H^{\gamma-1}. \] (3)

Moreover, all self-employed workers are assumed to be looking for a paid job and hence the number of jobseekers is equal to \( u = L_H \), the condition for which will be verified later (we relax this assumption in Appendix B). This setup implies that a larger self-employment sector is associated with lower incomes.

---

17Banerjee and Duflo (2007) describe how the poor choose to run their own-account businesses because they cannot find a steady well-paid job in the wage sector and not because of their entrepreneurial drive.

18This measure includes factors assumed to be fixed, such as land or aggregate capital that are rented to self-employed workers. Landlords or other owners of these factors are excluded from the income analysis. They receive the surplus not earned by the self-employed.

19For the agricultural sector, for example, this could be interpreted as the amount of land being fixed as in Matsuyama (1992). Alternative explanations that link a larger size to low marginal product assume instead that productivity decreases 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. In general, this feature of the model intends to capture the phenomenon described in Fields (2011) and Banerjee and Duflo (2007): workers in places where self-employment activities are a large share of employment express a desire
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 $w$ satisfies the following Bellman equation:

$$rW(w) = w + \delta(U - W(w)), \quad (4)$$

where 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

$$W(w) = \frac{w + \delta U}{r + \delta}. \quad (5)$$

The job search is a costly process that involves direct search costs, as well as 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:

$$rU = w_H(L_H) - z + \lambda(\theta) \int (\max\{W(w), U\} - U) dG(w), \quad (6)$$

where $r$ is the common firms’ and workers’ discount rate; $\lambda(\theta)$ 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 (4), we can solve for the worker’s reservation wage $w_R$ that equates the value of search with that of working:

$$w_R = w_H(L_H) - z + \frac{\lambda(\theta)}{r + \delta} \int_{w_R} (w - w_R) dG(w), \quad (7)$$

so that only wage offers above $w_R$ are accepted.

to move to employment in the wage sector.
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:

\[
rf(p) = p - w + \delta (V(p) - J(p)).
\]  
(8)

The first term on the right-hand side of equation (8) 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

\[
rV(p) = -c + q(\theta) (J(p) - V(p)),
\]  
(9)

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). 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:

\[
w = \arg \max_w \left( p - w \right)^{1-\bar{\beta}} \left( w - w_H \right)^{\bar{\beta}},
\]  
(10)

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

$$ w(p) = \beta p + (1 - \beta)w_H, \quad (11) $$

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 entry condition implies that

$$ -k + \int_{\hat{p}} V(p) d\Gamma(p) = 0, \quad (12) $$

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. $w(\hat{p}) \geq w_R$. Below, we derive the formal condition for this to be true in equilibrium.

Substituting for the wage function $w(p)$ into the value of a job given in equation (8), we can rewrite the flow value of a vacancy as follows

$$ rV(p) = -\left(\frac{r + \delta}{r + \delta + q(\theta)} c + \frac{q(\theta)}{r + \delta + q(\theta)} (1 - \beta) \left(p - w_H(L_H(\theta))\right)\right). \quad (13) $$

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 (12) as the expected gain relative to the outside option of exiting the market, $V(\hat{p})$:

$$ rk = \int_{\hat{p}} \left(rV(p) - rV(\hat{p})\right) d\Gamma(p) = \frac{q(\theta)}{r + \delta + q(\theta)} (1 - \beta) \int_{\hat{p}} \left(p - \hat{p}\right) d\Gamma(p), \quad (FE) $$
where we have used the fact that $V(\hat{p})$ is equal to zero. It is convenient to define surplus function as $\varphi(\hat{p}) = \int_{\hat{p}}^{\hat{p}} (p - \hat{p}) d\Gamma(p)$, 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 $\varphi(\hat{p}) = \int_{\hat{p}}^{\hat{p}} (1 - \Gamma(p)) dp$, 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.

$$\frac{c}{q(\theta)} = (1 - \beta) \frac{\hat{p} - w_H(L_H(\theta))}{r + \delta}. \quad (14)$$

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 $w_H$). Since the match surplus is an increasing function of $p$, all $p > \hat{p}$ matches are accepted. 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,

$$\lambda(\theta)L_H = \delta L_F = \delta(1 - L_H) \Rightarrow L_H = \frac{\delta}{\delta + \lambda(\theta)} \tag{15}$$

where $L_H$ is the mass of workers in the home production sector and $L_F$ is the mass of workers in 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 depicted in Figure 5.
Coming back to the workers’ reservation wage, we can re-write it in productivity terms based on wage determination rule defined in equation (10).

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

We are interested in an equilibrium, in which the reservation productivity \( p_R \) is not higher than \( \hat{p} \), so that the productivity distribution and hence the wage distribution are determined by the labor demand side. The formal condition for the reservation productivity to be equal to at most \( \hat{p} \) is the following:

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

This condition is likely to be satisfied if the searching costs \( z \) are relatively high, the worker’s bargaining power \( \beta \) is low and market tightness is low (or equivalently the home production sector is large). We consider these assumptions to be plausible in the developing countries’ context given the virtual absence of a social security system and a relatively high degree of information frictions and low mobility in the labor markets.

Finally, the value of search in equilibrium has to be higher than the value of engaging in home production and not participating in the labor market, and hence not incurring the search costs. That is, we assume that

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

We can show that one can always find a value of \( z \) such that both inequalities (17) and (18) are sat-
isfied (see Appendix B). Hence, in what follows we impose both of these conditions in equilibrium and discuss in detail the significance of this assumption in Section 3.8.

3.6 Comparative statics

There are a number of parameters that can potentially affect equilibrium values, and presumably generate the relation 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 parameter $m$, home sector productivity $A$, workers’ bargaining power $\beta$, and parameters in the 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 vacancy filling rate has to rise and hence market tightness has to fall, shifting the free entry curve downwards. As a result of the increase in $k$, both the reservation productivity and market tightness fall (see Figure 6). As it becomes more difficult to enter the market, the number of firms falls and so does the vacancy-to-unemployment ratio. A 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 the average productivity level of firms as the reservation productivity is now lower.

Figure 6: The effect of an increase in the entry costs $k$.

An increase in the 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 upwards. 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, with both ZP and FE shifting in the same direction as when \( m \) increases. Intuitively, a rise in matching efficiency, \( m \), or firms’ bargaining power, \( 1 - \beta \), has a positive impact on the value of a firm and induces firm entry, thus increasing market tightness \( \theta \). The effect on productivity threshold \( \hat{p} \) is less obvious from the graphs; however, below we show that changes to \( \hat{p} \) in this case are unambiguous.

Another interesting question is what happens for larger values of home production productivity \( A \). Larger values of \( A \) increase workers’ outside option and thus their wages. Hence, the marginal firm needs to be more productive now so that the Zero Profit condition shifts upwards. As a result, \( \hat{p} \) increases and \( \theta \) falls, leading to more workers moving to self-employment.

Finally, we inspect changes in the underlying productivity distribution. Suppose that distribution \( \Gamma_1 \) first-order stochastically dominates \( \Gamma_2 \). Consider, for example, a location shift as a special case of first-order dominance, which implies that \( \Gamma_1 \) has a higher mean, while the variance of underlying productivity is the same for both distributions. In that case, a greater mean value of productivity implies a greater expected gain from entry, thus shifting the Free Entry condition upwards. As a consequence, the equilibrium values are higher for both the reservation productivity and the market tightness under \( \Gamma_1 \).

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.** From the graphical analysis above, we showed that a greater \( k \) reduces \( \hat{p} \) and \( \theta \) and that an increase in \( A \) also increases \( \hat{p} \) but reduces \( \theta \). For changes in \( m \), note that the equilibrium market tightness unambiguously rises, as \( q \) is increasing in \( m \) and decreasing in \( \theta \). In order to see what happens to productivity threshold \( \hat{p} \) (and hence the wage dispersion), we combine equations (FE) and (ZP) to obtain
\[
\hat{p} = w_H(L_H(\theta)) - \frac{c}{1 - \beta} + \frac{c\hat{p}}{rk}.
\]
Differentiating \( \hat{p} \) with respect to \( m \), we get:
\[
\frac{\partial \hat{p}}{\partial m} = \left(1 - \frac{c\hat{p}'(\hat{p})}{rk}\right)^{-1} \frac{\partial w_H}{\partial L_H} \left(\frac{\partial L_H}{\partial m} + \frac{\partial L_H}{\partial \theta} \frac{\partial \theta}{\partial m}\right).
\]
We know that in equilibrium \( \frac{\partial \theta}{\partial m} > 0 \) then, given \( \frac{\partial w_H}{\partial L_H} < 0 \) due to a decreasing marginal product
of labor 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 $\frac{\partial \hat{p}}{\partial \theta} > 0$. 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:

$$\frac{\partial \hat{p}}{\partial \beta} = \left(1 - \frac{c\varphi'(\hat{p})}{rk}\right)^{-1} \left[\frac{\partial w_H}{\partial L_H} \frac{\partial L_H}{\partial \theta} - \frac{c}{(1 - \beta)^2}\right].$$

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 $\hat{p}$.

The result on first order stochastic dominance is simple to prove as $\varphi_1(\hat{p}) = \int_{\hat{p}} (1 - \Gamma_1(p)) dp \geq \int_{\hat{p}} (1 - \Gamma_2(p)) dp = \varphi_2(\hat{p})$. 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 $\hat{p}$.

Finally, suppose that $\Gamma_2$ is a mean preserving spread of $\Gamma_1$, i.e. $\int_{\hat{p}} \Gamma_1(p) dp \leq \int_{\hat{p}} \Gamma_2(p) dp$ for all $\hat{p}$ and $\int p \Gamma_1(p) = \int p \Gamma_2(p)$. We can show that $\varphi_1(\hat{p}) - \varphi_2(\hat{p}) = \int p d\Gamma_1(p) - \int p d\Gamma_2(p) + \int \hat{p} (\Gamma_1(p) - \Gamma_2(p)) dp$. Given the same mean, $\varphi_1(\hat{p}) < \varphi_2(\hat{p})$, which implies that the FE curve under $\Gamma_1$ lies below its counterpart under $\Gamma_2$, resulting in lower equilibrium values of market tightness and productivity threshold.

3.7 Wage employment, wages, and income dispersion

In this section 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. Because the link between $\theta$ and the size of the wage sector is unambiguous, it is clear from Lemma 1 how the model can deliver different levels of labor allocations.

In the model, wages are a linear combination of productivity and self-employment income. Therefore, the observed mean and variance in wages are affected by the equilibrium productivity threshold $\hat{p}$. That is, $E(w) = \beta E(p \mid p \geq \hat{p}) + (1 - \beta)w_H$. That means that the results of Lemma 1 have also a direct implication in terms of mean wages, as parameter differences that raise the productivity threshold and market tightness simultaneously, also increase wages (through average productivity and through a greater outside option).

The variance of wages can be written as $\text{Var}(W) = \beta^2 \text{Var}(p \mid p \geq \hat{p})$. Heckman and Honoré (1990) show that the variance is decreasing in $\hat{p}$ as shown in Figure 5 if the productivity distri-
bution 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 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 policy 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 10 to 14 in Appendix C.

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

**Proposition 1** Assume \( d\Gamma(p) \) is log-concave. 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.** Proposition 1 in [Heckman and Honoré (1990)] shows that for log-concave distributions, \( \frac{\partial \text{Var}(p | p \geq \hat{p})}{\partial \hat{p}} \leq 0 \). A simple Taylor expansion gives us the variance of log wages as \( \text{Var}(\log(w)) \approx \frac{\beta^2}{w_H^2} \text{Var}(p | p \geq \hat{p}) \). 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 the variance of observed productivity and an increase in \( w_H \).

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 economies. Log-concave distributions include normal, exponential, logistic, gamma (for shape parameter greater or equal than 1), beta, extreme value, among others.
production and the resulting wage distribution.

More formally, the variance of (log) income \( \ln I \) can be written as

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

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 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.

To show this more formally, we use Taylor approximation of \( \ln w \) around \( \ln w_H \) to linearize log wages:

\[
Var(\ln I) \approx (1 - L_H) \frac{\beta^2}{w_H} \left(L_H(E(\tilde{p}) - w_H)^2 + Var(\tilde{p})\right),
\]

where \( E(\tilde{p}) \) is the average productivity and \( Var(\tilde{p}) \) is the variance of the truncated distribution such that \( p > \tilde{p} \). 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,

\[
\lim_{L_H \to 0} \frac{\partial Var(\ln I)}{\partial k} \approx \frac{\beta^2}{w_H} \left( \frac{\partial L_H}{\partial k} (E(\tilde{p}) - w_H)^2 + \frac{\partial Var(\tilde{p})}{\partial k} - Var(\tilde{p}) \frac{\partial L_H}{\partial k} \left(1 + \frac{2w_H'}{w_H} \right) \right) > 0.
\]

This derivative is positive since we have showed that \( \frac{\partial Var(\tilde{p})}{\partial k} > 0 \) 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 due to an increase in the variance of wages due to a lower productivity threshold, and the fact that some workers now 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,

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

24
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. The same inverse U-shape relationship for income inequality can be observed for changes in labor market frictions $m$.

### 3.8 Discussion of equilibrium conditions

Our model’s 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 workers’ reservation wage, and (ii) all self-employed workers search for a job. The latter assumption is not crucial for the results we have shown above and the model can easily be extended to allow only a fraction of the self-employed workers to search in the wage sector (see Appendix B for details). 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.

The first condition, on the other hand, requires a more detailed discussion. 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 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)](Hornstein2011). 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

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

(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 unemployed 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 Mm 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 would be 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 (15) and assuming that
\( r \approx 0 \), we get

\[
Mm \approx \frac{1}{\frac{1}{u} + \rho}.
\]

This equation implies that there is a negative relationship between the share of wage employ-
ment in the labor market and the mean-min ratio. Firstly, Table 2 shows that this relationship
holds qualitatively in our data, using 50-10 wage percentile ratios. Secondly, also in terms of
magnitudes the Mm 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 Mm 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 Mm 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$^{23}$, and the interest rate of 15%, the model delivers the 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 will become apparent both in the simulation and estimation of the model, a standard search model applied to least developed economies can generate enough wage dispersion. As opposed to the case of the US, the model allows for more dispersion in wages than found in the data.

4 Estimation strategy

To estimate the model’s parameter, we use information on wages, the size of the wage sector, and workers’ characteristics drawn from the household survey data. We restrict our analysis to five countries - Ghana, Nigeria, Uganda, Tanzania, and South Africa - due to specific data requirements. Namely, 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. Below we outline the estimation procedure, paying a particular attention to the measurement of the entry costs in the data and the decreasing returns to scale parameter in self-employment.

$^{23}$In Ghana, the household survey has information on self-employment income. The replacement rate is computed as the ratio of the median self-employment income to median wages. Note that it ignores the searching costs $z$, so in reality it might be even lower.
4.1 Pre-determined parameters

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\(^{24}\). 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 in two groups. In the first group, parameters are set exogenously for the lack of necessary data or are estimated from other data sources outside of 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]).

*Fixed 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]). For the purpose of this simulation, 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 decreasing 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 payments, payments in kind or working for a family farm without pay. Instead, we choose to recover it from the production function in the agricultural sector, given that the majority of self-employed activities in developing countries are related to subsistence farming. In particular, we run the following regression across Sub-Saharan African countries:

$$\ln Y_{Hjt} = b_0 + b_1 \ln L_{Hjt} + b_2 \ln T_{jt} + \varepsilon_{jt},$$ (21)

where $Y_{Hjt}$ is the value of agricultural production in country $j$ at year $t$, $L_{Hjt}$ is the number of

\(^{24}\)We have tried several distributions in the family with 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.
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. That is, we assume that the returns to scale parameter $\gamma$ is the same across the five countries we analyze, while the overall home sector productivity $A$ (capturing the quality of land and other factors of production) is allowed to vary.

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. While both of these values of $\gamma$ might seem to be too low, they are appropriate for the types of self-employment occupations that we have in mind (traditional farming, casual jobs, petty retail, etc.). Moreover, similar values have been found in the literature (see Aragón and Rud, 2015).

### 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 4 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. In the other four countries, the job finding rate is lower that 0.4% a month. Again, this suggests that labor market inefficiencies are prevalent in this region and labor mobility is extremely low.

Using the flows into and out of self-employment, we can compute a steady-state self-employment share as predicted by the model (see equation (15)) 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 those economies are not too far from the steady state. Give 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

---

25 The fact that there exist substantial differences in the agricultural sector productivity across countries is well documented in Gollin et al. (2014), Lagakos and Waugh (2013) and Herrendorf and Teixeira (2011) among others.

26 “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.” Lewis (1954, p. 402).
Table 4: 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>45.2%</td>
<td>30.2%</td>
<td>40.0%</td>
<td>40.8%</td>
</tr>
</tbody>
</table>


4.2 Empirical targets

Using a set of the pre-defined parameters described above, we estimate the remaining vector of parameters $\theta = (c, k, m, A, \sigma, \beta)'$ by the simulated method of moments (see Gourieroux et al., 1993). We are matching the following moments in the data: the mean log wage, the standard deviation of log wage residuals, the self-employment share, the labor share, the entry costs per worker and the average hiring costs as a fraction of mean wages (see Table 6). The latter three moments warrant a more detailed description.

Labor share

We use the average labor share in the model to back out the workers’ bargaining power from

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

which depends on the equilibrium productivity cutoff $\hat{\beta}$ 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 Surveys data. The Enterprise Surveys collect firm-level data from business owners and top managers and cover a broad range of topics including firm’s costs, employment, and performance measures.\(^{27}\) 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

\(^{27}\)See http://www.enterprisesurveys.org for details. The data are available for 2006 in Uganda and Tanzania and for 2007 in Ghana, Nigeria, and South Africa.
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). These values produce a relatively low bargaining power parameter compared to what is commonly used in the literature, however, we expect the bargaining power of workers to be lower in developing countries than in industrialized economies. Moreover, 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 barriers to firm entry have been associated with higher employment in non-wage activities.

To illustrate the relationship between the entry costs and income, we use ‘Starting a Business’ indicators obtained from the World Bank’s Doing Business survey for 2010. Figure 7 confirms a negative relationship between the entry costs and a country’s level of GDP per capita, especially at high levels of entry costs.

Although the legal costs of starting a business align well with the results of our model, it 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 Sub-Saharan African 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 or very severe

---

28For example, Banerjee and Newman (1999) and Ghatak and Nien-Huey Jiang (2002) show how credit constraints can affect the occupational choice of individuals and determine both the size of the modern sector and the level of wages. Buera and Shin (2013) build a model in which financial frictions not only distort the allocation of production factors (capital and entrepreneurial talents) but also slow down their reallocation process. Ayyagari, Demirgüç-Kunt and Maksimovic (2008) examine which obstacles are more restrictive to growth and find that factors related to finance, crime, and political instability directly affect the growth rate of firms, with financial constraints being the most robust of the three.

29See Djankov et al. (2002) and Herrendorf and Teixeira (2011), among others.
obstacles to operations. Poor electricity supply and outages, access to finance, crime and disorder, corruption and informal sector competition are among the most commonly named barriers to operations. A typical case of the severity of each obstacle is presented below in Figure 8 for Ghana. More than 80% of firms state that problems with electricity and power outages represent a major or a very severe obstacle.

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. 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. Secondly, the entry barriers are expressed as per-worker costs in the model since the production function in the formal sector exhibits constant returns to scale. 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 higher than the number of workers at an average firm in the economy, which again understates the magnitude of the entry costs. The model’s

---

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 400% 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 Sub-Saharan African 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 substantial fixed cost of different nature.
Figure 8: Percentage of firms that perceive the following obstacles to operations to be major or very severe in Ghana.

Source: Authors’ tabulations based on the Enterprise Survey data, 2007.

analog of GDP is monthly output, expressed as total production in the wage sector and the home sector less vacancy posting costs.

The data used for the estimation are summarized in Table 5. With the exception of Tanzania, the electricity connection costs are monotonically decreasing with the level of economic development.

**Hiring costs**

There are no enough data available to estimate the vacancy posting costs $c$ directly. Instead, we think of the hiring costs in the model in a more general way - these costs might include direct recruiting costs, costs of screening and training new employees, as well as indirect costs of lost output when, for instance, current employees take their time away from production to train new recruits. The average hiring costs in our model are equal to the flow cost of opening a vacancy $c$ times the probability of filling it, which is equal to $1/q$. The estimates of the hiring costs vary even 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
Table 5: Entry costs

<table>
<thead>
<tr>
<th>Country</th>
<th>Electricity connection cost, % GDP per capita</th>
<th>Average firm size</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.

on data collected by PricewaterhouseCoopers. Abowd and Kramarz (2003) estimate firing and hiring costs directly based on survey data for a representative sample of French firms. They find that the average hiring costs (including the direct training costs) per hire are approximately equal to the median of monthly wages. In developing countries, the training costs might be lower if the jobs are less skill-intensive, on the other hand the recruiting and monitoring costs might be higher given the labor market inefficiencies. For the lack of more precise information, we estimate the vacancy posting costs $c$ so that the resulting hiring costs are equal to one month of mean wages, i.e. $c/q = E(w)$, and run a robustness check with the hiring costs of three months of wages to check the sensitivity of our analysis to this moment.

### 4.3 Estimating the model parameters

We are matching the following moments in the data: the mean log wage, the standard deviation of log residuals, the self-employment share, the labor share, the entry costs and the hiring costs (see Table 5). 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 the identification of the parameters. The variance of log wages is determined primarily by the variance of underlying productivity $\sigma$. However, a higher variance implies a higher mean wage (in contrast to what is observed 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 and the worker’s bargaining power determines the labor share in the model. Finally, we back out the entry costs $k$ from electricity connection costs as a percentage of output and the vacancy cost parameter $c$ from the hiring costs.

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 \( \hat{\psi}(\hat{\theta}) \). 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:

\[
\hat{\theta} = \underset{\theta}{\arg \min} \left( \hat{\psi}(\theta) - \psi \right)' \Omega (\hat{\psi}(\theta) - \psi),
\]

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:

\[
\hat{V} = (D' S D)^{-1},
\]

where \( D \) is the derivative of the vector of moments with respect to the parameter vector, i.e. \( D = \frac{\partial \psi(\theta)}{\partial \theta}' \), 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

\[
ASE(\hat{\theta}_j) = \sqrt{\frac{1}{N} \hat{V}_{jj}},
\]

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 in the working age. Note that for the entry costs and the hiring 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.
Table 6: Empirical targets

<table>
<thead>
<tr>
<th>Country</th>
<th>$E(\ln w)$</th>
<th>$sd(\ln w)$</th>
<th>$L_H$</th>
<th>$\frac{w}{p}$</th>
<th>$k/Y$</th>
<th>$\frac{c/q}{E(w)}$</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>1.0</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>1.0</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>1.0</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>1.0</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>1.0</td>
<td>21703</td>
</tr>
</tbody>
</table>

Note: The costs of electricity connection per worker (see Table 5) 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 $\frac{w}{p}$ 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 average hiring costs as a fraction of wages $\frac{c/q}{E(w)}$ is chosen ad hoc. 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.

5 Results

5.1 Estimated parameters

Table 7 shows the estimated parameters and their standard errors.\(^{31}\)

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 in our data. 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.\(^{32}\)

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

\(^{31}\)Note that the standard errors here understate statistical uncertainty as they do not account for the fact that $\gamma$ and $\delta$ are pre-estimated.

\(^{32}\)Hsieh and Klenow (2009), for example, find sizable gaps in marginal products of labor and capital across plants in the manufacturing sector in China and India relatively to the United States. They show that when capital and labor are hypothetically reallocated to equalize marginal products to the extent observed in the United States, the manufacturing total factor productivity increases by 30%-50% in China and 40%-60% in India. Bartelsman, Hallwanger and Scarpetta (2013b) 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 7: 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>566.3</td>
<td>37.6</td>
<td>0.0008</td>
<td>0.14</td>
<td>1109.4</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>(0.805)</td>
<td>(14.122)</td>
<td>(4.5 \times 10^{-5})</td>
<td>(0.006)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Uganda</td>
<td>391.5</td>
<td>63.7</td>
<td>0.0054</td>
<td>0.25</td>
<td>3737.8</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>(2.337)</td>
<td>(19.631)</td>
<td>(2.4 \times 10^{-4})</td>
<td>(0.011)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Ghana</td>
<td>361.1</td>
<td>123.1</td>
<td>0.0012</td>
<td>0.33</td>
<td>2534.8</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>(0.319)</td>
<td>(20.412)</td>
<td>(4.8 \times 10^{-5})</td>
<td>(0.011)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Nigeria</td>
<td>582.8</td>
<td>126.9</td>
<td>0.0015</td>
<td>0.26</td>
<td>740.8</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>(1.285)</td>
<td>(61.224)</td>
<td>(1.7 \times 10^{-4})</td>
<td>(0.021)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>South Africa</td>
<td>971.1</td>
<td>319.8</td>
<td>0.0074</td>
<td>0.34</td>
<td>1625.4</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>(0.526)</td>
<td>(7.771)</td>
<td>(7.4 \times 10^{-5})</td>
<td>(0.003)</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

Note: Asymptotic standard errors are given in parentheses.

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. Column (1) in Table 8 presents the 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 live of $1.25 and $2 a day. For example, our estimates for Tanzania generate self employment income of about $10 a month or $0.5 a day (assuming 20 working days in a month), which is consistent with the fact that about 70% of the population live below $1.25 a day. In addition, we compare our estimates of self employment income to the marginal product of labor in agriculture, derived from equation \( \text{(21)} \) using the two values of the decreasing returns to scale parameter, \( \gamma \).

Table 8: 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>$2 a day</th>
<th>Marginal product in agriculture, ( \gamma = 0.24 )</th>
<th>( \gamma = 0.15 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tanzania</td>
<td>9.7</td>
<td>67.9</td>
<td>87.9</td>
<td>14.6</td>
<td>8.9</td>
</tr>
<tr>
<td>Uganda</td>
<td>16.9</td>
<td>37.9</td>
<td>64.7</td>
<td>10.6</td>
<td>6.4</td>
</tr>
<tr>
<td>Ghana</td>
<td>33.1</td>
<td>28.6</td>
<td>51.8</td>
<td>31.5</td>
<td>19.2</td>
</tr>
<tr>
<td>Nigeria</td>
<td>32.0</td>
<td>62.0</td>
<td>82.2</td>
<td>139.1</td>
<td>84.6</td>
</tr>
<tr>
<td>South Africa</td>
<td>151.7</td>
<td>16.7</td>
<td>35.2</td>
<td>133.0</td>
<td>81.0</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.

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.\(^{33}\) Our param-

\(^{33}\)Below we perform a robustness check when only a fraction of self-employed workers actively search for a wage
eter values imply that South Africa has nine, 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 (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.

The estimates of workers’ bargaining power parameter $\beta$ range from 0.14 in Tanzania to 0.34 in South Africa. These values reflect the fact that workers’ bargaining position is relatively weak in poor countries due to a lower degree of unionization or individual bargaining stemming from relatively low levels of workers’ human capital. The Global Wage Report 2010/2011 (ILO) shows that the share of unionized workers as a fraction of the workforce is 1.1% in Uganda, 2.2% in Tanzania, 14% in Ghana and 29% in South Africa.

5.2 Outcome elasticities

Using the model, we can examine the empirical relevance of various determinants of job creation and wages and quantify their relative importance. We focus on four channels: the entry costs, $k$, the labor market efficiency, $m$, self-employment productivity, $A$, and underlying productivity dispersion, $\sigma$. We use the estimated parameters for each country to 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 9.

As predicted by the model, a reduction in market frictions (a reduction in $k$ or a rise in $m$) leads to a larger wage sector, higher wages and a lower wage dispersion. That is, lowering entry barriers or improving matching efficiency increases competition in the wage sector and makes it harder for low productivity firms to survive. 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 job. That specification generates a higher value of the matching efficiency parameter $m$; however, it is still much lower than found in more developed countries.

It 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). Cahuc, Postel-Vinay and Robin (2006a) estimate bargaining power parameters for different groups of workers by their skill levels using French matched employer-employee matched. They find that manual workers have a very low (close to zero) bargaining power parameter, while high-skilled workers had a much higher parameter.
adopt new technologies and replace less productive older firms and that restricting entry leads to lower average productivity and sluggish growth.

Productivity in self-employment activities is generally lower in poorer countries and exogenous positive shocks in that sector have been used to explain structural change (see for example Lewis (1954) and Matsuyama (1992)). We find that an increase in the home productivity $A$, while increasing the wage level and decreasing the wage dispersion, has a negative impact on the size of the wage sector. 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 curbs the growth of the modern 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.

1. **A reduction in labor market frictions has the largest impact on wage employment.** One percent increase in the matching efficiency parameter leads to about 0.8 percent increase in the size of the wage sector in South Africa and about 1.7-1.9 percent increase in the other four countries; while a reduction in the entry costs is about half as effective. Hence, if the policy goal is job creation then focusing on labor market frictions needs to be a policy makers’ priority.

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. For example, one percent increase in the matching efficiency leads to 0.26 percent increase in average wages in South Africa and only to 0.01 percent increase in Tanzania. Therefore, addressing both policy
Table 9: Elasticities of outcome variables with respect to changes in policy parameters.

<table>
<thead>
<tr>
<th>Country</th>
<th>k</th>
<th>m</th>
<th>A</th>
<th>σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tanzania</td>
<td>-0.93</td>
<td>1.86</td>
<td>0.00</td>
<td>0.93</td>
</tr>
<tr>
<td>Uganda</td>
<td>-1.06</td>
<td>1.79</td>
<td>-0.08</td>
<td>1.06</td>
</tr>
<tr>
<td>Ghana</td>
<td>-1.00</td>
<td>1.71</td>
<td>-0.07</td>
<td>1.07</td>
</tr>
<tr>
<td>Nigeria</td>
<td>-0.94</td>
<td>1.88</td>
<td>0.00</td>
<td>0.94</td>
</tr>
<tr>
<td>South Africa</td>
<td>-0.39</td>
<td>0.76</td>
<td>-0.07</td>
<td>0.46</td>
</tr>
</tbody>
</table>

variables at the same time amplifies their effects on labor market outcomes.

3. **Non-linear relationship between the reduction in frictions and income inequality.** The effect of the frictions parameters \( m \) and \( k \) on income inequality differs between South Africa and the other four countries: a reduction in market frictions in Ghana, Nigeria, Uganda, and Tanzania leads to an increase in overall income inequality, while the opposite is true in South Africa. This is consistent with our theoretical results. In the poorer countries in our sample the majority of workers (above 85 percent) are self-employed. Hence, 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, thus increasing income inequality.

4. **A substantial impact of the self-employment sector productivity on mean wages, income, and in-
equality. We find that one percent increase in $A$ brings about up to 0.75 percent increase in the overall income (Nigeria) and about 0.4 percent drop in income inequality. Even in South Africa, where the share of self-employed workers (those directly affected by changes in $A$) is relatively low, the increase in workers’ outside option has a substantial effect on mean income and income dispersion. While it is unsurprising that an increase in the home sector productivity leads to higher income levels, it also has a substantial effect on wages: for example, we find that one percent increase in $A$ causes about 0.2 percent rise in the mean wage and about 0.3 percent drop in the standard deviation of log wages. While a rise in home sector productivity has a negative impact on job creation, these effects are close to zero for the poorer countries in our sample. We thus consider this policy to be very effective in fighting poverty and increasing the overall income level in least developed countries, at least in the short run while other market frictions are still in place.

5.3 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 goal, be it the size of the wage sector or the overall income, through a reduction in entry barriers, labor market frictions or an increase in the home sector productivity.

The most obvious policy to achieve development goals is improving firms’ productivity through, for example, opening to 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.36

36Interestingly, 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
In what follows, we want to focus on alternative policies that might be considered as a potential substitute for technological improvement. As previously discussed, the sunk cost that must be incurred by a new firm is a key parameter in many models looking at productivity distribution of incumbent firms. However, quantitative analysis has been less precise in defining these costs, even in counterfactual analysis. For example, Bartelsman et al. (2013a) talk about a combination of factors, including entrepreneur’s effort and administrative fees; Aw et al. (2003) simply mentions regulatory and technological differences in entry costs in two countries (Taiwan and South Korea) and Ulyssea (2010) refers to technological determinants of entry for firms in Brazil. Direct microevidence on entry costs has also failed to provided 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.

We propose that a good way of looking at entry costs in least developed countries is to concentrate on a key input for production, namely electricity. There is a large body of evidence that shows that appropriate access, quality and pricing of electricity can increase firm entry and industrial output (Rud, 2012a), and firm productivity and growth (Abeberese, 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, we suggest that infrastructure policies reducing the cost of access or improving its quality can 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 is stronger in developed countries. 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 fail to induce the reallocation of workers from self-employment to the wage sector.
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. 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, as offer rejections are very high. More promising results are shown in Franklin (2015) and Abebe et al. (2016). The authors used a series of randomized experiment to assign treatments that reduce search costs, such as transport subsidies or training aiming at improving search effectiveness in spatially dislocated areas of Ethiopia. For example, Franklin (2015) shows that weekly transport costs in Ethiopia average about 20% of median total expenditure and that lowering transport costs increases the intensity of job search and increases the probability of finding permanent employment by 30 percent in the short run. Moreover, the subsidies reduce participation in temporary and informal work. This type of innovative labor market policies that include subsidies to workers or firms searching seem to be the most promising way of strengthening labor markets in least developed countries, as costs of search seem high enough to fragment labor markets.

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 increasing productivity in self-employment activities and could be thought as a policy to increase $A$ in terms of our model.

In the analysis below, we use the model to find out how a given country can achieve its development goal. 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. In particular, we set four separate goals - the size of the wage sector, the mean and standard deviation of log wages, and the average income - at the levels observed in South Africa. We then simulate the model for each country and compute what values of a given policy variable $k$, $m$ or $A$ are
Table 10: Policy experiments: Ghana to South Africa

<table>
<thead>
<tr>
<th>Parameter values</th>
<th>Targets</th>
<th>Wage sector size</th>
<th>Average income</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
<td>Baseline</td>
<td>2535</td>
<td>281</td>
</tr>
<tr>
<td>$m$</td>
<td>0.0012</td>
<td>0.0038</td>
<td>0.0123</td>
</tr>
<tr>
<td>$A$</td>
<td>123.1</td>
<td>–</td>
<td>1058.2</td>
</tr>
</tbody>
</table>

required to get to those targets.37

The detailed results of these policy experiments are presented in Table 12 in Appendix D for all countries, while here we concentrate on two targets for Ghana. Table 10 presents the baseline values of the policy parameters and the new values that are required if Ghana were to achieve a given target by using one policy instrument at a time. 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 ten times or, alternatively, increase its labor market efficiency three times. (Note that we do not use an increase in self-employment productivity 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 9 times, the labor market needs to become roughly 10 times more efficient, or, alternatively, the entry costs have to be reduced 80 times.

Table 10 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 subsection, 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 the matching efficiency $m$ is kept constant then $k$ needs to be lowered from 2,535 to 281 international dollars to reach South Africa’s level of wage employment, hence the total gap in $k$ is 2,254 international dollars. However, if $m$ is increased at the same time the required change in $k$ will be smaller.

Figure 9 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

37 Note that we do not consider transition dynamics after the changes are introduced, instead these experiments should be thought of as a comparison between two steady state economies - the baseline versus a counterfactual economy with alternative parameter values.
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 suggest 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 a half and reducing the gap in $A$ by only 12% at the same time.

5.4 Robustness checks

Some of the empirical moments used to estimate the model’s parameters are not readily available in the data. To check the sensitivity of our results to these moments we perform several robustness exercises. First, our estimate of the vacancy posting costs $c$ is pinned down by the average hiring costs in the model. Given the lack of data on the magnitude of the hiring costs in developing countries, the choice of one month of wages is ad hoc. In order to check how sensitive the results are to this moment, we run a robustness check with the hiring costs of 3 months of average wages. Under this assumption, the estimated vacancy posting costs are higher, while the other parameters remain practically the same. This does not alter any of the policy experiment results substantially.

Secondly, we re-estimate the model with an alternative value of the returns to scale parameter in self employment, i.e. $\gamma = 0.15$. This specification increases our estimate of the home productivity parameter $A$, while leaving other parameter values unchanged. Again, the results of the policy experiments and the obtained elasticities of job creation, wages and income are virtually
the same.

Finally, we check the sensitivity of our analysis to the assumption that the search costs are low enough for every self-employed worker to engage in active job search. From the fact that the transition rates from self-employment to wage employment in the data are very low, we can infer that the degree of labor market frictions in least developed countries is very high. An alternative interpretation of this finding is that the searching costs are relatively high compared to the benefits of search so that some workers are discouraged to look for wage employment. In this case, the pool of actively searching workers among the self-employed is smaller resulting in low transition rates from the subsistence sector to the wage sector. We can easily extend our model to allow for the possibility that not all self-employed workers engage in search (see the details in Appendix B). In practice, however, it is very difficult to identify from the data what fraction of self-employed workers are looking for a job, or alternatively what the searching costs are, to estimate the model numerically.

We proceed as follows. Ghana Living Standards Survey 2005 has an underemployment section that asks respondents whether they seek to change their work situation and what steps they take to do it. In our sample, 19.2% of self-employed workers report that they would like to change their work situation and out of those 77% take some active steps towards it (such as applying to employers, searching for newspaper adverts, asking friends or relatives, etc.). We then re-estimate the model under the assumption that the fraction of self-employed workers that actively engage in search is \(0.192 \times 0.77 = 0.148\). We assume that the searching costs are proportional to self-employment income, i.e. \(z = w_H\), and recover the implied value of \(z\). We find that in order to fit the above fraction of active job seekers, the searching costs need to be relatively high: our estimate is equal to 2.6 times the self-employment income \(w_H\), or 86 international dollars. As a result, the matching efficiency parameter is now higher: \(m = 0.0032\) compared to the baseline value of 0.0012.

In addition to changing the matching efficiency parameter, this specification affects other results as well. In particular, all of our outcome variables become more sensitive to changes in policy parameters. Table 11 shows the elasticities of wage employment, mean wages, wage dispersion, mean income, and income inequality, as well as the share of active jobseekers among the self-employed in this alternative specification. There are two reasons for why the elasticities are much higher in magnitude. First, the estimated value of matching efficiency is now higher, and as we discussed above, the existing complementarities between the model parameters makes
other policies more effective. In addition to that, another margin of adjustment in this model is the share of active jobseekers. Consider for example a decrease in the entry costs parameter $k$. A reduction in the entry costs leads to more firms entering the market and thus increasing the benefits from search to workers. The number of the job seekers increases, which makes it easier for firms to fill a vacancy (it decreases the so-called congestion externality when firms by posting more vacancies congest the market and the probability to find a match falls). Hence, the effect on all outcome variables is much larger in the presence of this additional workers’ participation margin.

The ordering of the effects of market frictions parameters is the same as in the baseline specification. For example, increasing matching efficiency is still a more effective way of job creation than a reduction in the entry costs or the searching costs of workers. Intuitively, the effect of a reduction in $k$ or in $z$ has a direct effect only on one side of the market (firms in the former case and workers in the latter case), while an increase in matching efficiency simultaneously improves the job finding probability of workers and the vacancy filling rate of firms, thus having a larger overall effect on job creation. Given these differences, we think of our baseline estimates as a lower bound on the outcome elasticities with respect to $k$ and $m$.

The effect of changes in home productivity parameter $A$ on average wages and wage dispersion is strikingly different. For workers to be indifferent between home production and search, the reservation wage $w_R$ has to be equal to the outside option $w_H$. Since the search costs $z$ and other parameters affecting the firms’ decision (such as $k$ and $m$) are not affected, the reservation wage is unchanged. That is, the number of workers searching will adjust to keep market tightness at the same level. Therefore, even though the productivity in the home sector increases, so does the number of self-employed workers, cancelling the positive productivity effect on self-employment income. Hence, the outside option, the average wage, and the dispersion of wages are not affected by changes in home productivity, $A$. At the same time, the reallocation of workers from wage employment to self-employment impacts the overall income negatively. Thus, increasing self-employment productivity is a counterproductive policy at least until all self-employed workers start actively searching for a job.
Table 11: Robustness check for Ghana: fraction $\alpha < 1$ of self-employed workers search for a job

<table>
<thead>
<tr>
<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>
<th>Share of job-seekers among SE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline specification ($\alpha = 1.0$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k$</td>
<td>-1.00</td>
<td>-0.03</td>
<td>0.04</td>
<td>-0.41</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>0.66</td>
<td></td>
</tr>
<tr>
<td>$A$</td>
<td>-0.07</td>
<td>0.21</td>
<td>-0.30</td>
<td>0.62</td>
<td>-0.48</td>
<td></td>
</tr>
<tr>
<td>$\sigma$</td>
<td>1.07</td>
<td>0.82</td>
<td>0.26</td>
<td>0.80</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Alternative specification ($\alpha = 0.148$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k$</td>
<td>-8.57</td>
<td>-0.23</td>
<td>0.33</td>
<td>-3.59</td>
<td>-3.41</td>
<td>-8.85</td>
</tr>
<tr>
<td>$m$</td>
<td>14.36</td>
<td>0.39</td>
<td>-0.54</td>
<td>6.06</td>
<td>5.15</td>
<td>15.0</td>
</tr>
<tr>
<td>$A$</td>
<td>-8.07</td>
<td>0.00</td>
<td>0.00</td>
<td>-2.74</td>
<td>-3.67</td>
<td>-9.32</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>16.29</td>
<td>1.23</td>
<td>-0.31</td>
<td>7.29</td>
<td>6.27</td>
<td>18.24</td>
</tr>
<tr>
<td>$z$</td>
<td>-7.29</td>
<td>-0.19</td>
<td>0.27</td>
<td>-3.07</td>
<td>-2.89</td>
<td>-8.51</td>
</tr>
</tbody>
</table>

6 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, differences 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


Besley, Timothy and Maitreesh Ghatak, Property Rights and Economic Development, Vol. Volume 5, Elsevier,


Kalemli-Ozcan, Sebnem and Bent E. Sorensen, “Misallocation, Property Rights, and Access to
Finance: Evidence from Within and Across Africa,” Working Paper 18030, National Bureau of

Koelle, Michael and Simon Quinn, “Self-Employment, Capital and Job Search: Evidence from a

Lagakos, David and Michael E. Waugh, “Selection, Agriculture, and Cross-Country Productivity


Lentz, Rasmus and Dale T. Mortensen, “Labor Market Models of Worker and Firm Heterogeneity,”


Magruder, Jeremy R., “High Unemployment Yet Few Small Firms: The Role of Centralized Bar-

Matsuyama, Kiminori, “Agricultural productivity, comparative advantage, and economic

Meghir, Costas, Renata Narita, and Jean-Marc Robin, “Wages and Informality in Developing

Melitz, Marc J., “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry

Mortensen, Dale T. and Christopher A. Pissarides, “Job Creation and Job Destruction in the

Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny, “The Allocation of Talent: Implica-

Pallais, Amanda, “Inefficient Hiring in Entry-Level Labor Markets,” American Economic Review,
2014, 104 (11), 3565–99.

Petrongolo, Barbara and Christopher A. Pissarides, “Looking into the Black Box: A Survey of


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 earning data are given at monthly frequency. However, when wages refer for 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 the residuals in 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, 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. However, 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: Equilibrium conditions

We are looking for an equilibrium, in which the productivity distribution, and hence the wage distribution, are determined by the labor demand side. The worker’s reservation productivity is
given by

\[
p_R = w_H(L_H(\theta)) - \frac{z}{\beta} + \frac{\lambda(\theta)}{r + \delta} \int_{p_R} (p - p_R) \frac{d\Gamma(p)}{1 - \Gamma(\hat{p})} \\
\leq w_H(L_H(\theta)) - \frac{z}{\beta} + \frac{\lambda(\theta)}{r + \delta} \int_{\hat{p}} (p - \hat{p}) \frac{d\Gamma(p)}{1 - \Gamma(\hat{p})} \\
= \hat{p} - \frac{z}{\beta} - \frac{c(r + \delta)}{q(\theta)(1 - \beta)} + \frac{\lambda(\theta)}{r + \delta} \left( 1 + \frac{r + \delta}{q(\theta)(1 - \beta)} \right) \frac{rk}{1 - \Gamma(\hat{p})},
\]

where the first step uses the fact that the surplus function \( \varphi(\cdot) \) 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:

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

(22)

The second condition that has to be satisfied in equilibrium is the participation constraint. That is, the value of search in equilibrium, \( U \), 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 \( w_H/r \). In the latter case the worker does not incur the 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 relatively 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 value of search \( U = w_R/r \) has to be at least as high as the value of home production \( w_H/r \), which implies that the searching costs need to satisfy the following inequality:

\[
z \leq \frac{\lambda(\theta)\beta}{r + \delta} \int_{\hat{p}} \left( p - w_H(L_H(\theta)) \right) \frac{d\Gamma(p)}{1 - \Gamma(\hat{p})} \\
= \frac{\lambda(\theta)\beta}{r + \delta} \left( r + \delta + q(\theta) \frac{rk}{q(\theta)(1 - \beta) 1 - \Gamma(\hat{p})} + \frac{c(r + \delta)}{q(\theta)(1 - \beta)} \right) \\
= \frac{\beta}{1 - \beta} \left( \frac{rk}{1 - \Gamma(\hat{p})} \left( \theta + \frac{\lambda(\theta)}{r + \delta} \right) + c\theta \right),
\]

(23)

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 $z$

$$\frac{\beta}{1-\beta} \left( \frac{rk}{1-\Gamma(\hat{p})} \left( \theta + \frac{\lambda(\theta)}{r+\delta} \right) + c\theta \right) \geq z \geq \frac{\beta}{1-\beta} \left( \frac{rk}{1-\Gamma(\hat{p})} \left( \theta + \frac{\lambda(\theta)}{r+\delta} \right) - \frac{c(r+\delta)}{q(\theta)} \right), \quad (24)$$

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

Based on our parameter estimates (and in line with the exposition presented in Section 3.8), we find that the first inequality (22) is not binding for our model in any country, except South Africa. The implied searching costs in SA need to be higher than $w_H$, resulting in a negative consumption flow while searching, to match the observed magnitudes of wage dispersion. 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 then a more advanced model is needed, which allows for either workers’ heterogeneity or on-the-job search and that accounts separately for the large informal sector (around a third of workers) to fit the data.

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 $m$, which leads to an increase the productivity threshold $\hat{p}$ and the job finding rate $\lambda$. It means that the searching costs $z$ will have to be higher for inequality (22) to hold. This assumption is not very restrictive due to two reasons. First, we can think of searching costs $z$ as a fraction of home production income (e.g. due to lost output when workers actively search instead of producing at home) and hence we can expect an increase in $z$ related to an increase 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 on the effect of policy variables on wage levels and dispersion.

Finally, suppose that inequality (23) 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 in the labor market 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, so that workers are indifferent between searching in the

---

38 Meghir et al. (2015) and Ulyssea (2010) consider a large informal sector that is on average less productive than the formal sector and where transitions between informal and formal sector are important.
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 $\theta = \frac{v}{\delta_H}$ and the steady state self-employment is determined by $L_H = \frac{v}{\delta + a \lambda(\theta)}$. Now, in addition to equilibrium equations (ZP) and (FE), we have one additional condition that $w_H = w_R$ and one additional unknown $\alpha$. We can express the participation constraint as

$$z = \frac{\lambda(\theta)}{r + \delta} \left( \int_{\hat{p}} p \frac{d\Gamma(p)}{1 - \Gamma(\hat{p})} - \hat{p} + \frac{c(r + \delta)}{q(\theta)(1 - \beta)} \right),$$

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

**Appendix C: 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, together with the level of labor market competition, 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 low productivity firms.

The left panel of Figure 10 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 10 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
Figure 10: 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 11 shows that an increase in the 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 11). 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 constant wages) 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 11 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 11) 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 12 illustrates the effect of changing the location parameter $\mu$ and the scale parameter
First, we find that a country with a higher average underlying productivity will also have a lower wage dispersion (see the left panel of 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 13 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. While

\[^{39}\text{We use a logistic productivity distribution, because it allows us to analyze separately the effects of changes in the mean from changes in the variance of underlying productivity. Moreover, a logistic distribution belongs to the family of log-concave densities; hence, the results of our Proposition 1 are applicable here.}\]
we expect unionization to increase mean wages, and the model delivers that, it would also reduce job creation.\footnote{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.} This would be hard to reconcile with our stylized fact where countries like South Africa have larger wage employment than presumably less unionized countries.

The second factor that determines a worker’s bargaining position is her outside option $w_H$, which is affected by productivity in home production, $A$, and may vary across countries. In our framework, an increase in productivity in self-employment increases the outside option, and hence workers’ wages, leading to a smaller wage sector. 

The top panels of Figure 14 illustrate the effects of bargaining power parameter $\beta$ and productivity parameter $A$ on the size of the wage sector. Both of these variables increase the wage cost to the firm, which discourages entry into the market and leads to a smaller value market tightness, thus reducing the wage sector. This result is similar in spirit to standard competitive labor market models, which predict that unions are facing a tradeoff of increasing wages at the expense of lower employment.

Based on these results, we conclude that the hypothesis that economic development leads to an increase in unionization, which in turn compresses wages, is not supported in our model. In fact, as can be seen on the bottom left panel of Figure 14, a higher degree of unionization that is captured by higher $\beta$ increases the variance of wages (through a lower value of $\hat{p}$ and a larger weight put on productivity as opposed to the outside option). Additionally, similarly to changes in the underlying firm productivity across countries, differences in self-employment productivity...
Figure 14: The effect of changes in workers’ bargaining position on the size of the wage sector and wage dispersion

alone cannot explain the stylized facts we observe in the data, in particular the existence of a negative relationship between the size of the wage sector and the variance of wages (the bottom right panel of Figure 14).

Appendix D: Policy experiments

Table D1 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 12: Policy experiments: Targeting South Africa’s values

<table>
<thead>
<tr>
<th>Parameter values required to fit the following targets</th>
<th>Baseline size</th>
<th>St. dev of log wages</th>
<th>Mean income</th>
<th>Mean wage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tanzania</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k</td>
<td>1109.4</td>
<td>68.5</td>
<td>27.9</td>
<td>1.0</td>
</tr>
<tr>
<td>m</td>
<td>0.0008</td>
<td>0.0033</td>
<td>0.0068</td>
<td>0.0254</td>
</tr>
<tr>
<td>A</td>
<td>37.6</td>
<td>–</td>
<td>150.5</td>
<td>1078.8</td>
</tr>
<tr>
<td>Uganda</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k</td>
<td>3737.8</td>
<td>357.0</td>
<td>330.7</td>
<td>14.8</td>
</tr>
<tr>
<td>m</td>
<td>0.0054</td>
<td>0.0176</td>
<td>0.0273</td>
<td>0.0983</td>
</tr>
<tr>
<td>A</td>
<td>63.7</td>
<td>–</td>
<td>176.2</td>
<td>1073.7</td>
</tr>
<tr>
<td>Ghana</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k</td>
<td>2534.8</td>
<td>280.9</td>
<td>408.9</td>
<td>32.6</td>
</tr>
<tr>
<td>m</td>
<td>0.0012</td>
<td>0.0038</td>
<td>0.0037</td>
<td>0.0123</td>
</tr>
<tr>
<td>A</td>
<td>123.1</td>
<td>–</td>
<td>215.1</td>
<td>1058.2</td>
</tr>
<tr>
<td>Nigeria</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k</td>
<td>740.8</td>
<td>23.1</td>
<td>61.8</td>
<td>5.3</td>
</tr>
<tr>
<td>m</td>
<td>0.0015</td>
<td>0.0070</td>
<td>0.0085</td>
<td>0.0181</td>
</tr>
<tr>
<td>A</td>
<td>126.9</td>
<td>–</td>
<td>287.3</td>
<td>1084.3</td>
</tr>
</tbody>
</table>