Wage employment, unemployment and self-employment across countries*

Markus Poschke†

McGill University, CIREQ and IZA

April 2018

Abstract

The variation of employment status across countries follows three broad patterns: Labor markets in poor countries exhibit (1) systematically higher self-employment rates and (2) much higher rates of unemployment relative to wage employment. In addition, (3) countries with high unemployment relative to wage employment have higher self-employment rates even conditional on GDP per capita. I interpret high unemployment to employment ratios as evidence of labor market frictions, and develop a simple heterogeneous-firm search and matching model with choice between job search and self-employment to analyze their effect. Quantitative analysis of the model, separately calibrated to eight countries, suggests that variation in labor market frictions can explain almost the entire variation in not only unemployment, but also self-employment across the calibration countries. The model also generates joint variation in unemployment and self-employment accounting for a third or more of their relationship in the data. In addition, the analysis shows that labor market frictions affect output not only via their effect on employment, but also by pushing searchers into low-productivity own-account work.

Keywords: unemployment, self-employment, occupational choice, entrepreneurship, firm size, productivity

*I would like to thank Christoph Hedtrich, Andrei Munteanu, Javad Samieenia, Xian Zhang and, in particular, Masaya Takano for excellent research assistance. I would also like to thank Girum Abebe, Stefano Caria, Simon Franklin, Douglas Gollin, Christian Moser, Alemayehu Seyoum Taffesse, Ruth Vargas Hill and seminar participants at the IGC Workshop on Dimensions of Structural Transformation (Addis Ababa, September 2016), the 2016 SMAUG Workshop on Macroeconomics, Search & Matching, the 2017 Conference on Human Capital and Financial Frictions at Georgetown University, the 7th European Search and Matching Network meeting at the Barcelona GSE Summer Forum, and the University of Toronto for valuable comments and suggestions. I gratefully acknowledge financial support from the International Growth Centre (IGC), and thank the Ethiopian Central Statistical Agency, the IGC, the Minnesota Population Center and the statistical agencies participating in IPUMS International for facilitating data access.

†Contact: McGill University, Economics Department, 855 Sherbrooke St West, Montreal QC H3A 2T7, Canada. e-mail: markus.poschke@mcgill.ca

1 Introduction

Labor markets in low income countries differ fundamentally from those in advanced economies. They feature a very large number of self-employed and own account workers, and a preponderance of small firms. Mirroring this, the rate of wage employment is low, and there is a large number of job searchers. In Addis Ababa, the capital of Ethiopia, for example, the self-employment rate (own-account workers plus employers) exceeds 30%, and the unemployment rate exceeds 20%.

The employment structure in poor countries contrasts with that in rich countries, where most workers are employed in large firms. In the United States for example, own-account workers account for only about 5% of employment, whereas about half of all employment is in firms with more than 500 employees (Hipple (2010), Census Business Dynamics Statistics).

These two examples are not isolated cases. The high prevalence of self-employment in poor countries is well known (see Gollin 2007). The first contribution of this paper is to document two additional, new facts on self-employment and unemployment in the cross-section of countries using harmonized census data from more than 60 countries provided by IPUMS International (Minnesota Population Center 2017). First, while the unemployment rate (unemployment/labor force) does not vary systematically with income per capita (in line with Caselli (2005)), the ratio of unemployment to the labor force excluding self-employment is much higher in poorer countries.[1] On average, it increases by two and a half percentage points every time income per capita doubles. As a result, it is almost 10 percentage points larger in the poorest countries compared to the richest ones. Second, in urban labor markets, self-employment is particularly high in countries with high unemployment u relative to employment n, even after controlling for GDP per capita. The relationship is quantitatively strong: an increase in $u/(u+n)$ by one percentage point is associated to an increase in the self-employment rate by around 0.7 percentage points.

High values of unemployment relative to wage employment suggest that job seekers are not very successful, indicating strong labor market frictions. As these make job search less attractive, they make self-employment relatively more attractive. High self-employment in poor countries may thus at least partly be due to lower attractiveness of job search.

Since small firms and own-account workers typically are less productive, the difference in employment structure may have implications for aggregate productivity and welfare. Not surprisingly, the high prevalence of own-account work and the small size of firms is an important concern to policy makers in poor countries. It has also given rise to an academic literature examining the link between the firm size distribution and aggregate productivity

[1] Note that the unemployment/employment ratio is a monotonic transformation of this ratio.

(see notably Restuccia and Rogerson 2008, Hsieh and Klenow 2009).

This literature has exclusively analyzed the allocation of employment across employer firms, and largely ignored both unemployment and self-employment – despite their importance in poor economies. The second contribution of this paper consists in addressing this gap. To do so, I develop a theoretical framework that allows linking unemployment, self-employment, and productivity, and allows exploring their connections via counterfactual analysis.

Existing models of frictional labor markets have been tailored to advanced economies, and therefore are not suitable for analyzing economies with high self-employment. (The few exceptions are discussed below.) For analyzing labor markets in advanced economies, it may be acceptable to abstract from self-employment, given its low cyclicality there. This abstraction is less acceptable in countries where a third of the labor force is self-employed, and there are almost as many own-account workers as employees. To fill this gap, I extend the standard Diamond-Mortensen-Pissarides (DMP) search and matching model to be able to capture the high rates of self-employment and the importance of small firms typical of low income countries.

In doing so, I aim to deviate as little as possible from a standard DMP model. I add two features to the model: (1) choice between job search on the one hand and entry into entrepreneurship on the other hand, combined with (2) firm heterogeneity. The first feature clearly is needed to be able to say anything about entrepreneurship and self-employment. I assume that while job search is subject to search and matching frictions, entry into entrepreneurship is always possible at a cost. Success, however, is uncertain, as the second feature implies that entrants differ in productivity. This assumption delivers a meaningful distinction between own-account workers and employers, and also allows addressing the determinants of the small size of firms in low income economies. The firm size distribution and the entry rate into entrepreneurship then are endogenous model outcomes. Finally, I also model casual jobs in a very simple way, to reflect their importance in poor countries.

I then calibrate the model using data on labor market states and flows and the firm size distribution for the urban areas of eight countries, ranging in income level from Addis Ababa, the capital of Ethiopia, via Mexico, Indonesia and some European economies to the United States. The use of information on labor market flows in poor countries is an important, novel feature of the analysis. Calibrating the model to various countries shows how it can accommodate very different labor market conditions. It also permits analyzing quantitatively which cross-country differences, out of a large set of potential candidates, are the determinants of the strongly dispersed unemployment and entrepreneurship rates observed in the data.

This analysis points to variation in labor market frictions as the main determinant of cross-country differences not only in unemployment, but also in self-employment. Differences in labor market frictions explain almost all the variation in unemployment and self-employment across the eight calibration economies. The model also accounts for at least a third of the relationship between self-employment and the unemployment-employment ratio found in the data. In contrast to this, variation in parameters more directly related to self-employment, like entry costs or the relative productivity of own-account workers compared to employer firms, could also explain observed patterns in self-employment, but generates counterfactual variation in unemployment.

Stronger labor market frictions make job search less attractive, and thereby promote self-employment. They also discourage firms from hiring, leading to an economy with more, smaller firms and more own-account workers. This effect is particularly strong in economies with low entry costs for firms. This implies that while variation in labor market frictions in a high-entry cost economy, like the US, mostly affects the unemployment rate, changes in labor market frictions can have a stronger effect on self-employment than on unemployment in a low-entry cost economy.

Labor market frictions also affect aggregate output. Part of this comes simply from their effect on unemployment. This effect is largest in developed economies, where firms face high entry costs. But another part, which is quantitatively very important in poor, low-entry cost economies, comes from the fact that strong labor market frictions induce individuals to take up low-productivity own-account work instead of searching for employment.

Finally, there are other factors besides labor market frictions that affect the relative attractiveness of job search and self-employment. Several of them can be affected by economic policy. For this reason, I also simulate the reaction of the economy to changes in the flow value of unemployment, in the relative profitability of own-account work, and to flat and size-dependent distortions (SDDs). The latter may arise from variation in tax rates and in the enforcement of regulation with firm size. A key theme that emerges from these exercises is that occupational choice is a very important margin of adjustment. For example, transfers to job searchers prompt a reduction in entrepreneurial entry in favor of job search, leading to an increase in the unemployment rate that far exceeds the one in a model with fixed occupational choices. This implies that occupational choice is a margin that needs to be addressed when evaluating the potential effects of policies. I also find that distortions have very different effects on occupational choices compared to labor market frictions. While both reduce average firm size, distortions reduce self-employment by discouraging entry overall, unlike labor market frictions, which encourage own-account work. Distortions, while potentially important, thus are an unlikely explanation of high self-employment rates in poor

countries.

To summarize, there is a strong relationship between self-employment and unemployment in cross-country data. There also is a clear theoretical link: potential job seekers or entrants compare the two options, so that their relative attractiveness affects the number of people engaging in each activity. My quantitative findings suggest that this channel is important, and that variation in labor market frictions can account for a large fraction of the univariate and joint variation in self-employment and unemployment rates across countries observed in the data. Combined with the effect of labor market frictions on output, this calls for more attention to systematic variation in labor market frictions across countries as a determinant of cross-country differences in economic outcomes. Improving labor market functioning in low income economies can thus have multiple benefits: not only reduced unemployment, but also a lower incidence of low-profit own-account work.

Two aspects of the model and results merit discussion at this point. First, a key assumption in both the theoretical and the empirical analysis is that self-employment and job search constitute distinct activities between which individuals need to choose – i.e., they cannot engage in both at the same time. Of course, the assumption that individuals can engage in only one activity at a time is typical for models of occupational choice. It is relaxed in models with on the job search, but even those typically assume that search on the job is less effective than full-time search. This appears to be particularly true for job search in poor countries. In Addis Ababa, for example, job search requires time consuming travel to peruse job ads at centralized job boards, and to drop off CVs in person at companies (Franklin 2014). The cost of job search is substantial (Abebe, Caria and Ortiz-Ospina 2017).

Abebe, Caria, Fafchamps, Falco and Franklin (2016) show that even over longer time spans, it is rare for the unemployed to engage in self-employment. In fact, the unemployed report working only an average 1.3 hours per week in the Ethiopian Urban Employment and Unemployment Survey for 2012 used in this paper. The self-employed in contrast report working an average of 50 hours per week, similar to employees. Self-employment also is highly persistent – substantially more persistent than wage employment – and the self-employed are less likely to transition to wage employment than to unemployment (Bigsten, Mengistae and Shimeles 2007). Self-employment thus truly appears to be a distinct activity from job search, in line with my analysis.

A possible reason for this is that self-employment typically requires some amount of capital, and therefore is not practical as a temporary activity intended to financially sustain job search. It is more common to see occasional casual employment, often day labor, used to finance job search. This does not require the worker to have capital.

A second issue is that the importance of labor market frictions found in the quantitative analysis may at first sight appear to conflict with the notion that firms in poor countries can fill vacancies quickly. For example, Blattman and Dercon (2016) document that manufacturing firms in Ethiopia do not face a shortage of applicants to their vacancies. At the same time, these authors find a very high level of quits and turnover. This suggests that while it may be easy for firms to hire *some* worker, it is much harder to find the *right* worker, and one who will stay. That is, a productive, durable match is hard to form. In this sense, there is no conflict between the quantitative findings from the model and evidence on hiring from the literature.

My findings naturally lead to the question of the precise nature of frictions in urban labor markets of poor countries. Since the model used for the analysis was on purpose kept simple, this question goes beyond the scope of this paper, and should be the subject of future research. There is no shortage of competing candidate explanations. Are matches hard to form because information on vacancy and worker attributes is costly or difficult to convey, e.g. because of low levels of use of information technology, or low levels of skill certification? Does something prevent workers from exercising the optimal amount of search effort? Or do workers have unrealistic expectations, leading them to search in suboptimal market segments or to have high reservation wages? Some of the experimental work cited in the literature discussion on the next few pages takes a first stab at these questions.

Related literature. While existing work on unemployment and job search in developed economies is abundant, there are only a few papers studying poorer economies.[2] Albrecht, Navarro and Vroman (2009), Margolis, Navarro and Robalino (2012), Narita (2014), Bradley (2016) and Galindo da Fonseca (2018) are most closely related to this paper, in that they also allow for self-employment.[3] Yet, their focus is not on labor market frictions and self-

[2]It is also true that little of the work on labor market search in developed countries considers self-employment. Two recent exceptions are Kredler, Millan and Visschers (2014) and Delacroix, Fonseca, Poschke and Ševčík (2016), who study the joint determination of unemployment and self-employment over the business cycle in the United States, Canada and Europe. To the best of my knowledge, they seem to be the first to do so since the earlier paper by Fonseca, Lopez-Garcia and Pissarides (2001), who focus on the effect of entry barriers in the OECD.

[3]Zenou (2008), Ulyssea (2010), Bosch and Esteban-Pretel (2012), Meghir, Narita and Robin (2015) model firms' choice of formality versus informality in macroeconomic models of search and analyze how policies, in particular the enforcement of regulations, affect the share of formal jobs, unemployment and aggregate output. None of them allows for an occupational choice by workers or job seekers, ruling out the analysis of self-employment by construction. Rud and Trapeznikova (2016) do allow for self-employment, but do not model occupational choice. They assume that all workers who do not find a job in a constant-returns sector engage in self-employment. Finally, Gollin (2007) is a key paper that shows how the self-employment rate declines with income per capita across countries. The paper quantitatively analyzes the relationship in a span of control model building on Lucas (1978), but does not address labor market frictions.

employment, but on the effect of taxes, unemployment insurance benefits, severance pay and entry costs on output and/or the size of the informal sector.

The present paper is also different in terms of methodology. First, none of the papers mentioned conducts a cross-country analysis. Second, the papers all assume that self-employment or entrepreneurship opportunities arrive at a fixed, exogenous rate. The exogenous arrival rate implies that the self-employment rate can respond to changes in the environment only via a selection effect. This limits variation in the self-employment rate, and limits the impact of occupational choice on aggregate outcomes, which I find to be large.

To the best of my knowledge, the only other paper analyzing unemployment across countries spanning the distribution of income is the very recent paper by Feng, Lagakos and Rauch (2017).[4] These authors find that in data for the entire country, unemployment rates are higher in more developed countries, and attribute this to the higher productivity of the formal sector there. I discuss the relationship between their and my findings in Section 2.2. In a nutshell, my findings and theory are consistent with theirs. However, their empirical findings only hold at the level of the entire country, and are influenced by the presence of unpaid workers. When focussing on urban areas, which allow for a better comparison across countries at different levels of development, the finding of higher unemployment relative to wage employment in poor countries is robust.

Finally, a few recent papers study search behavior, labor market frictions and self-employment in developing economies at the micro level. Both Franklin (2016) and Abebe et al. (2016) find that reducing search frictions at the individual level improves job search outcomes. Bassi and Nansamba (2018) find that certifying worker skills affects labor market outcomes. Blattman and Dercon (2016) show that unpleasant jobs are often taken temporarily, to cope with adverse shocks or finance search for better jobs or future self-employment, and that self-employment is considered desirable by many. Lagakos, Moll, Porzio, Qian and Schoellman's (2018) finding of flatter experience-wage profiles in poorer countries is also consistent with more severe search frictions in poorer countries. Donovan et al. (2017) also interpret their findings as suggesting that self-employment constitutes an important alternative to job search. This work is highly complementary to this paper, and helps to indicate potential manifestations of the precise nature of frictions in urban labor markets in poor countries.

The paper is organized as follows. The next section documents the joint relationship of self-employment and unemployment rates and GDP per capita across countries, and briefly

[4]Donovan, Lu and Schoellman (2017) document labor market *flows* for 13 countries at different levels of development. Bick, Fuchs-Schündeln and Lagakos (2018) document how hours worked vary with income per capita within and across countries.

discusses related literature. Section 3 presents the model. Quantitative results are shown in Section 4 to 7. Section 4 describes the calibration of the model economy using data from eight countries. Section 5 determines which factors are the main quantitative determinants of cross-country differences in unemployment and self-employment. Section 6 analyzes the effects of labor market frictions on unemployment, self-employment and productivity in more detail, while Section 7 shows the effects of other changes in the environment on these outcomes. Section 8 concludes. Appendices contain additional figures and tables, and additional details on theory and numerical methods.

2 Key features of labor markets in low income economies

This section presents evidence on the relationship between self-employment, wage employment and unemployment across the income distribution of countries. I begin by describing data sources.

2.1 Data sources and measurement issues

My main source of data for comparing self-employment and unemployment across a broad set of countries consists in the censuses available via IPUMS International. IPUMS International provides access to micro data from almost 200 censuses collected in more than 60 economies since 1960 (Minnesota Population Center 2017). While the bulk of the data was collected after 1980, there are 40 censuses collected between 1960 and 1980. The number of censuses per country ranges from one to nine, with a median of four. Censuses typically take place every ten years. This data source is very versatile, as it allows computing measures of self-employment and unemployment not only for the aggregate economy, but also for subgroups (like urban residents, young workers, etc.) for many countries. For example, an urban unemployment rate can be computed for 137 censuses from 55 countries.[5]

A critical issue in cross-country comparisons is harmonized measurement. This is why IPUMS harmonizes country data, aiming to provide comparable measures. This being said, unemployment and self-employment are inherently elastic concepts, and care needs to be taken. In general, the census data has the additional advantage for researchers that some measurement issues can be overcome by suitable sample restrictions.

In the IPUMS census data, unemployment is computed from the harmonized EMPSTAT (employment status) variable. This classifies individuals as employed, unemployed, or inac-

[5]Throughout, I limit the analysis to countries with a population of at least one million.

tive.[6] The union of the employed and the unemployed constitutes the labor force.

Unemployment is known to be difficult to compare consistently across countries. IPUMS aims to apply UN and ILO standards, defining the unemployed as persons out of work who are actively searching for a job. The search criterion is difficult to apply in poor countries, where the job search process for many workers may not be very formal. Reflecting this, the IPUMS census data allow defining both a narrow and a relaxed unemployment rate. The difference between the two consists in the categories of "unemployed because no work available" or "inactive unemployed", which are separately identified in the IPUMS data (where available).[7] My main results use the relaxed definition of unemployment, which many statistical agencies in developing economies regard as the most useful one for their setting (see e.g. Central Statistical Agency of Ethiopia 2015).[8] However, results are very similar for the narrow definition.[9]

To assess whether differences in measurement conventions across censuses matter, I group the observations in three tiers of comparability, like Feng et al. (2017). The top tier contains censuses where the reference period for the employment status question is clearly specified as the past week. In the second tier, the reference period consists of the last four weeks. Censuses using any other reference period, or lacking a clear specification of one, make up the third tier. Robustness checks reported below show that, apart from somewhat smaller statistical significance due to lower sample size, results are generally similar when restricting the analysis to the top comparability tier.

The measure of self-employment is derived from the IPUMS CLASSWK (class of worker) variable, which categorizes the employed as either self-employed, wage or salary workers, unpaid workers, or other. Typically, those who worked at least one hour in the reference period, including informal work or day labor, are considered employed. In many cases, more detailed information on the employer and the type of contract is available. For the self-employed, most censuses distinguish employers and own-account workers.

[6] Age barriers defining the universe that is asked about employment status can differ slightly across countries. I deal with this by verifying that results are not driven by specific age groups.

[7] Overall, the relaxed and narrow measures of unemployment are strongly correlated, with a correlation coefficient of 0.97. But for some countries, in particular poorer ones, the relaxed unemployment rate may exceed the narrow one by up to ten percentage points. (Across censuses, the median difference is 0, the 75th percentile is 0.016, and the 90th percentile is 0.041.)

[8] In line with this, the 19th International Conference of Labor Statisticians in 2013 passed a resolution to introduce the "potential labor force" as a desirable additional measure at the aggregate level. In addition to the employed and the unemployed, defined in the usual way, this includes both unavailable job seekers (searching but temporarily unavailable) and available potential job seekers (not actively searching but available). The latter group includes "discouraged job seekers". Unfortunately, no ILO measures of the potential labor force or of relaxed unemployment are available at the time of writing.

[9] It would be desirable to also measure an underemployment rate, defined by hours worked. Unfortunately, this information is only available in few censuses.

Finally, countries differ strongly in their economic structure and, as is well known, the structural composition of the economy is strongly associated with development (see e.g. Herrendorf, Rogerson and Valentinyi 2014). For instance, in poor countries, many workers work in agriculture, often on the family farm. To minimize the effect of these differences, my main analysis uses data not for the entire country, but for urban areas, which are more similar across countries both in their economic structure and in the functioning of labor markets. The IPUMS data are key for being able to do this. I report results for the entire country when it is informative.[10]

For robustness, I also consult aggregate measures of unemployment and self-employment from the ILO. These are mostly computed from labor force surveys, and are typically annual. An important disadvantage of this source is that only country-level measures are available. Given the importance of agriculture in poor countries, these are less comparable across countries than the measures for urban areas computed using IPUMS data.

2.2 The distribution of employment status and development

Figure 1 depicts the prevalence of different types of employment status in urban areas by country log income per capita.[11] Employment status is a detailed measure that combines information from EMPSTAT and CLASSWK. It can take one of four values for any member of the labor force: unemployed (not working, but searching for a job), wage or salary worker (employed in somebody else's firm), self-employed (own-account worker or employer), or "other". Individuals who do not fall into any of these four groups are considered not to be in the labor force.

The figure shows, for each country, cumulative shares. For any country, the lowest marker (triangles) shows the proportion of unemployed labor force members (the unemployment rate), the difference between the black dot and the triangle shows the share of wage/salary workers, and the difference between the grey dot (at the top of the figure) and the black dot shows the fraction of the labor force that is self-employed. Finally, the difference between the grey dot and one gives the fraction of "other". Since this is negligible, I ignore this category in the following. I also exclude unpaid workers.

For each set of points, I plot a line of best fit for an OLS regression on log GDP per capita. The shading of areas makes the prevalence of different employment statuses across

[10] Ideally, one might also want to account for sectors directly. However, apart from the conceptual difficulty of assigning job seekers to a particular sector, the number of censuses reporting the sector of (un)employment is also much more limited than that reporting urban versus rural status, at 88 compared to 150.

[11] Income per capita throughout is from Penn World Tables 9 (Feenstra, Inklaar and Timmer 2015).

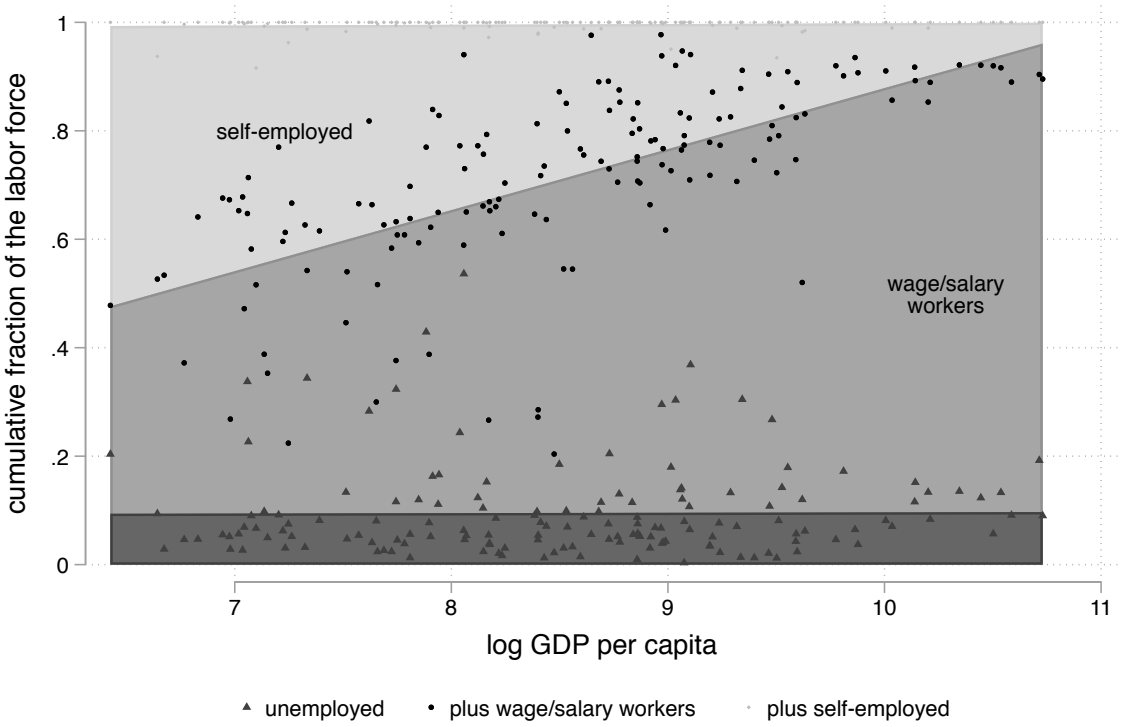


Figure 1: Composition of the labor force and development

Sources: GDP per capita: PWT 9.0. Employment status: IPUMS International. 196 censuses covering 64 countries over the years 1960 to 2011. Data for urban areas. Bottom area: unemployment rate.

the country income distribution very clear.

It is immediate from the figure that the self-employment rate declines strongly with development, echoing the well-known finding of Gollin (2007). Self-employment rates range from almost 80% of the labor force in the poorest countries to about 10% in the richest ones.

The unemployment rate, in contrast, does not vary systematically with development, although it is quite variable across countries. The counterpart of this is that the fraction of wage and salary workers in the labor force strongly increases with development, from only about 10% of the labor force in the poorest economies to around 80% in the richest ones.

Regression results underlying the lines in Figure 1 are reported in Table 1. They are similar no matter whether the regression is run on country averages (as in the table), or whether censuses are pooled (as in the figure and in Table 17 in the Appendix). The unemployment rate does not vary systematically with log income per capita, whereas the employment rate and the self-employment rate vary symmetrically: the self-employment rate declines by 0.13 percentage points for each 1% increase in income per capita, and the

Table 1: Composition of the labor force and development

dependent variable:	self-employment rate	rate of wage employment	unemployment rate	UN ratio
Urban areas:				
log GDP per capita	-0.132***	0.138***	0.003	-0.035**
	(0.017)	(0.017)	(0.009)	(0.014)
R^2	0.507	0.543	0.002	0.099
observations	150	150	165	150
countries	58	58	65	58
Entire country:				
log GDP per capita	-0.187***	0.183***	0.012*	-0.033***
	(0.016)	(0.014)	(0.007)	(0.011)
R^2	0.670	0.718	0.041	0.121
observations	214	214	235	214
countries	68	68	77	68

Notes: The table shows regression coefficients from regressions of the dependent variable on log GDP per capita, using time averages of country data (between regression). Constant not reported. * (**) [***] indicates $p < 0.1$ (< 0.05) [< 0.01]. Data sources as in Figure 1. Results for a regression using pooled data are similar and are shown in Table 17.

employment rate increases by roughly the same amount. This translates into a decline in the self-employment rate, and an equivalent increase in the employment rate, by 9 percentage points every time income per capita doubles. The lower panel of the table shows that regression results for the entire country are similar, with even larger coefficients in absolute terms. Figure 10 shows results for the entire country graphically. That figure and Figure 11 also show the role of unpaid workers. Table 18 shows that results are essentially identical when only information from countries in the top tier of data comparability is used.

Table 2 shows that the pattern in self-employment is driven by own-account workers. The fraction of employers actually is higher in richer economies. These two results hold both for urban areas and overall. Since on average, employers account for only 18% of the self-employed, and account for less than half almost everywhere, it is clear that the overall pattern for the self-employed is driven by own-account workers.

Figure 1 clearly shows the importance of self-employment in poor economies. It also shows another pattern: while unemployment as a fraction of the labor force does not vary with income per capita, the ratio of job seekers to employees clearly does. To analyze this

Table 2: The relationship between entrepreneurship rates and income per capita

dependent variable:	fraction own-account workers, urban	fraction employers, urban	fraction own-account workers, entire country	fraction employers, entire country
log GDP per capita	-0.143***	0.012***	-0.190***	0.010***
	(0.020)	(0.003)	(0.019)	(0.002)
R^2	0.512	0.236	0.629	0.273
observations	140	140	189	189
countries	53	53	63	63

Notes: The table shows regression coefficients from regressions of the dependent variable on log GDP per capita, using time averages of country data (between regression). Constant not reported. Standard errors in parentheses. * (**) [***] indicates $p < 0.1$ (< 0.05) [< 0.01]. Data sources as in Figure 1. Results for a regression using pooled data are similar (not reported).

object, define the "UN ratio" as $u/(u+n)$, where u denotes the unemployment rate and n the employment rate. Since the UN ratio differs from the unemployment rate only in its denominator, it has a similar scale. While the unemployment rate has a median of 7% (10th percentile: 2%, 90th percentile: 19%) in the IPUMS data, the UN ratio has a median of 11% (10th percentile: 4%, 90th percentile: 33%).

Given how few employees there are in poor countries, it is clear from the figure that the UN ratio attains systematically higher values in these countries. This is corroborated by the regression coefficients in the last columns of Table 1, which are economically and statistically significant. They show that the UN ratio declines by 2.5 percentage points as country income per capita doubles.

Tables 3 in the main text and 21 in the Appendix show that this finding is robust to several potential concerns. First, the pattern is not due to differences in demographics, since it holds within age group, both in urban areas and at the level of the entire country. Second, the relationship between the non-participation rate and GDP per capita is very similar to that between the UN ratio and GDP per capita. This implies that even if there may be some misclassification between unemployment and non-participation, in particular in classifying the inactive unemployed, the negative relationship between the UN ratio and GDP per capita should still go through.[12] Finally, the relationships between the unemployment rate, the UN

[12]The relationship established here differs from that in Bick et al. (2018), who find higher employment rates (including self-employment) in poorer countries. The difference is not driven by data quality or sample period: even when only using tier 1 data and limiting the sample to the year 2000 and later, I still find significantly lower participation in urban areas of poor countries. Instead, the difference appears to be driven by sample composition. Notably, Bick et al.'s (2018) sample does not include several poor, low-participation countries from the IPUMS data because of lack of comparability of their hours data, whereas it includes

Table 3: Unemployment and development, subsamples

dependent	unemployment rate			UN ratio		
variable:	age 20-29	age 30-60	age 61-65	age 20-29	age 30-60	age 61-65
Urban areas:						
log GDP per capita	0.004	0.006	0.009	-0.052***	-0.022*	-0.034**
	(0.013)	(0.008)	(0.008)	(0.018)	(0.013)	(0.014)
R^2	0.001	0.008	0.023	0.123	0.053	0.095
observations	165	165	159	150	150	145
countries	65	65	62	58	58	56
Entire country:						
log GDP per capita	0.018*	0.011**	0.013**	-0.046***	-0.023**	-0.036***
	(0.010)	(0.005)	(0.005)	(0.015)	(0.010)	(0.012)
R^2	0.044	0.051	0.081	0.127	0.078	0.123
observations	235	235	226	214	214	208
countries	77	77	75	68	68	68

Notes: The table shows regression coefficients from regressions of the dependent variable on log GDP per capita, using time averages of country data (between regression). Constant not reported. Standard errors in parentheses. * (**) [***] indicates $p < 0.1$ (< 0.05) [< 0.01]. Data sources as in Figure 1.

ratio, and log GDP per capita are similar when a narrow measure of the unemployment rate is used. All of this holds both for the entire country and for urban areas only.

Table 19 in the Appendix shows that the relationships between the self-employment rate, the unemployment rate and GDP per capita are similar in ILO data.

Having stressed the strong, robust, negative relationship between the UN ratio and development, I conclude the analysis of the relationship between the distribution of employment status and development with a closer look at the relationship between the unemployment rate and development. The regression coefficient capturing this relationship tends to be positive, but insignificant in data for urban areas. It is slightly larger and therefore significant in data for the entire country. The coefficient of 0.012 in the lower panel of Table 1 is very close to that of 0.015 obtained by Feng et al. (2017) using similar data. Figure 2 and Table 20 in the Appendix show the importance of the denominator of the unemployment rate in driving this result. The table shows that the fraction of unpaid workers is much smaller in richer countries, in particular when data for the entire country is used. (The fraction of unpaid workers

some poor, high-participation countries which are not in IPUMS International.

is smaller in urban areas across countries, and therefore matters less there.) As a result, including unpaid workers (who are considered "employed" according to the CLASSWK and EMPSTAT variables in the IPUMS censuses) in the analysis drives up the denominator of the unemployment rate in poor countries, suggesting lower unemployment there. Excluding them from the analysis (and thus from the denominator of the unemployment rate) leads to an entirely flat relationship between unemployment and development. Further excluding the self-employed from the denominator yields the UN ratio, with its negative relationship with development established above. Figure 2 shows these three relationships visually using data for urban areas. (See Figure 12 for a similar figure using data for the entire country.)

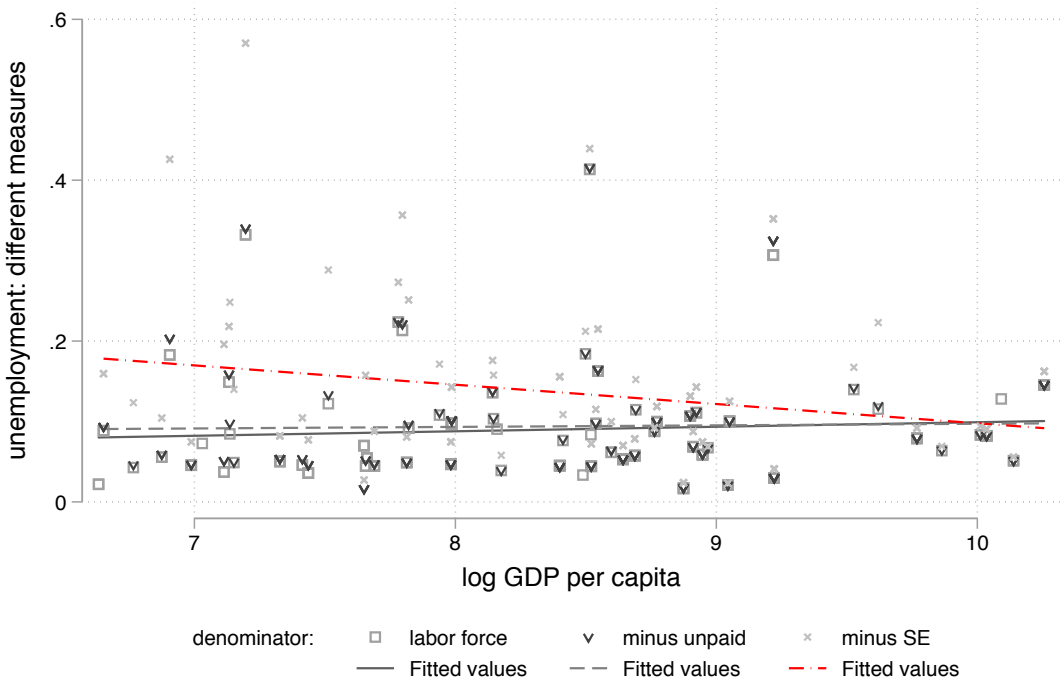


Figure 2: Different measures of unemployment and development

Notes: Data sources as in Figure 1. Data for urban areas. Regression outputs underlying the lines of best fit reported in Table 20.

To summarize, there are large differences in the distribution of employment status across countries at different points of the global income distribution. Comparing countries, doubling income per capita goes along with a reduction in the fraction of the labor force engaged in self-employment by 9 percentage points, a corresponding increase in the fraction engaged as wage or salary workers, and a decline in the UN ratio of 2.5 percentage points.

These patterns imply that the apparent constancy of the unemployment rate with development is misleading: among those who are not self-employed, the share of unemployed people is much higher in poorer countries. This suggests that the functioning of labor markets differs systematically with development: while the fraction of the labor force searching for a job does not vary systematically with income per capita, the fraction that actually ends up with a job is much lower in poorer countries.

This failure to transform job seekers into employees suggests either very limited hiring by firms, worse functioning of labor markets, very quick destruction of jobs, or any combination of these. All of these imply that job search is less attractive in poorer countries, either because it is less likely to be successful, or because jobs, once found, do not last long. This should affect occupational choice, pushing the unemployed away from job search and encouraging own-account work. High self-employment in poor countries may thus at least partly be due to lower attractiveness of job search. This argument suggests that there should be an independent connection between the UN ratio, as a measure of the (un)attractiveness of search, and self-employment. I now turn to examining this relationship.

2.3 Self-employment and unemployment

Figure 3 shows the bivariate relationship between the self-employment rate and the UN ratio, again using IPUMS census data.[13] It is clear that there is a positive relationship between the two variables, both in urban areas (left panel) and in countries as a whole (right panel). The figures show this relationship up to the 90th percentile of the UN ratio. (For urban data, the relationship flattens above this level of the UN ratio due to the influence of a few censuses; see Figure 13 in the Appendix.) The relationship is both economically and statistically significant, with a regression coefficient of 0.79 for both samples, implying an almost one-to-one relationship between the self-employment rate and the UN ratio.

Table 4 shows that this relationship is robust to also controlling for log GDP per capita. The table reports results for urban areas, again for a sample truncated at the 90th percentile of the UN ratio, in line with the findings in Figure 13. This table shows that the coefficient on the UN ratio is positive, and economically and statistically significant. It is clear that the relationship is driven by own-account workers. These results imply that an increase in

[13] It would not make much sense to analyze the relationship between the self-employment rate and the unemployment rate u. Since these two rates plus the employment rate must by construction add up to one, there is a strong mechanical negative relationship between them: when one rises, the other one must fall, unless the employment rate absorbs all the variation. This mechanical relationship dominates any other pattern there might be. The UN ratio can also be thought of as a tool to analyze the relationship between self-employment and unemployment, while eliminating the mechanical relationship.

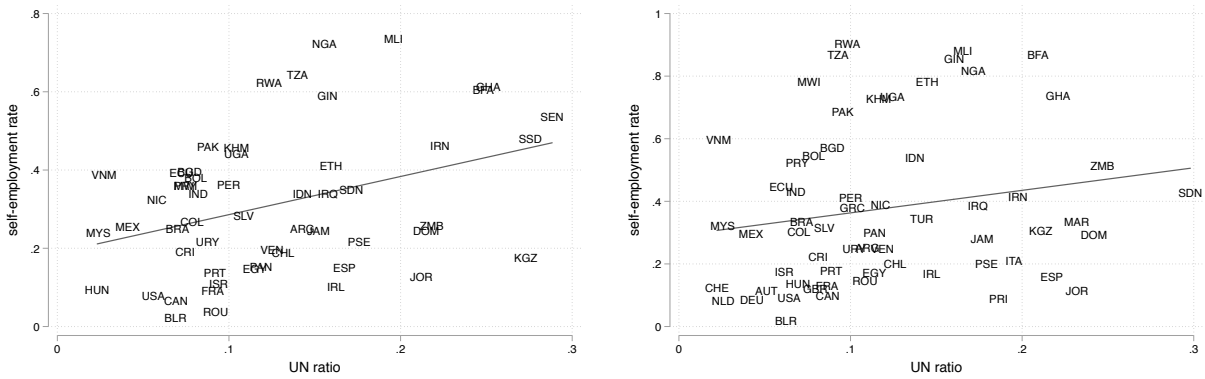


Figure 3: The self-employment rate versus the UN ratio $u/(u+n)$, urban (left) and overall (right)

Notes: The solid line shows the fit from an OLS regression. Graphs and regressions exclude observations of UN ratio above the 90th percentile of the variable (0.31). Full range shown in Figure 13 in the Appendix. The regression coefficients are 0.97 (standard error 0.35) for urban areas and 0.72 (standard error 0.49) for the entire country.

the UN ratio by one percentage point, at a constant level of GDP per capita, is associated with an increase in the self-employment rate by half a percentage point, due to an increase in the fraction of own-account workers by 0.6 percentage points. Results also indicate that self-employment is lower in richer countries, with a coefficient that is similar to that of the bivariate relationship between the self-employment rate and income per capita. Results for a pooled regression are similar (see Table 22 in the Appendix). When using only data for countries in the top data comparability tier, the point estimate in the first column is essentially identical, only the standard error a bit larger, as the sample is a third smaller (see Table 23).

Results are different when using data for the entire country. Here, the inclusion of GDP per capita in the regression leads to an insignificant coefficient on the UN ratio (see Table 24 in the Appendix, and also Table 25 using ILO data). This is not entirely surprising. When using data for the entire country, data for poor countries includes many respondents in rural areas, where small-scale agriculture is highly prevalent, and where there are few large employers, limiting opportunities for wage employment. As a result, it is plausible that in these areas, opportunities for wage employment will have hardly any effect on employment choices by individuals. To ensure comparability across countries, I will focus on the results for urban areas shown in Table 4.

Summarizing the analysis in this section, the comparison of urban labor markets of

Table 4: The relationship between self-employment and the UN ratio, controlling for GDP per capita, urban areas

dependent variable:	self-employment rate	fraction own-account workers	fraction employers
UN ratio	0.702**	0.802**	0.058
	(0.285)	(0.312)	(0.051)
log GDP per capita	-0.122***	-0.136***	0.012***
	(0.018)	(0.020)	(0.003)
R^2	0.556	0.575	0.229
observations	136	126	126
countries	54	48	48

Notes: The table shows regression coefficients from regressions of the dependent variable on the UN ratio and log GDP per capita, using time averages of data (between regression). Constant not reported. Standard errors in parentheses. * (**) [***] indicates $p < 0.1$ (< 0.05) [< 0.01]. Data sources as in Figure 1. Results for a regression using pooled data are similar (Table 22).

countries at different stages of development reveals three regularities: Labor markets in poor countries feature (1) systematically higher self-employment rates and (2) higher rates of unemployment relative to wage employment (a higher UN ratio), and (3) self-employment is higher in countries with high unemployment relative to wage employment, even conditional on GDP per capita.

2.4 A concrete example: Ethiopia

To conclude this section, I present data describing the labor market in Addis Ababa, the capital and major city of Ethiopia. This is a particularly poor economy, with very high levels of self-employment and unemployment.[14] It is also one of the countries used for calibrating the model in Section 4.

Table 5 shows the distribution of employment status across categories in Addis Ababa in 2015. First, it is evident that a large share of individuals reports to be unemployed.[15] Among

[14]Over the last decade (with the exception of 2016), economic growth in Ethiopia has been rapid, with increasing urbanization and a slight increase in the size of the private sector. This makes Ethiopia in 2015 reasonably similar to other East African economies.

[15]This figure is in line with that directly reported by the Ethiopian CSA for Addis Ababa, for example in Central Statistical Agency of Ethiopia (2015). Note that both figures use a "relaxed" definition of unemployment, as advocated by the Ethiopian CSA as appropriate for the local context. The relaxed concept includes persons without work who are available for work, but not actively searching. According to ILO figures, which use a narrower concept, the median unemployment rate in Ethiopia over the period 2004 to 2014 is 16%. Another difference is that the ILO figure includes rural areas, where unemployment is

the employed, less than a quarter of those working in the private sector hold permanent employment of the type typical in advanced economies. A group that is just as large is engaged in casual or temporary employment, while an even larger group engages in self-employment (employers and own-account workers). Evidently, permanent private sector jobs, while considered desirable, are only one of several common forms of employment. Entrepreneurship is just as important, and many workers are engaged in casual work. Anecdotal evidence suggests that this work often serves to finance job search, another very common activity (see e.g. Franklin 2014).

Table 5: Distribution of labor force status and private sector employment status in 2015 in Addis Ababa, Ethiopia

Labor force status	share (%)	Employment status	share (%)
Employed	45.6	Private permanent	24.2
Unemployed	14.2	Self-employed	33.6
Not in the labor force	40.2	Own-account workers	24.2
		Employers	9.4
Unemployment rate	23.8	Private casual or temporary	22.7
		Domestic	12.8
		Other	6.7

Source: Urban Employment and Unemployment Survey for 2015 collected by the Ethiopian Central Statistical Agency. Excluding the public sector and government enterprises, which account for 21.1% of total employment, and unpaid family workers (1.9%). Details on data treatment are provided in Appendix C.

Due to scarcity of panel data, information on flows is more limited.[16] For the case of Ethiopia, Bigsten et al. (2007) show significant flows across employment categories over the four-year interval from 2000 to 2004. For example, 54% of those who are self-employed in 2000 are still (or again) self-employed in 2004. 7% of the unemployed become entrepreneurs, and 26% become employees. 22% of employees become unemployed. The full transition matrix is reported in Section 4, where I compare it to flows predicted by the model.

systematically lower for low income economies.

[16]While the World Bank's Living Standards Measurement Study has a panel component, it is too small to be informative for the case of urban Ethiopia.

3 Theory

Having documented the relationship between self-employment and unemployment, the second objective of this paper is to develop a simple benchmark model that can account for key features of labor markets not just in advanced economies, but for a broad cross section of countries. This section sets out such a model.

I base the model on the Diamond-Mortensen-Pissarides (DMP) model of random search and matching in labor markets, a workhorse model for the analysis of labor markets in developed economies. I extend the model in three ways. First, the unemployed can choose whether to search for a job or enter entrepreneurship (occupational choice). Second, firms are heterogeneous. It is not optimal for all entrants to continue in business. Among those who continue, some become own-account workers and others employers. The latter in turn differ in the optimal size of their firms. Finally, the unemployed periodically engage in casual work to sustain their job search. As a result, the model generates an equilibrium partition of the population into the unemployed, employees, own-account workers, employers and causal workers, as well as a distribution of firm sizes.

These features constitute the minimum extension of the DMP model required to be able to reproduce the above-mentioned facts, and to study the effect of labor market frictions on unemployment, self-employment, and firm sizes. Clearly, endogenizing the entrepreneurship rate requires giving model agents the ability to choose between entrepreneurship and employment or job search.[17] Allowing for firm heterogeneity allows capturing the difference between own-account workers and employer firms, and it also allows frictions to affect not only the quantity of entrepreneurs, but also their quality and size. It also enables the analysis to address the observed small size of firms in low income economies. Finally, casual jobs are introduced in a simple way because they are so common in poor economies. Their presence allows the unemployed to sustain job search for prolonged periods of time.

3.1 States, flows and the labor market

Time is discrete. The economy consists of a measure one of homogeneous individuals. They value the net present value of income, discounting future income using a discount rate r. In any period, individuals die with a fixed, exogenous probability ϕ, and a measure ϕ of newborn individuals enter unemployment. An individual can be in exactly one of four states:

[17] I also explored a version of the model where not only the unemployed can become self-employed, but where the employed can also leave their jobs to engage in entrepreneurship. (For this to occur in equilibrium, it has to be the case that entry is more favorable for them compared to the unemployed, for example because they are on average better entrepreneurs.) Quantitative results for that model are broadly similar, but it is computationally more cumbersome.

unemployment, employment, own-account work, or being an employer. Let their measures be u, n, e_s and e_f. A fraction of the unemployed engages in casual work in any period.

Flows. Any period, a number of endogenous and exogenous flows across the four states in the economy can occur. The exogenous flows occur with fixed, exogenous rates, and are as follows. Existing matches dissolve with a probability ξ. Own-account workers and employers need to close their business with probabilities λ_s and λ_f, respectively. All of these flows move the affected individuals into the unemployment pool. For firm closures, employees also lose their jobs and move to unemployment. To simplify notation, denote the total job separation rate for workers by $s \equiv 1 - (1-\phi)^2(1-\xi)(1-\lambda_f)$, and the exit rates for firms by $\tilde{\lambda}_s \equiv \lambda_s + (1-\lambda_s)\phi$ and $\tilde{\lambda}_f \equiv \lambda_f + (1-\lambda_f)\phi$, respectively. Separations can be caused by death of either the worker or the employer, by firm shutdown, or by an exogenous match separation.

Any period, a fraction δ of individuals in the unemployment pool need to engage in casual work. I model this state as a result of a shock instead of a choice to keep the model simple. Modeling it as a choice would require introducing saving, which would substantially complicate the model. While engaged in casual work, individuals cannot search for jobs. In the following period, they return to the unemployment pool and again face the probability δ of casual work. Given its exogenous nature, income from casual work does not affect equilibrium outcomes unless it is so high that individuals would voluntarily choose it over job search. Hence, to save on notation, I assume that both the unemployed and individuals in casual work enjoy an income flow of b.

In addition to these exogenous flows, there are two key endogenous flows. As usual in such models, the job finding rate for job seekers is an equilibrium object. In addition, the entry rate into entrepreneurship, h, is endogenous. Its determination is described below.

The labor market. Job seekers and vacancies posted by employer firms intending to hire meet in a standard labor market with matching frictions. Employers posting a vacancy incur a per period cost of k_v. I assume that the number of matches per period is given by a standard Cobb-Douglas matching function. Let the number of vacancies be v. The measure of job seekers is $\bar{u} = (1-\delta)(1-h)(1-\phi)u$. Defining labor market tightness as $\theta \equiv v/\bar{u}$, the probability that a vacancy is filled in any given period is $q(\theta) \equiv A\theta^{-\mu}$, and the probability that a job seeker finds a job is θq, where μ is the exponent on vacancies in the matching function. A parameterizes the efficiency of the matching process.[18]

[18]This process describes the creation of productive matches, which then survive until destroyed at a common match destruction rate s. As usual, the process does not describe in detail how these matches are

The distribution of employment states. These flows generate a partition of individuals in the economy into the four states. I will focus on stationary equilibria of this economy. In a stationary equilibrium, the measure of agents in each state is constant. Each measure can be derived by equating flows into and out of a state. In this way, the equilibrium measures of own-account workers and employers can be obtained as

$$e_s = \frac{(1-\delta)h(1-\phi)p_s}{\tilde{\lambda}_s}u \tag{1}$$

and

$$e_f = \frac{(1-\delta)h(1-\phi)p_f}{\tilde{\lambda}_f}u, \tag{2}$$

where p_s and p_f denote the probability that an entrant chooses to become an own-account worker or an employer, respectively. These two endogenous objects are described below.

The unemployment rate in a stationary equilibrium is given by the *modified Beveridge curve* (MBC)

$$u = \frac{(1-e_f-e_s)s + e_f\tilde{\lambda}_f + e_s\tilde{\lambda}_s}{s + (1-\delta)(1-h)(1-\phi)\theta q + (1-\delta)(1-\phi)h(p_f+p_s)}. \tag{3}$$

For $\lambda_f = \lambda_s$, this simplifies to

$$u = \frac{s}{s + (1-\delta)(1-h)(1-\phi)\theta q + (1-\delta)h(1-\phi)(p_f+p_s)s/\tilde{\lambda}_f}. \tag{4}$$

This expression is analogous to the usual Beveridge curve, with two differences. First, unemployment outflows occur not only to employment (at a rate θq for searchers), but also to entrepreneurship. As a result, the job finding rate and the unemployment outflow rate are different in this economy. Second, employees and entrepreneurs have different flow rates into unemployment. This is captured in the different terms in the numerator of equation (3), and results in the final fraction in the denominator in equation (4). Intuitively, if the flow rate into unemployment is lower for entrepreneurs than for employees, then a larger entrepreneurship rate tends to reduce unemployment.

formed. That is, it is not designed to capture the high rates of turnover that may occur in the first days of a match (as documented by Blattman and Dercon (2016) for some Ethiopian manufacturing firms), and it does not exclude that successful matches are discovered, at some cost, in a high-frequency process of selection.

Finally, the measure of employees follows as

$$n = 1 - u - e_s - e_f. \tag{5}$$

Next, I describe the values and optimal behavior for firms, employees, and the unemployed.

3.2 Agents' problems, value functions, and occupational choice

Firms. All firms produce a homogeneous good that they sell in a perfectly competitive market. Firms differ in their productivity z. An entrepreneur learns about the current firms' productivity when starting the firm, and keeps that level of productivity as long as the firm is active. Given z, an entrepreneur can decide to hire employees, to become an own-account worker, or to exit to unemployment.

Employer firms produce with the production function $y = zn^\gamma, \gamma \in (0, 1)$, where y denotes the firm's output, and n denotes its employment. The parameter γ captures the degree of decreasing returns to scale in production. In this setting, optimal firm employment is an endogenous, determinate object that depends on the expected wage, labor market tightness, and on a firm's productivity. The model can thus generate employers of different sizes, which coexist with own-account workers.

Own-account workers produce with the production function $y = \zeta z$. ζ is a parameter governing relative productivity of own-account workers. It could be either smaller than one, as the self-employed have to spend some time managing their business and therefore produce less than a single employee without management duties, or larger than one, as own-account workers are not subject to the same incentive and contracting problems employers face. In addition, they may *de jure* or *de facto* be treated differently in terms of regulations and taxes. A typical presumption is that own-account workers are much less subject to regulatory oversight and taxation (see e.g. Albrecht et al. 2009).

At optimal size $n(z)$, the values of own-account work and being an employer are given by

$$F_s(z) = \zeta z + \frac{(1-\phi)(1-\lambda_s)}{1+r} F_s(z) + \frac{(1-\phi)\lambda_s}{1+r} U \tag{6}$$

$$F_f(z) = zn(z)^\gamma - wn(z) - \frac{k_v}{q}[\xi + (1-\xi)\phi]n(z) + \frac{(1-\phi)(1-\lambda_f)}{1+r} F_f(z) + \frac{(1-\phi)\lambda_f}{1+r} U \tag{7}$$

respectively. They consist in flow profits plus the expected, discounted continuation value. For own-account workers, flow profits are simply equal to output. For employers, they equal output minus the wage bill, minus the cost of rehiring workers who depart, either due to

match destruction or due to death.

Firm entry and type decision. The unemployed can decide to start a firm instead of searching for a job. Doing so involves first paying an entry cost k_f. They then draw their productivity z from a known distribution $G(z)$.[19] Based on the realization of z, they decide whether to hire workers and become an employer, whether to continue as own-account workers, or whether to return to unemployment.

The optimal choice is characterized by two thresholds, z_s and z_f. (See Figure 4.) It is clear that the value of unemployment, U, is independent of z. It is also clear from equation (6) that the value of own-account work increases linearly in productivity z. Finally, given optimal employment choices discussed below, the net value of operating an employer firm at optimal employment, net of the cost $n(z)k_v/q$ of reaching that level, is increasing and convex in z.[20] As a result, continuation values as a function of z are as depicted in Figure 4. Entrants with productivity above z_f become employers. Those with productivity below z_s exit, and those with z between z_s and z_f become own-account workers. (This structure is analogous to that in Gollin (2007).) Given a productivity distribution $G(z)$ for new entrants, this implies that new entrants exit with probability $G(z_s)$, and become employers with probability $p_f \equiv 1 - G(z_f)$. With the remaining probability p_s, they become own-account workers. The definition of p implies that the productivity distribution of employers is

$$\tilde{g}(z) = \frac{g(z)}{1 - G(z_f)}, \quad z \geq z_f, \tag{8}$$

where g is the *pdf* associated to G. There are no employers with $z < z_f$.

Combining these possibilities, the value of entry is given by

$$Q = \frac{1-\phi}{1+r}\left[-k_f + \int \max\left(F_f\left(z\right) - \frac{k_v}{q\left(\theta\right)}n\left(z\right),\ F_s\left(z\right),\ U\right)\mathrm{d}G(z)\right] \tag{9}$$

I now turn to workers and the unemployed.

[19] The assumption of uncertainty about post-entry productivity is in line with the literature on firm dynamics, and is motivated by the large rates of turnover of young firms.

[20] Convexity reflects the ability of employers to leverage their own productivity z by hiring workers accordingly. Given constant firm-level productivity and constant, linear hiring costs due to labor market frictions, it is optimal for firms to move to optimal employment directly upon entry.

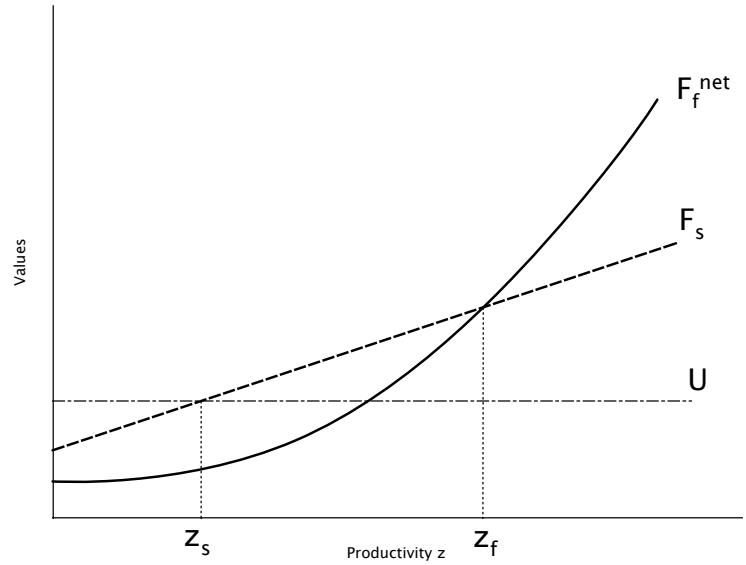


Figure 4: The values of unemployment (U), self-employment (F_s), and the value of being an employer net of hiring costs at entry ($F_f^{\text{net}}(z) = F_f(z) - n(z)k_v/q$), with associated productivity cutoffs

Workers. Employed workers receive a wage w per period. They lose their job with the combined separation probability s, and keep it otherwise. Wage determination is discussed below. Since wages are common across jobs in this economy, workers have no incentive to leave a job voluntarily. As a result, the value of employment is given by

$$W = w + \frac{1-s}{1+r}W + \frac{s-\phi}{1+r}U. \tag{10}$$

The unemployed, and occupational choice. Recall that a fraction δ of the unemployed needs to engage in casual work in any period. The remainder can choose between job search and entrepreneurial entry. Job search yields a per period flow value of b, and results in success with probability θq. As a result, the values of search, S, and that of casual employment, $\underline{U}$, are given by

$$S = b + \frac{1-\phi}{1+r}\left[\theta q W + (1-\theta q)U\right] \tag{11}$$

$$\underline{U} = b + \frac{1-\phi}{1+r}U. \tag{12}$$

With occupational choice, the value of unemployment is given by

$$U = \delta \underline{U} + (1-\delta)\max\left\{S, Q\right\}. \tag{13}$$

With probability δ, the unemployed need to engage in casual work and cannot search. With the complementary probability, they can either search, or choose to start a firm. Since workers are ex ante identical, it is clear that in an equilibrium with entry it must be true that $S = Q$. If this holds, an endogenous fraction h of the unemployed start a firm.[21] In the following, I focus on such an equilibrium.[22]

3.3 Wage determination and vacancy posting

Upon matching, a firm and a worker bargain over the wage. Like Cahuc, Marque and Wasmer (2004) and Elsby and Michaels (2013), I assume that workers and firms split the surplus from a match, with workers receiving a fixed share proportional to their bargaining weight η.[23] Wages are bargained upon hiring, and remain constant thereafter. Then it can be shown (see Appendix B.2 for a detailed derivation) that

$$w = \frac{r+\phi}{1+r}U + \frac{\eta}{1-\eta}\left[1 - \frac{(1-\phi)(1-\lambda_f)}{1+r} + \xi + (1-\xi)\phi\right] \cdot \frac{k_v}{q(\theta)}. \tag{14}$$

Two remarks are in order. First, the wage curve given by equation (14) is analogous to the wage curve in a standard DMP model, with the exception of the constants. In particular, wages increase in labor market tightness θ, reflecting the fact that match surplus is larger when the expected hiring cost k_v/q is larger. Second, self-employment opportunities enter bargaining workers' outside option U, and can affect wages in this way. Finally, although firms vary in productivity, all matches are paid the same wage. This is because upon hiring, any worker is marginal, and the relevant surplus to consider in bargaining is that of a marginal job. When firms are at their optimal employment, more productive firms have more employees, and the marginal surplus is equalized across firms. As a consequence, wages are also equalized across firms of heterogeneous productivity.

A firm's optimal employment is given by

$$n(z) = (z\gamma)^{\frac{1}{1-\gamma}} \left\{ (\eta(\gamma-1)+1) \left[\left(1 - \frac{(1-\phi)(1-\lambda_f)}{1+r} + \xi + (1-\xi)\phi\right) \frac{k_v}{q} + w \right] \right\}^{\frac{-1}{1-\gamma}}. \tag{15}$$

[21] Technically, one can think of the unemployed as following a mixed strategy that specifies starting a firm with probability h, where h is a choice object.

[22] In principle, an equilibrium with only own-account work may also arise. This could be the case if the relative productivity of own-account workers is very high. I abstract from this equilibrium for lack of empirical relevance for urban labor markets.

[23] See Stole and Zwiebel (1996) and Bruegemann, Gautier and Menzio (2015) for the game-theoretic foundations of this assumption.

Optimal firm size increases in productivity, and decreases in the cost of employing a worker, which comprises both the wage and the expected cost of replacing departing workers.

Continuing employer firms face departures of workers at a rate of $\tilde{\xi} \equiv \xi + (1-\xi)\phi$ per period, and thus need to post $\tilde{\xi} n(z)/q$ vacancies per period to replace them. New entrants find it optimal to hire $n(z)$ workers all at once, and therefore post $n(z)/q$ vacancies. From equation (2), new entrants account for a fraction $\tilde{\lambda}_f$ of employers. As a result, total vacancies in the economy are given by

$$v = \frac{\tilde{\lambda}_f + (1-\tilde{\lambda}_f)\tilde{\xi}}{q} e_f \int n(z)\tilde{g}(z)\mathrm{d}z. \tag{16}$$

3.4 Equilibrium

A stationary equilibrium consists in values $W, U, S, \underline{U}, F_f(z), F_s(z), Q$, a distribution described by u, n, e_s, e_f and $\tilde{g}(z)$, probabilities h, p_f and p_s, a function $n(z)$, and numbers v, θ, w such that

1. values $W, U, S, \underline{U}, F_f(z), F_s(z), Q$ are given by equations (6) to (7) and (9) to (13),
2. households are indifferent between occupational choices: $Q = S$,
3. wages fulfill equation (14),
4. the equilibrium distributions are generated by household choices and are stationary, according to equations (1) to (5) and (8),
5. firms post vacancies optimally (equations (15) and (16)), and
6. labor market tightness $\theta = v/[(1-\delta)(1-h)(1-\phi)u]$ is generated by unemployment in- and outflows and by firms' vacancy posting decisions.

The key equilibrium objects are θ, w, and h. The values W, U, S, $\underline{U}$, F_s, F_f and Q depend only on w and θ. Hence, the same holds for the thresholds z_s and z_f and for the probabilities p_s and p_f. Tightness and the wage also determine each firm's optimal employment $n(z)$ and the productivity distribution of employers, and hence also the average size of employer firms. The entry rate h then has to take a value such that the number of employers e_f generates a consistent value of tightness, combining equations (2), (4) and (16).

Figure 5 depicts the key equilibrium relationships, and how they determine the equilibrium values of θ, w and h. The top panel plots the wage curve and the occupational choice (OC) condition in θ, w-space. The wage curve, given by equation (14), is familiar

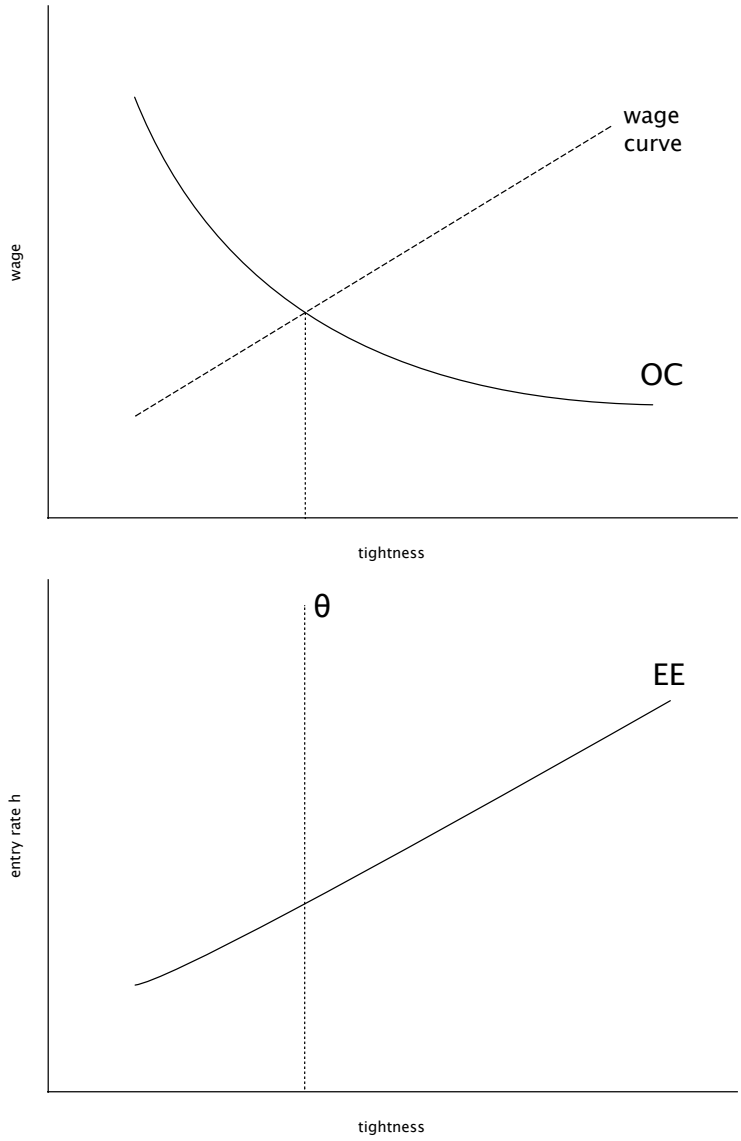


Figure 5: A sketch of equilibrium determination

from the standard DMP model. It shows that workers can bargain higher wages when the labor market is tighter. The OC curve depicts the combinations of θ and w at which the equilibrium condition $Q = S$ holds. Since the value of search S increases in both θ and w, while the value of firm entry declines in both θ and w, it is clear that this locus is negatively sloped. Neither the wage curve nor the OC condition depends on h. As a result, these two conditions on their own determine equilibrium θ and w.

The third key condition, the entrepreneurial entry (EE) condition, then determines the equilibrium entry rate h. Tightness and the wage pin down average firm size and vacancies per firm. However, the aggregate number of vacancies given in equation (16) and thus market tightness depend on the entry rate, which enters equation (16) via equations (2) and (4) (determining u). It is a consistency condition, showing the value of h required to generate equilibrium tightness. Intuitively, for a given wage, higher θ implies more costly hiring and thus smaller firms. Then many firms, and thus a high entry rate h, are needed to actually generate a high θ. This is depicted in the upward-sloping EE curve in the lower panel of Figure 5. Given θ from the upper panel, equilibrium h can be read off the EE curve in the lower panel.

Given the evidence shown in Section 2, the comparative statics I focus on are those with respect to the cost of posting vacancies. Lower vacancy posting costs raise the value of entry relative to that of unemployment, shifting OC up. They also reduce rents from matches, implying that the wage curve tilts down. As a result, tightness clearly increases, while the change in the wage is ambiguous. It can be shown that optimal firm size increases, shifting the EE curve down. Since equilibrium θ increases, the overall change in h is ambiguous.

The changes in equilibrium variables also affect entrants' continuation decisions, and the composition of the population of firms. Lower vacancy posting costs raise the value of being an employer, and higher tightness raises the value of unemployment. The value of being an own-account workers, in contrast, is only affected via the value U, which is obtained in case of an exit shock. As a result, the threshold z_f shifts down and z_s shifts up, implying an increase in the probability p_f that an entrant becomes an employer. The probability of becoming an own-account worker decreases.

It is clear that higher tightness, by increasing job finding, tends to reduce unemployment. At the same time, a decline in self-employment entry and an increase in the number of job seekers mitigates the decline in unemployment. In quantitative simulations, it is generally the case that the first effect dominates, and unemployment declines, and that self-employment also declines. To quantify all effects, I next turn to an empirically guided examination of the quantitative properties of the model.

4 Calibration

In the remaining sections of the paper, I analyze the quantitative properties of the model, and assess its ability to account for cross-country variation in unemployment and self-employment. To do so, I calibrate the model separately for eight economies at very different levels of development and with very different levels of unemployment and self-employment, ranging from Ethiopia to the United States. I then analyze which model parameters are central in driving observed variation in these labor market outcomes. This analysis suggests that differences in labor market frictions are key. To obtain a more nuanced understanding of their role and functioning, I then explore the effect of varying labor market frictions in a variety of settings. Finally, I assess the effect of varying several policy-relevant parameters in the model: the relative attractiveness of own-account work, the flow value of unemployment, and the extent of size-dependent distortions.

Can the model account for the strong variation in unemployment and self-employment across countries shown in Section 2? To verify this, I calibrate the model for eight economies at very different stages of economic development: Ethiopia, Indonesia, Mexico, Italy, France, Germany, Canada, and the US (in increasing order of GDP per capita). The choice of countries is driven by data availability. These countries essentially span the entire spectrum of country GDP per capita, with a ratio of US GDP per capita to that for Ethiopia of about 60 in 2010, for example. Rates of self-employment and unemployment also differ widely across these economies: self-employment ranges from around 9% in Germany to almost 50% in Indonesia, and the unemployment rate ranges from about 4% in Mexico to over 20% in Ethiopia. Fitting the model to this broad range of settings is a challenge, but also an opportunity: it reveal which parameters are key in driving the observed large differences in unemployment and self-employment.

Calibrating the model requires using statistics on the structure of employment, on some flows between different employment statuses, and on the firm size distribution. The choice of statistics and of calibration countries is limited by data availability. As discussed below, the nine key target moments required for the calibration are the unemployment outflow rate, the unemployment rate, the self-employment rate and fraction of own-account workers, the fraction of casual workers, the firm exit rate, the share of employment in firms with at least 10 employees, the labor income share, and the ratio of income in unemployment to the wage. The statistics that define the set of calibration countries – because their availability is most limited – are the unemployment outflow rate, information on *urban* (as opposed to country-wide) unemployment, self-employment, and own-account work, employment concentration,

and the firm exit rate.[24]

Next, I describe source for the calibration targets. To begin, there is no "easy" source giving unemployment outflow rates for a broad range of countries.[25] For the US, I take the postwar average US unemployment outflow rate from Shimer (2012). I take information on the unemployment outflow rate for Ethiopia from the 2015 Urban Employment and Unemployment Surveys (UEUS) conducted by the Ethiopian Central Statistical Agency. Data processing is described in Appendix C. For the remaining countries, I compute the unemployment outflow rate using ILO data on unemployment by duration and the method of Elsby, Hobijn and Şahin (2013).[26] Urban self-employment, own-account work, and unemployment rates are from IPUMS Censuses, using the lastest available census for each country. For Ethiopia, they are taken from the UEUS. For the US, they are computed using information from Hipple (2010). Information on the concentration of employment is from Poschke (2018) for most countries, from Berry, Rodriguez and Sandee (2002) for Indonesia, and from Bartelsman, Haltiwanger and Scarpetta (2004) for Mexico. For the US, it is computed by combining data from Hipple (2010) with information from the Statistics of US Businesses (SUSB) published by the US Census Bureau. Finally, the firm exit rate is from Bartelsman et al. (2004) for most countries, and from Bigsten et al. (2007) for Ethiopia (see also below). For Indonesia, I assume it to be identical to that for Mexico. Finally, I set the rate of casual employment by job seekers to zero for European countries, Canada, and the US, take it from the UEUS for Ethiopia, and from IPUMS Censuses for Indonesia and Mexico. I set targets for the labor income share and for b/w to common values of 0.67 and 0.4, respectively. The former is in line with levels of the labor income share documented by Gollin (2002) for a very broad range of countries. The latter essentially reflects lack of information. Direct information on job destruction rates or the length of employment relationships would help, but is not available for such a broad set of countries. The eight countries included in the

[24]For some of the developed economies, only country-level statistics are available. Since urbanization rates in these countries are very high, this is less of a concern. It is however key to use information on urban areas for the poor countries.

[25]For an ambitious attempt at studying job finding rates and development, see Donovan et al. (2017).

[26]I compute the steady state unemployment exit hazard using information on the unemployment rate and the fraction of spells of less than six months for the maximum available years for each country. (Depending on country, this spans 2003 to 2013 up to 2015 or 2016.) Unlike the US Bureau of Labor Statistics or the OECD, the ILO unfortunately does not report unemployment by duration for shorter durations, like one month. Yet, for the OECD member countries in the sample, my measures are generally very close to those computed by Elsby et al. (2013) using durations up to one month, which is to be expected if there is no or only weak duration dependence. I am therefore confident in using the ILO-based figures. For the US, where evidence for duration dependence is strong, there is a larger discrepancy, and I use the figure from Shimer (2012). Note that the use of a six-month rate implies that very short job spells will be missed. This implies that the calibrated matching function does not capture all hires, but only produces moderately durable matches that last at least six months. (See also footnote 18.)

calibration are the ones for which all these target moments are available.

As usual in such models, some parameters need to be calibrated outside the model. The model time period is set to one month. I set the interest rate such that the annual interest rate is 4%. I set the retirement probability ϕ such that the expected duration of working life is 40 years. I set μ, the exponent on unemployment in the matching function, to 0.5, and γ, the exponent on labor in the production function, to 0.85 (Atkeson and Kehoe 2005). Finally, I impose that the exogenous firm exit rates λ_f and λ_s are equal within each country.

Next, I normalize two parameters. These are the average productivity draw of an entrant and the productivity of the matching function, A. First, with homothetic preferences, the overall level of productivity in the model is not identified. I thus normalize the mean productivity draw of entrants to one. The levels of the other parameters that are in the same units, namely the standard deviation of $G(z)$, the flow value of unemployment b, and the cost levels k_f and k_v, then are to be interpreted relative to this mean productivity. Second, as is typical in search and matching models, the matching function productivity and the vacancy posting cost k_v cannot be identified separately without direct information either on the cost of hiring, or on tightness or the number of vacancies. Such information is only available for a few, rich countries. I therefore normalize A to one. This implies that differences in k_v discussed below combine the effect of differences in the vacancy posting cost and differences in the productivity of the matching function. That is, a calibrated high level of k_v could either reflect a truly high cost of posting vacancies, low efficiency of matching, or a combination of the two. Hence, the exercises analyzing the effect of varying k_v conducted in the following sections should be interpreted as varying frictions in labor markets overall, not necessarily k_v specifically.

The remaining parameters are calibrated internally to match a set of nine targets. Heuristically, one can think of a mapping of targets to parameters as follows.[27] First, the key parameters controlling flows between unemployment and employment are the per period cost of posting a vacancy k_v, and the match destruction rate ξ. Given a productivity level of the matching function, the vacancy posting cost is key for employers' hiring efforts, and thus for the unemployment outflow rate of the unemployed. Hence, I use the unemployment outflow rate as a target for k_v. This outflow rate ranges from 4.5% in Ethiopia to almost 45% in the US. Given the unemployment outflow rate and moments on entrepreneurship, the level of the unemployment rate identifies the job destruction rate ξ.

A second set of moments relates to self-employment and entrepreneurship. Here, I set

[27] Of course, it is actually the case that targets have to be matched jointly by setting all seven parameters, and cannot be matched individually one by one. Nevertheless, each parameter clearly affects some targets more strongly than others.

the parameters k_f, ζ, λ_f and σ_z to match the self-employment rate, the fraction of own-account workers, the firm exit rate, and the share of employment in large firms. Clearly, higher fixed entry costs k_f discourage entrepreneurship, and thus affect the overall level of entrepreneurship (own-account workers plus employers). The parameter ζ controls the relative productivity of own-account workers. Higher ζ thus leads to a higher level of own-account work given an overall level of entrepreneurship. The fraction of employers is around 4-5% of employment in almost all countries, and is slightly lower in poorer countries.[28] Own-account workers account for the remainder of the self-employed. Their fraction of employment ranges from 4% in Germany to 45% in Indonesia, in line with the broad variation in self-employment rates. The mapping between the exogenous firm exit rate in the model, λ_f, and the data exit rate is immediate. Exit rates from Bartelsman et al. (2004) range from 5% per year in Germany to 14% in Mexico. Finally, since most firms in the model are (very) small, a higher dispersion of the productivity draws of entrants, generated by higher σ_z, generates more employment in large firms. The share of employment in firms with at least 10 employees lies between 80 and 90% in rich countries. Employment is less concentrated in the poorer countries.[29]

Three further moments are closely related. Conditional on the unemployment rate, the rate of casual employment in an economy identifies δ. The labor income share is informative about workers' bargaining power η. To pin down the flow value of unemployment, b, I set b/w to 0.4 in all economies (see the discussion above).

As a benchmark for the analysis below, I also calibrate the model to an average economy, described by average values of all target moments. For the few statistics that are not consistently available for all countries, like the share of employment in firms with at least 10 employees, I take the average using actual data where available, and model-predicted data from the country calibration for those countries where data availability forced us to use a slightly different, related moment in the country calibration.

To save space, I do not report all calibration results and parameters in the main text – see Table 27 for these. Here, I discuss the calibration for the most extreme case, Ethiopia, in some detail, and then compare it to the calibrations for the other extreme, the US, and for the average economy.

[28] I take this target from IPUMS Census data. For Ethiopia, there is a large discrepancy between the Census figure and that from the UEUS, which contains more detailed information on firm employment, so I target the average of the two values.

[29] For Mexico and Indonesia, I have information on the share of employment in firms with at least 20 employees. (76% and 33%, respectively, from Bartelsman et al. (2004) and Berry et al. (2002).) For Ethiopia, I target the share of employer firms with less than 10 persons engaged, which is 87% in UEUS data.

Table 6: Calibration: model and data moments (Ethiopia)

	model	data
Targeted moments:		
Unemployment outflow rate	0.044	0.045
Unemployment rate	0.237	0.237
Casual employment	0.245	0.236
Fraction own-account workers	0.288	0.29
Fraction employers	0.05	0.048
4-year entrepreneurship persistence	0.582	0.54
Share firms with $n \leq 10$	0.871	0.874
Labor income share	0.67	0.67
b/w	0.4	0.4
Not targeted:		
UN ratio	0.320	0.308
Entry rate h	0.0138	
Job finding rate	0.063	
Total job separation rate	0.046	
Annual firm exit rate	0.142	
Mean firm employment	2.2	
Mean employment (employers)	7.3	
Share of employment in firms with $n > 10$	0.089	
Mean SE income/w	1.1	
Mean employer income/w	5.1	
Business income/Y	0.656	
Own-account income/Y	0.250	

Table 7: Four-year transition matrix between the states of entrepreneurship, employment and unemployment (Ethiopia). Data values in parentheses.

	e'	n'	u'
e	0.582	0.114	0.208
	(0.538)	(0.107)	(0.221)
n	0.101	0.387	0.417
	(0.065)	(0.597)	(0.219)
u	0.152	0.343	0.410
	(0.068)	(0.261)	(0.528)

Source: Bigsten et al. (2007). Remaining probability is retirement/transition out of the labor force.

Table 6 shows the model fit for Ethiopia. It is overall very close. The table also shows model predictions for some non-targeted moments. For the ones shown in Table 6, no direct data counterparts are available, but their orders of magnitude are still instructive. First, the entrepreneurial entry rate from unemployment is 1.5% per month, whereas the job finding rate for searchers is 6%. This implies that about one fifth of the outflows from unemployment are due to entry into self-employment. (Note that for Ethiopia, the overall unemployment outflow rate is below the job finding rate since unemployed workers engaging in casual work cannot search.) The mean size of employer firms is 7, in line with UEUS data and much below mean firm sizes in rich economies. Due to the high self-employment rate, the fraction of business income (income of own-account workers plus employer profits) in aggregate output is 65%, and that of own-account workers is 25%.

Table 7 compares model predictions for flows across the states of entrepreneurship, employment and unemployment to data for the period from 2000 to 2004. The data matrix is adapted from Bigsten et al. (2007); see Appendix C for details. Unfortunately, no more recent flow matrix is available. In addition, the available data combine own-account workers and employers in one group. Only the top left element of the matrix, showing persistence in entrepreneurship, is targeted in the calibration. In spite of this, model and data transitions overall have similar orders of magnitude. In particular the transitions out of entrepreneurship to both unemployment and employment are replicated very closely by the model, despite the fact that the latter can only occur indirectly in the model (via unemployment). In contrast, the model overstates entry rates into entrepreneurship, from both employment and unemployment, overstates employment to unemployment transitions, and understates unemployment persistence. This is due to the fact that the transition matrix is for the years 2000 to 2004, a period when the Ethiopian economy was significantly poorer. More specifically, it reflects the fact that the ergodic distribution over entrepreneurship, employment and unemployment implied by the data transition matrix is $[0.13, 0.55, 0.32]$, i.e. it implies much less entrepreneurship and higher unemployment than what is observed in more recent data. As a result, it is necessarily the case that when the model is calibrated to match recently observed entrepreneurship and unemployment rates (which are higher and lower, respectively), it will generate more entrepreneurship entry, larger unemployment outflows, and a lower persistence of unemployment than found in the data a decade earlier.

Table 8 shows the parameters generated by the calibration exercise. This reveals why the unemployment rate is so high in Ethiopia: the combination of a job finding rate that is low by global standards (6%, close to continental European levels) with a job destruction rate that is high by global standards (3.2%, close to US levels) results in a high level of

Table 8: Calibration: parameter values (Ethiopia)

externally calibrated:		
r	discount rate (annualized)	0.04
ϕ	retirement probability (annualized)	1/40
μ	matching function	0.50
γ	decreasing returns to scale	0.85
internally calibrated:		
k_f	entry cost	13.54
k_v	vacancy posting cost	69
η	worker bargaining power	0.432
b	utility flow of unemployment	0.188
λ_f, λ_s	firm exit rates (annualized)	0.12
ξ	separation rate	0.032
σ_z	productivity distribution	0.0224
ζ	relative own-account productivity	0.519
δ	casual job probability	0.440

unemployment.

Overall, the model clearly replicates key features of the Ethiopian economy: high rates of unemployment, self-employment and casual work, and a preponderance of small or tiny firms. Table 9 compares calibration results for Ethiopia to those for the US and for the average economy. Target moments are not shown, since they are almost identical to model moments. (See Table 26 for details.) This table shows how the model is able to replicate the vastly different structures of the three calibrated economies. It also supports the arguments about how model moments identify parameters made above.

The table shows a subset of five parameters, to stress five salient differences across the calibrations. First, vacancy posting costs relative to productivity are very high in Ethiopia, and very low in the US. This is the first key reason for the high unemployment rate in Ethiopia. Second, the job destruction rate is high in Ethiopia relative to the US. This is the second key reason for the high unemployment rate in Ethiopia. Third, the entry cost is low in Ethiopia, and high in the US. On the face of it, this is the key reason for the high self-employment rate in Ethiopia. (Results below will show that labor market frictions, parameterized by k_v, also play a large role.) Fourth, the relative productivity of own-account workers, ζ, is low in Ethiopia. This indicates that the fraction of own-account workers in Ethiopia is high not because this state is very attractive here compared to other countries, but despite its low attractiveness. Finally, the dispersion of productivity in Ethiopia is tiny relative to the other countries. This is what is required to generate a small share of

Table 9: Comparing calibrations – highlights

country:	Ethiopia	USA	average
Model moments:			
Unemployment outflow rate	0.044	0.453	0.180
Unemployment rate	0.237	0.051	0.106
Self-employment rate	0.348	0.098	0.193
Fraction own-account workers	0.288	0.050	0.149
Fraction employers	0.05	0.048	0.044
Share of employment in firms with $n > 10$	0.089	0.848	0.740
Parameter values:			
Vacancy posting cost k_v	69	12	45.4
Firm entry cost k_f	13.54	56	7.5
Job destruction rate ξ	0.032	0.0136	0.0143
Productivity dispersion σ_z	0.0224	0.164	0.32
Relative own-account productivity ζ	0.519	0.657	0.605

The top panel shows model moments for three calibrations: the ones targeting Ethiopia and the US, respectively, and that targeting average values of data moments. The model moments shown here are generally close to the targeted data moments. (See Table 26 for details.)

employment in large firms. It should be noted that size-dependent distortions (SDDs) – i.e., a burden of taxes, regulation, or other costs or frictions that increases in firm size – could generate a similar outcome.[30]

How does the model stack up compared to dimensions of the data that were not directly targeted in the calibration? This comparison can be made for the total job separation rate, which can be compared to separation rates computed from ILO data or those reported in Elsby et al. (2013). The latter source allows comparing unemployment inflow rates for the five countries in the set of calibration countries that are OECD members. The correlation between model-implied separation rates and empirical ones is above 0.9. Some differences arise due to differences between the sample period for the data used by Elsby et al. (2013) and that used for the calibration targets here.

The fact that the model can be calibrated to a set of very different countries shows its versatility. In the following sections, I use it to analyze quantitatively the determinants of

[30] All country calibrations assume that there are no SDDs, and let σ_z be country-specific. An alternative approach would be to assume that σ_z is common, and that SDDs are country-specific. What both approaches have in common is that they can only identify variation in one of the two dimensions, productivity variation or SDDs, and not both at the same time.

cross-country differences in labor market outcomes.

5 Which factors drive cross-country differences in unemployment and self-employment?

It is clear from the figures shown in Section 2 that unemployment and self-employment rates vary very strongly across countries. Which factors account for this variation? To answer this question, I conduct the following exercise. Starting from the average country calibration, I recalibrate the model for each country, keeping all parameters as in the calibration for the average target, except for one or a combination of few parameters. That is, I separately find which values of k_f, k_v, etc., give the model the best fit to the country-specific calibration targets for Ethiopia, the US, etc., when all remaining parameters are as in the "average" calibration. I then compute the share of variation in outcomes of interest in the data that the model can explain in these different scenarios. The question is: how much of the variation can be explained by optimally varying just a single parameter, or a small set of parameters?

Results for this exercise are shown in Table 10. The first column shows the fit of the model when one, two, or three parameters are country-specific. The fit is computed as one minus the ratio of the sum of the calibration loss statistic across countries when some parameters are country specific to the value of the statistic when all parameters are common. It ranges between zero and one, and a larger number describes a better fit. A number of one indicates that varying a limited number of parameters fully explains the variation in the data, while a number of zero indicates that doing so is no better than using the parameters for the average economy for all countries. By construction, letting all internally calibrated parameters adjust would allow the model to fit all countries perfectly, implying a statistic of one. Subsequent columns show the model's explanatory power in terms of individual variables. These statistics are computed as one minus the ratio of the sum of squared deviations between model and data values with country-specific parameters to the sum of squared deviations with common parameters. (This statistic is akin to a coefficient of determination, or R^2.)[31]

It is very clear from these results that for the overall fit of the model, variation in k_v is key. Letting k_v adjust to allow the model to fit the calibration targets for each country as closely as possible results in a reduction of the calibration loss function by almost half compared to

[31] As also discussed by Asker, Collard-Wexler and De Loecker (2014), this measure is akin to the uncentered R^2 in regression analysis. By definition, the measure cannot exceed one. (Nothing prevents it from being negative here.)

the case with common parameters (drawn from the average target calibration). Additionally allowing for the job destruction rate ξ to be country-specific results in a reduction in the loss by half again. Variation in only these two parameters can thus account for almost three quarters of the variation in calibration targets in the data. Finally, also allowing the utility flow in unemployment parameter b to be country-specific reduces the loss by more than half again, bringing it to one tenth of its value with common parameters.

In terms of individual outcome variables, the combination of k_v and ξ is also very powerful. Together, they explain almost the entire variation in the unemployment outflow rate, and a third of the variation in the UN ratio. They also explain 80% of the variation in the self-employment rate. Further allowing b to be country-specific allows the model to explain 90% or more of the variation in both the unemployment-related variables and in self-employment.[32]

For some individual outcome variables, other parameters have more explanatory power. For example, allowing for only country-specific ζ explains more than 90% of the variation in the self-employment rate. However, this scenario worsens the model's fit in terms of the unemployment rate compared to the situation with common parameters for all countries. The reason is that while high ζ implies high self-employment, it also reduces unemployment, generating a correlation between self-employment and unemployment that runs counter to the data. The same occurs for country-specific entry costs k_v.

Overall, these results suggest that cross-country variation in parameters encapsulating labor market frictions is key for understanding variation in labor market outcomes across countries. This is the case not only for the unemployment outflow rate (which is directly affected by the vacancy posting cost k_v) and the unemployment rate and the UN ratio (which are directly affected by k_v and by the job destruction rate ξ), but also for the self-employment rate.

Before turning to a more detailed analysis of the effect of labor market frictions in the model, I investigate whether, beyond the dispersion in unemployment and self-employment rates in the data, they can also account for the relationship between the UN ratio and self-employment shown in Section 2.

Figure 6 depicts the relationship between self-employment and the UN ratio in model and data. It shows model results for two cases: one ("2 specific parameters") where k_v and ξ are country-specific and chosen to best fit each country's set of calibration targets, and

[32]There are some particularities. For example, while country-specific k_v leads to a large improvement in the fit across countries, the effect is particularly pronounced in Ethiopia, the US and Mexico (not shown in the table). In France, in contrast, the improvement in fit from changing k_f is largest.

Table 10: Explanatory power of the model when only a subset of parameters is country-specific

Outcome:	Fit of all calibration targets	unemployment outflow rate	unemployment rate	UN ratio	self-employment rate
One country-specific parameter:					
k_f	0.173	0.099	-0.075	0.143	0.701
k_v	0.438	0.715	0.306	0.370	0.105
η	0.118	0.209	0.213	0.117	-0.141
b	0.124	0.167	0.003	-0.013	0.224
λ_f	0.065	0.001	0.100	0.202	0.315
ξ	0.190	0.021	0.284	0.413	0.883
σ_z	0.159	0.079	0.019	0.204	0.591
ζ	0.138	-0.017	-0.113	0.003	0.915
Two country-specific parameters:					
k_v, ξ	0.708	0.939	0.191	0.336	0.808
Three country-specific parameters:					
k_v, b, ξ	0.915	0.987	0.984	0.988	0.890

Notes: The first column reports the reduction in the sum of the calibration loss statistic for all eight countries when one, two or three parameters are chosen to minimize each country's loss function, relative to the sum of loss statistics when the parameters for the calibration for the average target are used in all countries. The subsequent columns report the reduction in the sum of squared deviations between model predictions and data values for the indicated statistics. These numbers are akin to the R^2 of the model for these outcome variables.

Table 11: Bivariate regression coefficients, model and data

Dependent variable:	self-employment rate	UN ratio	self-employment rate
Independent variable:	GDP per capita	GDP per capita	UN ratio
data:			
all countries	-0.132	-0.035	1.177
calibration countries	-0.100	-0.029	1.925
model with country-specific values of ...			
k_v	-0.043	-0.001	1.484
k_v and ξ	-0.051	-0.006	0.881
k_v, ξ and b	-0.052	-0.041	0.374

Notes: Data numbers are from Table 1 and the regression underlying the line of fit shown in Figure 3. They are for 58 (54 in column 3) observations, and statistically significant at least at the 5% level. Model numbers are from specifications allowing a limited number of parameters (indicated in the table) to vary across countries. They are from a regression using the eight simulated countries. The figures in the last two rows correspond to the model regression lines shown in Figure 6.

one ("3 specific parameters") where in addition, b is also country-specific. It shows the data as small dots, data for the eight countries used in the calibration as triangles, and model outcomes for three (two) country-specific parameters as black squares (grey diamonds). (The fit of the country calibrations and the model explanatory power for each individual variable separately thus is given in the two bottom rows in Table 10.) The solid and dashed lines in each figure show best fits of a linear regression of the variable on the vertical axis on that on the horizontal axis. The regression coefficients underlying these lines are reported in Table 11.

It is immediately clear from the lines of best fit that the model outcomes are qualitatively in line with the data. Quantitatively, the relationship between the self-employment rate and the UN ratio in the model is about one third as strong as it is in the data.

With variation in variables capturing labor market frictions only, the model thus does an excellent job in reproducing not only each country's levels of self-employment and the UN ratio individually – implying variation in the self-employment rate across countries of more than 20 percentage points –, but also the bivariate relationship of these two variables. This suggests that variation in labor market frictions across countries is not only a driver of differences in unemployment, but also in other labor market outcomes, in particular self-employment. The only driver required for this is strong variation in labor market frictions, in line with empirically observed variation in unemployment rates.

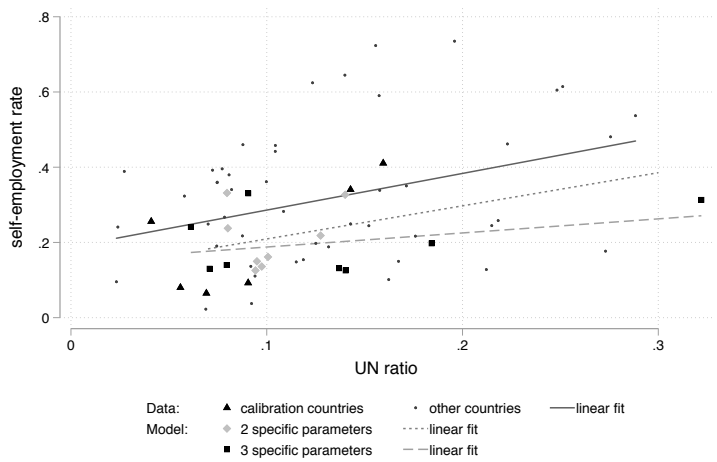


Figure 6: Self-employment and the UN ratio: data and model outcomes

Notes: Points labelled "model" show model outcomes with parameters from the calibration for the average economy, except for k_v and ξ, which are country-specific (series labelled "2 specific parameters"). In the series labelled "3 specific parameters", the parameters b is also country-specific.

Finally, note that the analysis here does not allow for a direct effect of the level of development on occupational choice and unemployment. Previous work has suggested that this effect may be important (see in particular Gollin (2007) and Feng et al. (2017)). This merits some discussion. First, in Gollin (2007), development affects self-employment choices through the following channel: when capital and labor are complements in production, higher aggregate TFP, by stimulating capital accumulation, raises wages relative to profits, leading marginal entrepreneurs to exit. A similar effect could be incorporated into the present analysis. However, it turns out that the economic intuition just given does not go through in a setting with bargaining. In a standard twist, the fact that firms share the marginal product of each job with workers discourages investment. Higher aggregate TFP still raises the capital-labor ratio, and thus the labor income share, in this setting. But because underinvestment by firms implies elevated levels of the marginal and average product of capital, it reduces the capital income share, and raises the profit share. (See Bauducco and Janiak (2018) for a related result.) As a result, higher aggregate productivity can actually lead to increased self-employment in a model with bargaining, even when capital and labor are complements. Given these complications, and the strong effect of labor market frictions just shown, I focus on the latter and leave the effect of aggregate TFP to future analysis.

Second, Feng et al. (2017) assume that richer countries are characterized by higher productivity of employer firms, but not of own-account workers. An increase in the productivity

of employers attracts workers into job search, increasing unemployment and reducing own-account work. Increasing ζ in the model analyzed here would have a similar effect. And indeed, results given in Table 10 show that variation in ζ alone could explain almost the entire variation in self-employment among the calibration countries. However, in the urban sample I am focussing on, its prediction for unemployment are counterfactual. (They are however, in line with the findings of Feng et al. (2017) at the country level; see the discussion in Section 2.)

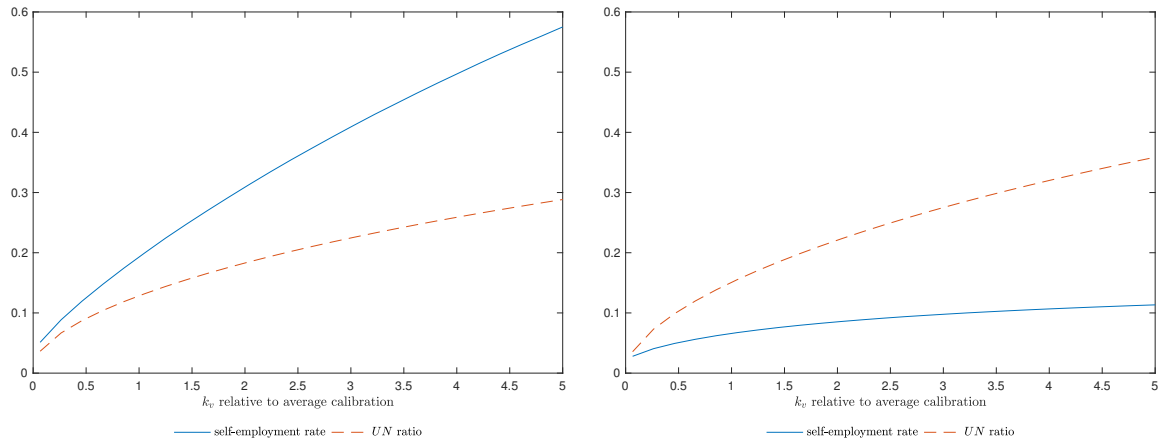
6 Labor market frictions, self-employment, and productivity

Having shown the importance of labor market frictions in accounting for cross-country differences in labor market outcomes, I next illustrate their effects in more detail. I focus on the effects of hiring costs k_v, since they are the individual parameter with the greatest explanatory power. How do labor market frictions affect occupational choices and aggregate outcomes?

Lower hiring costs make running a business more profitable, and thus attractive. (OC shifts up in the top panel of Figure 5.) Lower costs of creating a match also reduce match surplus, shifting the wage curve down. The net effect is higher tightness, and an ambiguous change in the wage.

Figure 7 shows the effect of changes in k_v on the self-employment rate and the UN ratio. It is clear from Figure 7a that lower k_v not only leads to a lower UN ratio – this is as expected in a standard DMP model – but, by making job search more attractive, also reduces the self-employment rate. The second effect is sizeable: at the average country calibration, the self-employment rate declines more than the unemployment rate for a given change in k_v. For example, reducing k_v by half from its value in the calibration for the average country results in a reduction in the UN ratio by 3.8 and the self-employment rate by 6.8 percentage points.

Which margin reacts more strongly depends on parameters, in particular the the cost of establishing new firms, as is clear from comparing the two panels of Figure 7. When entry is costly (k_f is high), lower k_v prompts fewer entrepreneurs to go for search instead. But when entry is cheap, the self-employment rate can vary very strongly with k_v. Figure 7 shows this for a low value of k_f, corresponding to the one from the calibration for the average target, and a high value, the one from the calibration for the US. For example, for a country as in the average calibration but with the (high) level of the entry cost of the US, the self-employment



(a) Low k_f (from average country calibration)

(b) High k_f (from US calibration)

Figure 7: The effect of vacancy posting costs k_v on labor market outcomes for different levels of the entry cost k_f (benchmark: average country)

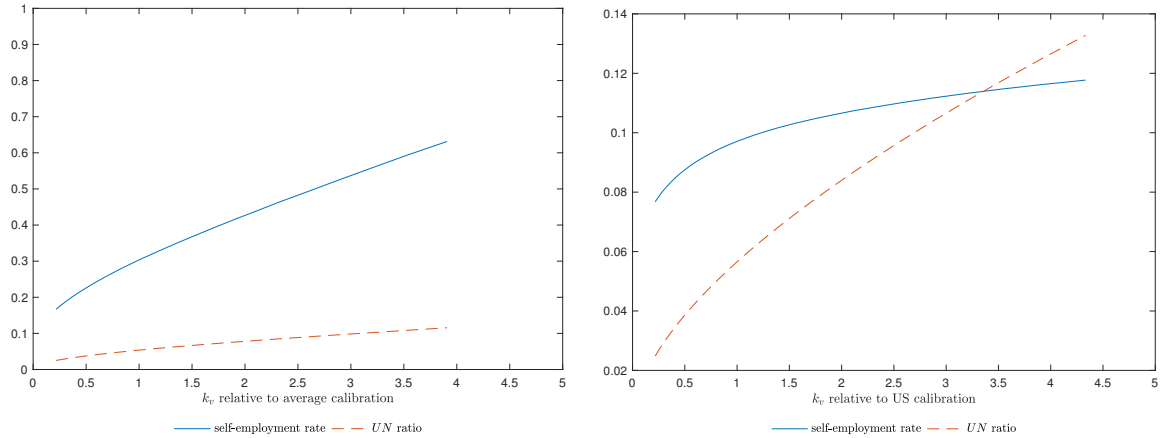
Notes: All parameters except k_v and k_f as in the calibration to the average target (see Tables 9 and 27 for parameter values). k_f as in the calibration for average targets in the left panel, and as in the calibration for the US in the right panel.

rate only falls by 1.5 percentage points as k_v is reduced by half.

Results are very similar when beginning from the calibration for the US, and not for a synthetic country with average target values. Since this calibration involves fairly high fixed costs, it implies a moderate reaction of the self-employment rate to labor market frictions, in particular when compared to the induced changes in the UN ratio (Figure 8b). Figure 8a shows the role of entry costs in this: a country that is like the US, but with a much lower value for the entry cost (taken from the calibration for Ethiopia), not only has a higher level of self-employment at the US value of k_v, but also experiences much larger changes in self-employment in reaction to changes in hiring costs.

Table 12 gives more detailed information on how these changes come about, for several different calibrations. Lower vacancy posting costs induce employer firms to post more vacancies, driving up labor market tightness. As in a standard DMP model, this results in higher job finding and unemployment outflow rates, higher wages, and a lower UN ratio.

Self-employment choices also change. First of all, despite the reactions of wage and tightness, lower k_v still implies a lower user cost of labor for employer firms, so that average firm size grows. This also prompts a larger fraction of entrants to become employers (except in the US calibration). As a consequence, the new equilibrium features slightly more, larger employer firms, and significantly fewer own-account workers. The fraction of own-account



(a) Low k_f (b) High k_f

Figure 8: The effect of vacancy posting costs k_v on labor market outcomes for different levels of the entry cost k_f (benchmark: US)

Notes: All parameters except k_v and k_f as in the calibration to the US (see Tables 9 and 27 for parameter values). k_f as in the calibration for Ethiopia in the left panel, and as in the calibration for the US in the right panel. Note different scales of the vertical axes.

workers declines partly because more entrants decide to become employers, but even more because the outside option of search becomes more valuable – so much so that the lowest level of productivity at which own-account work is optimal, z_s, increases by 8% in the calibration to the average target.

The reduction in vacancy posting costs leads to an increase in aggregate output. This effect is shown in Table 13. To understand its sources, I show the effect of lower k_v on output for four different model calibrations, as in the previous table. Aggregate output gains range from 1.4 to 6 percent. Changes in output can stem from the increase in employment, changes in wages and profits due to lower k_v, and the changes in firm size and occupational choice induced by lower k_v. The relative importance of these channels is illustrated by the decomposition of output gains in the bottom rows of the table. This shows two main results. First, output gains are entirely due to changes in the amount and composition of employment, and not due to output gains within groups, which are close to zero throughout. The reason for this is that while lower hiring costs lead to higher wages, they also entice new, lower-productivity employers to enter, implying that average firm output does not rise. Second, both lower unemployment and changes in self-employment propensities and composition drive overall output gains. Their relative importance differs across economies.

The four calibrations for which output effects are shown in Table 13 differ mainly in

Table 12: The effect of labor market frictions

change in	calibration to			
	average economy	average economy, high k_f	Ethiopia	US
tightness θ (%)	129.1	140.4	173.8	126.0
UN ratio (% pts)	-3.8	-4.8	-7.1	-1.8
u (% pts)	-2.6	-4.4	-1.0	-1.6
u outflow rate (% pts)	8.2	8.0	1.8	21.3
job finding rate θq (% pts)	9.7	8.1	4.1	21.6
fraction employees (% pts)	8.4	5.5	19.1	2.4
entrepreneurship (% pts)	-6.8	-1.5	-23.0	-1.0
own-account (% pts)	-7.0	-1.5	-25.3	-0.9
employers (% pts)	0.2	0.0	2.3	-0.1
mean firm n (%)	54.4	29.8	217.2	11.0
incomes:				
w (%)	5.6	6.3	3.1	6.4
SE/w (%)	-1.2	-2.0	-4.8	-4.7
employer/w (%)	-13.2	-13.9	-6.3	-13.8

Notes: The table shows the reaction of the model economy to a reduction in vacancy posting costs by half. Parameters for the respective benchmarks are given in Tables 9 and 27. In the second column, k_f takes on the value from the calibration for the US, as also seen in Figure 7b.

Table 13: The output effect of lower labor market frictions

% change in	calibration to			
	average economy	average economy, high k_f	Ethiopia	US
output:				
aggregate output	4.0	5.2	6.1	1.4
aggregate output net of k_v	7.7	10.0	2.3	6.1
aggregate output net of k_v and k_f	10.6	9.3	9.3	9.7
output of employer firms/employee	-1.5	-0.5	-13.1	-0.7
counterfactual output:				
group sizes as in benchmark	0.0	0.1	-0.1	0.0
only u changes	2.9	5.1	0.6	1.7
only self-employment rates change	1.2	0.0	5.2	-0.4
all group sizes change (average group output as in benchmark)	4.1	5.1	5.9	1.3

Notes: The table shows the reaction of a set of model economies to a reduction in vacancy posting costs by half. Parameters for the respective benchmarks are given in Tables 9 and 27. In the second column, k_f takes on the value from the calibration for the US, as also seen in Figure 7b. The last four rows of the table show counterfactual results. In these rows, "group" refers to the three groups of employees, own-account workers and employers. In the first of the four rows, counterfactual aggregate output is computed using group sizes from the benchmark, but average group output from the low-k_v economy (including spending on hiring). In the remaining rows, average output for each group is taken from the benchmark. In the second of the four rows, relative group sizes are as in the benchmark, but the unemployment rate is taken from the low-k_v economy. In the next row, the unemployment rate is taken from the benchmark, but relative group sizes (fractions of own-account workers and employers among those in work) from the low-k_v economy. In the final row, all group sizes are taken from the low-k_v economy.

their levels of k_v and k_f. The output changes and their sources reflect these differences. In economies with high entry costs, essentially the entire output gains come from lower unemployment. This is natural, given the small changes in self-employment in these economies, shown in Table 12. But in economies with low entry costs and high self-employment, changes in the self-employment rate can account for a third (average economy) up to almost the entirety (Ethiopia) of overall output gains. This is due to the large reduction in the rate of own-account work in response to lower k_v in these economies, combined with their relatively high output of employees relative to the self-employed.

To summarize, the model not only predicts a strong effect of labor market frictions on unemployment and self-employment, but also a strong effect on output. A substantial part of that comes from the effect of labor market frictions on occupational choices. This effects is particularly large in economies with strong labor market frictions and low entry costs.

7 The value of unemployment, regulation, and self-employment

The high rate of self-employment in poor countries may be affected by many factors other than labor market frictions. Therefore, I use the model to assess the effect of three such factors on self-employment, unemployment, and output: the flow value of unemployment (which is sensitive to policy), the relative productivity of self-employment (which reflects, among other things, the differential enforcement of regulations across different groups of firms), and the effect of distortions. Here, I consider both linear and size-dependent distortions, which may play an important role in the low productivity of poor countries (see e.g. Guner, Ventura and Xu 2008, Restuccia and Rogerson 2008, Hsieh and Klenow 2009).

Table 14 shows the effect of a higher flow value of unemployment, b, on occupational choices, labor market outcomes, and output. The table shows the effect of increasing b by 20% in the "average economy", in the Ethiopia calibration, and in an economy that is like the average economy, but has the higher job destruction rate of Ethiopia. For each economy, two columns are shown. The first one shows outcomes for the high-b equilibrium relative to the respective benchmark. The second one shows outcomes for high b, keeping the firm entry rate as in the benchmark. This column essentially captures what the effect of changing b would be if there was no response to it in terms of occupational choices.

It is immediately clear from the table that, unsurprisingly, higher b reduces labor market tightness and raises wages and unemployment. Note that the decline in tightness is not

Table 14: The effect of changes in b

change in	average economy		average economy, high ξ		Ethiopia	
	higher b	higher b, fixed OC	higher b	higher b, fixed OC	higher b	higher b, fixed OC
tightness (%)	-11.8	-10.0	-13.7	-10.4	-16.3	-11.8
UN ratio (% pts)	0.6	0.6	1.1	0.9	1.1	0.9
u rate (% pts)	0.8	0.7	1.8	1.2	6.2	0.5
entrepreneurship (% pts)	-1.9	-1.3	-4.8	-2.5	-20.1	0.9
own-account work (% pts)	-2.0	-1.4	-5.0	-2.7	-21.2	1.1
fraction employers (% pts)	0.1	0.1	0.2	0.2	1.1	-0.2
incomes:						
w (%)	1.3	2.4	1.7	3.4	1.0	2.9
SE/w (%)	-0.3	-0.3	-0.2	-0.4	-2.0	-2.8
employer/w (%)	-3.3	-5.4	-2.6	-6.2	-2.6	-5.9
aggregate output (%)	-0.5	0.3	-0.7	0.7	-2.8	-0.1
... net of k_v (%)	0.5	1.3	-0.1	1.6	-4.2	1.4
... net of k_v and k_f (%)	1.2	1.4	0.7	1.2	2.5	1.8
counterfactual output:						
only effect of higher u (%)	-0.9	-0.8	-2.1	-1.4	-4.0	-0.3

Note: Each column shows the effect of a single change relative to the benchmark, the calibration to the average economy. The columns labelled "higher b" shows the effect of raising b by 20%. The columns labelled "higher b, fixed OC" shows the effect of raising b by 20%, keeping the firm entry rate as in the benchmark. Changes are shown either in in percent or in percentage points, as indicated.

only due to less vacancy posting by firms, but also due to an increase in the number of job searchers. Higher b also affects occupational choices: as higher b makes job search more attractive and reduces incomes of employers, fewer individuals choose own-account work.

In the average economy, output net of entry costs and hiring costs actually increases very slightly. This indicates that the opportunity of self-employment leads to an inefficiently large number of entrants and small pool of job seekers in the benchmark.[33] The output loss from higher unemployment offsets these gains. Output gross of these costs declines slightly.[34]

[33]Recall however that the increase in b here is purely for illustrative purposes. These results should not be used to draw welfare conclusions, since I have abstracted from the source of the increase in b (e.g. tax financing).

[34]These findings are similar in flavor to those in Galindo da Fonseca (2018), who analyzes the effects of a policy of subsidies to startups.

All changes are more pronounced in the Ethiopia calibration. Here, the effect of changes in occupational choice is particularly salient: while the entrepreneurship rate rises slightly when the entry rate is fixed (this occurs because of the larger unemployment pool), it drops precipitously when optimal occupational choice is allowed for. This occurs as the weakest own-account workers switch to job search. This scenario also highlights clearly the difference between the unemployment rate and the UN ratio: The switch of own-account workers to job search hardly affects the UN ratio (it does not change frictions), but it leads to a large increase in the unemployment rate.

In this calibration, the output effects are also larger: The increase in the pool of job seekers leads to a substantial reduction in output. This is somewhat offset by the movement of own-account workers into more productive wage employment.

This exercise illustrates clearly how important it is to take changes in occupational choice in response to changes in policy or in the environment into account: an assessment that ignored the switch of own-account workers to job search would be completely off, not only in terms of the direct effect of the change on self-employment, but also in terms of its effects on unemployment and output.

Just like changes in the attractiveness of unemployment, changes in the attractiveness of self-employment also affect occupational choices. One way of thinking of these is that for example stricter enforcement of regulations for the self-employed could lead to a reduction in ζ. Table 15 shows the reaction of occupational choices to such a change.

Of course, lower attractiveness of own-account work will imply more job search, thus higher unemployment and, ultimately, higher employment. However, the strength of these changes depends on parameters, as can be seen by comparing columns. First, the reduction in the self-employment rate is larger the lower the entry cost.[35] Second, where the people leaving self-employment end up depends on labor market frictions. While most of them enter employment in the average economy calibration, the exact fraction varies with labor market frictions (87% in the benchmark versus 84% when hiring costs are slightly higher). Third, in an extreme calibration like that for Ethiopia, most of those leaving self-employment end up in either unemployment or casual employment, and not wage employment. This analysis again shows that in assessing the effect of changes in policy or the environment, it is important to take changes in occupational choice, and particularly movements between self-employment and job search, into account.

Finally, I assess the effect of distortions on occupational choices and outputs. A broad

[35]The reduction is also larger when k_v is higher, but it occurs from a higher base. The relative reduction in own-account work is very similar in the first two columns.

Table 15: The effect of lower ζ on occupational choices

	average economy			Ethiopia
change in	benchmark	with higher k_v	higher k_f	
UN ratio	0.0	0.0	0.0	0.1
u rate	0.5	0.7	0.1	5.3
fraction of				
employees	3.3	3.8	0.6	6.9
self-employed	-3.8	-4.5	-0.8	-18.2
own-account workers	-4.0	-4.8	-0.9	-19.4
employers	0.3	0.3	0.1	1.2

Note: Each column shows the effect of a reduction in ζ by five percent relative to the respective benchmark, in percentage points. The column labelled "higher k_v" uses the value for k_v for Ethiopia instead of the (lower) one of the average economy calibration. The column labelled "higher k_f" uses the value for k_f for the US instead of the (lower) one of the average economy calibration.

recent literature has documented distortions in firms' input choices, in particular in poor countries, and has argued that they contribute substantially to the low aggregate total factor productivity (TFP) of their economies. Particularly important are size- or productivity-dependent distortions: if input choices of highly productive firms are particularly far from optimal, the effect on aggregate output is going to be much larger than if all firms are distorted in a similar way.

In assessing the effect of distortions in the model studied here, I am not only interested in their effect on occupational choices and output, but also want to analyze whether the importance of labor market frictions found in the decomposition exercise could instead be attributed to distortions. To this end, Table 16 shows results from two experiments: a linear tax on revenue, and a size-dependent distortion, where the strength of the distortion increases parametrically with the relative productivity of a firm. In each case, the distortion raises unemployment. For comparison, I show for each case both the effect of the distortion and the effect of an increase in vacancy posting costs that leads to the same change in unemployment.

The first column of Table 16 shows the effect of a linear tax of 5% on revenue of both employers and own-account workers. The second column shows the effect of an increase in k_v that leads to the same unemployment rate. The third column shows the effect of introducing a size-dependent distortion in the benchmark economy. As is common, I model SDDs as

Table 16: Comparing the effect of k_v to that of taxes and distortions

change in	5% linear tax	equivalent increase in k_v	size-dependent distortion	equivalent increase in k_v
tightness (%)	-14.3	-27.1	-17.1	-28.5
UN ratio (% pts)	0.5	1.8	0.6	1.9
u rate (% pts)	1.1	1.1	1.1	1.1
incomes:				
w (%)	-8.0	-2.2	-9.6	-2.3
OA/w (%)	0.2	0.4	0.3	0.4
employer/w (%)	7.5	5.0	0.0	5.4
mean firm n	40.0	-15.5	36.6	-16.5
entrepreneurship (% pts)	-5.5	3.5	-5.2	3.8
own-account work (% pts)	-4.5	3.7	-4.1	3.9
fraction employers (% pts)	-1.0	-0.1	-1.1	-0.1

Note: The first column shows the effect of a universal tax of 5% on both employers and own-account workers, compared to the benchmark (the average economy calibration). The second column shows the effect of an increase in k_v that leads to the same unemployment rate. This requires an increase in k_v by 29%. The third column shows the effect of size-dependent distortions, parameterized by $\nu = 0.0315$. The fourth column shows the effect of an increase in k_v that leads to the same unemployment rate. This requires an increase in k_v by 31%.

productivity-specific taxes on firm revenue.[36] To be precise, I use a specification as in Buera and Fattal-Jaef (2016), and assume that firm revenue is taxed at a rate τ such that

$$1 - \tau(z) = \left(\frac{z}{\bar{z}}\right)^{-\frac{\nu}{1-\gamma}}, \nu \geq 0. \tag{17}$$

Since $1 - \tau$ can be larger or smaller than one, depending on z, this specification allows for both taxes (for highly productive firms with high levels of z) and subsidies (for low z). For comparability to results with linear taxes, I set the constant $\bar{z}$ to 1, and set ν to 0.0315, such that the average tax on employer firms is 5%. (The output-weighted average tax rate on employers is 5.8%. In general, the relationship between ν and the average tax rate depends on parameters, in particular σ_z, and on equilibrium outcomes, in particular the employer threshold z_f.)[37] The final column again shows the effect of an increase in k_v generating the same unemployment rate as the SDDs.

As expected, both types of taxes reduce the incentive to post vacancies, thereby reduce market tightness, and lead to higher unemployment. They also strongly reduce wages. In fact, because lower taxes in this model reduce both match surplus and the value of workers' outside option (via a lower value of self-employment), wages decline by more than 5%. (High vacancy posting costs, in contrast, raise match surplus.) The incomes of own-account workers decline by 5% for a given level of productivity, and therefore less than wages. Employers' incomes decline less than wages with the linear tax (because of workers' inferior outside option), but just as much with SDDs, because of a large decline in the highest incomes.

With linear taxes, lower wages imply larger firm size, and thus lead to a reduction in equilibrium entrepreneurship and entry. This is why the fraction of own-account workers declines despite the slight increase in their relative income.[38] In this particular case, SDDs raise average firm size as they strongly discourage entry and reduce the number of employers.

[36] The modeling device of size-specific taxes can capture both factors like a burden of taxes and regulation that is in fact higher for larger firms (an interpretation taken by e.g. Guner et al. (2008)) or internal frictions that affect larger firms more strongly and limit their expansion, like frictions in delegation (see e.g. Akcigit, Alp and Peters (2017) and Grobovšek (2017)). Another potential source of misallocation consists in financial frictions (see e.g. Banerjee and Duflo (2005) for evidence from India, and Cagetti and De Nardi (2006) and Buera (2009) for a macroeconomic approach).

[37] In the calibration, ν and the variance of productivity σ_z cannot be identified separately. This is why ν was set to zero in all calibrations. Instead, one could for example have chosen common σ_z and allowed for differences in ν. For reference, a calibration with σ_z of 0.2 and ν of 0.3 fits Ethiopia similarly well as the calibration shown in Table 6 and Table 8.

[38] This result is due to the modeling assumption that productivity is unknown before entry, and the same entry decision can lead to entry as an own-account worker or an employer, depending on the realization of productivity. For the benchmark economy, the result goes through even when only employers are taxed (not reported in the table).

(This effect differs a bit across calibrations.)[39]

It is clear from Table 16 that, although distortions and hiring frictions both raise unemployment, their further effects on the economy are different. In particular, frictions have a much smaller effect on wages (they do not reduce the match surplus; to the contrary), and therefore lead to smaller firms in all settings. Because own-account workers are not affected by frictions, own-account work becomes more attractive relative to both wage work and being an employer. As a consequence, stronger labor market frictions always raise the entrepreneurship rate, via an increase in the fraction of own-account workers.

The distortions analyzed here do not have this effect. In general, this of course depends on the type of distortion one considers. One could, for instance, envisage a distortion that takes the form of a subsidy to own-account workers. This could be modelled as an increase in ζ. Table 15 shows that this would raise the fraction of own-account workers. However, it would also reduce the unemployment rate, reiterating the finding from Section 5 that variation in ζ is inconsistent with the observed covariation of self-employment and unemployment in the data.

The exercises in this section have one key point in common: they have shown that changes to the relative attractiveness of search and entrepreneurial entry affect occupational choice, with direct consequences for the rate of unemployment and for output. When the self-employment rate is high, there is a large pool of individuals who could change their occupational choices in response to changes in the environment or in policies, with implications for the unemployment rate and for aggregate output. Neglecting these would lead to very misleading results, since occupational choice constitutes an important margin of adjustment.

8 Conclusion

The distribution of employment states varies strongly with income per capita. Labor markets in poor countries are characterized not only by higher levels of self-employment, but also by more unemployment relative to wage employment (a high UN ratio), indicating difficulty of job search. In addition, the self-employment rate is particularly high where the UN ratio is high. A search and matching model with occupational choice is flexible enough to be able to reproduce these patterns and match labor market outcomes in a very diverse set of countries.

A quantitative analysis of the model points to variation in labor market frictions as the dominant driver of differences in unemployment and self-employment across countries. This

[39] The effect of distortions on output and aggregate productivity found here is in line with those reported in the literature. For reasons of space, I do not discuss them in detail.

is true both for the univariate and joint distribution of unemployment and self-employment. This analysis points to high hiring costs or low matching efficiency and a high job destruction rate as the root causes of not only high UN ratios, but also high self-employment in poor countries. The analysis also shows that reduced labor market frictions would not only imply more wage employment and less self-employment in poor economies, but also substantial output gains. These stem not only from reduced unemployment, but also from a more efficient allocation of resources, with fewer own-account workers and more wage employees, employed in relatively more productive firms. Evidently, changes in occupational choice are central for these results.

The theoretical analysis in this paper was guided by the objective to stay as close as possible to a standard DMP model, and to add only the minimum extensions required to capture key features of the economic environment under study. The quantitative performance of the model shows that these simple extensions already go very far. Nevertheless, identifying more precisely what kind of labor market frictions are so large in poor countries would clearly be valuable. Doing so would require using richer data and a richer model. In return, it would allow analyzing more specific policies than the present, fairly abstract setting. Two particular directions for further work come to mind.

First, part of the reason unemployment is so high relative to employment in a country like Ethiopia is that the job destruction rate is high, while the job finding rate is low. It is not clear why the destruction rate is so high, in particular given the high cost of creating productive matches. One possibility is that match quality is very uncertain, leading both to a high destruction rate and a high cost of creating a lasting match.

Second, the analysis in this paper assumed that worker are homogeneous. However, self-employment rates, unemployment rates, and their age patterns differ by education group. Not all groups face the same situation, and there can be spillovers across groups, if e.g. the highly skilled act as employers for others.

Hence, extending both the empirical and the theoretical analysis is a promising avenue for future research. In poor countries, self-employment and unemployment are intimately linked, and further joint analysis of the two could greatly improve understanding of and policy towards these phenomena.

References

Abebe, G., Caria, S., Fafchamps, M., Falco, P. and Franklin, S. (2016), 'Curse of Anonymity or Tyranny of Distance? The Impacts of Job-Search Support in Urban Ethiopia', *NBER Working Paper* **22409**.

Abebe, G., Caria, S. and Ortiz-Ospina, E. (2017), 'The Selection of Talent: Experimental and Structural Evidence from Ethiopia', *Mimeo, University of Oxford* .

Akcigit, U., Alp, H. and Peters, M. (2017), 'Lack of Selection and Limits to Delegation: Firm Dynamics in Developing Countries ', *NBER Working Paper* **21905**.

Albrecht, J., Navarro, L. and Vroman, S. (2009), 'The effects of labour market policies in an economy with an informal sector', *The Economic Journal* **119**(539), 1105–1129.

Asker, J., Collard-Wexler, A. and De Loecker, J. (2014), 'Dynamic inputs and resource (mis)allocation', *Journal of Political Economy* **122**(5), 1013–1063.

Atkeson, A. and Kehoe, P. (2005), 'Modeling and measuring organization capital', *Journal of Political Economy* **113**(5), 1026–1053.

Banerjee, A. V. and Duflo, E. (2005), Growth Theory through the Lens of Development Economics, *in* P. Aghion and S. N. Durlauf, eds, 'Handbook of Economic Growth', Vol. 1A, Elsevier, Amsterdam, pp. 473–552.

Bartelsman, E., Haltiwanger, J. and Scarpetta, S. (2004), 'Microeconomic Evidence of Creative Destruction in Industrial and Developing Countries', *World Bank Policy Research Working Paper* **3464**.

Bassi, V. and Nansamba, A. (2018), 'Information on skills in the labor market: Experimental evidence from uganda', *mimeo, University of Southern California* .

Bauducco, S. and Janiak, A. (2018), 'The macroeconomic consequences of raising the minimum wage: capital accumulation, employment and the wage distribution', *European Economic Review* **101**, 57–76.

Berry, A., Rodriguez, E. and Sandee, H. (2002), 'Firm and group dynamics in the small and medium enterprise sector in Indonesia', *Small Business Economics* **18**(1), 141–161.

Bick, A., Fuchs-Schündeln, N. and Lagakos, D. (2018), 'How Do Hours Worked Vary with Income? Cross-Country Evidence and Implications', *American Economic Review* **108**(1), 170–99.

Bigsten, A., Mengistae, T. and Shimeles, A. (2007), 'Mobility and earnings in ethiopia's urban labor markets: 1994-2004', *World Bank Policy Research Working Paper* **4168**.

Blattman, C. and Dercon, S. (2016), 'Occupational choice in early industrializing societies: Experimental evidence on the income and health effects of industrial and entrepreneurial work', *IZA Discussion Paper* **10255**.

Bosch, M. and Esteban-Pretel, J. (2012), 'Job creation and job destruction in the presence of informal markets', *Journal of Development Economics* **98**(2), 270–286.

Bradley, J. (2016), 'Self-employment in an equilibrium model of the labor market', *IZA Journal of Labor Economics* **5**(1), 6.

Bruegemann, B., Gautier, P. and Menzio, G. (2015), 'Intra Firm Bargaining and Shapley Values', *NBER Working Paper* **21508**.

Buera, F. and Fattal-Jaef, R. (2016), 'The dynamics of development: Entrepreneurship, innovation, and reallocation', *manuscript* .

Buera, F. J. (2009), 'A Dynamic Model of Entrepreneurship with Borrowing Constraints', *Annals of Finance* **5**(3-4), 443–464.

Cagetti, M. and De Nardi, M. (2006), 'Entrepreneurship, Frictions, and Wealth', *Journal of Political Economy* **114**(5), 835–870.

Cahuc, P., Marque, F. and Wasmer, E. (2004), 'A Theory of Wages and Labor Demand with Intrafirm Bargaining and Matching Frictions', *CEPR Discussion Paper* **4605**.

Caselli, F. (2005), Accounting for Cross-Country Income Differences, *in* P. Aghion and S. N. Durlauf, eds, 'Handbook of Economic Growth', North Holland, Amsterdam.

Central Statistical Agency of Ethiopia (2015), *Statistical report on the 2015 Urban Employment Unemployment Survey (UEUS)*, Addis Ababa.
URL: *http://www.csa.gov.et/images/general/news/2015*

Delacroix, A., Fonseca, R., Poschke, M. and Ševčík, P. (2016), 'The cyclical dynamics of entrepreneurship and self-employment', *mimeo* .

Donovan, K., Lu, J. and Schoellman, T. (2017), 'Labor Force Attachment Across Countries', *mimeo, University of Notre Dame* .

Elsby, M. W. L., Hobijn, B. and Şahin, A. (2013), 'Unemployment Dynamics in the OECD', *Review of Economics and Statistics* **95**, 530–548.

Elsby, M. W. and Michaels, R. (2013), 'Marginal jobs, heterogeneous firms, and unemployment flows', *American Economic Journal: Macroeconomics* **5**(1), 1–48.

Feenstra, R., Inklaar, R. and Timmer, M. (2015), 'The next generation of the penn world table', *American Economic Review* **105**(10), 3150–3182.

Feng, Y., Lagakos, D. and Rauch, J. E. (2017), 'Unemployment and development', *mimeo, University of California in San Diego* .

Fonseca, R., Lopez-Garcia, P. and Pissarides, C. A. (2001), 'Entrepreneurship, start-up costs and employment', *European Economic Review* **45**(4-6), 692–705.

Franklin, S. (2014), 'Work, unemployment and job search among the youth in ethiopia', *World Bank Poverty Assessment Background Paper* .

Franklin, S. (2016), 'Location, Search Costs and Youth Unemployment: A Randomized Trial of Subsidized Transport', *mimeo, Oxford* .

Galindo da Fonseca, J. (2018), 'Unemployment, Entrepreneurship and Firm Outcomes', *mimeo, UBC* .

Gollin, D. (2002), 'Getting income shares right', *Journal of Political Economy* **110**(2), 458–474.

Gollin, D. (2007), 'Nobody's business but my own: Self-employment and small enterprise in economic development', *Journal of Monetary Economics* **55**(2), 219–233.

Grobovšek, J. (2017), 'Managerial Delegation, Law Enforcement, and Aggregate Productivity', *mimeo, University of Edinburgh* .

Guner, N., Ventura, G. and Xu, Y. (2008), 'Macroeconomic Implications of Size-Dependent Policies', *Review of Economic Dynamics* **11**(4), 721–744.

Herrendorf, B., Rogerson, R. and Valentinyi, A. (2014), Growth and structural transformation, *in* 'Handbook of Economic Growth', Vol. 2, Elsevier, pp. 855–941.

Hipple, S. (2010), 'Self-employment in the United States', *Monthly Labor Review* **133**(9), 17–32.

Hsieh, C.-T. and Klenow, P. (2009), 'Misallocation and Manufacturing TFP in China and India', *Quarterly Journal of Economics* **124**, 1403–1448.

Kredler, M., Millan, A. and Visschers, L. (2014), 'Great opportunities or poor alternatives: self-employment, unemployment and paid employment over the business cycle', *Society for Economic Dynamics 2014 Meeting Papers* .

Lagakos, D., Moll, B., Porzio, T., Qian, N. and Schoellman, T. (2018), 'Life cycle wage growth across countries', *Journal of Political Economy* **126**(2).

Lucas, R. (1978), 'On the size distribution of business firms', *Bell Journal of Economics* **9**, 508–523.

Margolis, D. N., Navarro, L. and Robalino, D. A. (2012), 'Unemployment insurance, job search and informal employment', *IZA Discussion Paper* **6660**.

Meghir, C., Narita, R. and Robin, J.-M. (2015), 'Wages and Informality in Developing Countries', *American Economic Review* **105**(4), 1509–1546.

Minnesota Population Center (2017), *Integrated Public Use Microdata Series, International: Version 6.5 [dataset]*, University of Minnesota, Minneapolis.

Narita, R. (2014), 'Self Employment in Developing Countries: a Search-Equilibrium Approach', *mimeo, University of São Paulo* .

Poschke, M. (2018), 'The firm size distribution across countries and skill-biased change in entrepreneurial technology', *American Economic Journal: Macroeconomics* .

Restuccia, D. and Rogerson, R. (2008), 'Policy Distortions and Aggregate Productivity with Heterogeneous Establishments', *Review of Economic Dynamics* **11**(4), 707–720.

Rud, J. P. and Trapeznikova, I. (2016), 'Wage Dispersion, Job Creation and Development: Evidence from Sub-Saharan Africa ', *mimeo, Royal Holloway* .

Shimer, R. (2012), 'Reassessing the ins and outs of unemployment', *Review of Economic Dynamics* **15**(2), 127–148.

Stole, L. A. and Zwiebel, J. (1996), 'Intra-firm bargaining under non-binding contracts', *The Review of Economic Studies* **63**(3), 375–410.

Ulyssea, G. (2010), 'Regulation of entry, labor market institutions and the informal sector', *Journal of Development Economics* **91**, 87–99.

Zenou, Y. (2008), 'Job search and mobility in developing countries. Theory and policy implications', *Journal of Development Economics* **86**(2), 336–355.

Appendix

A Additional Tables and Figures

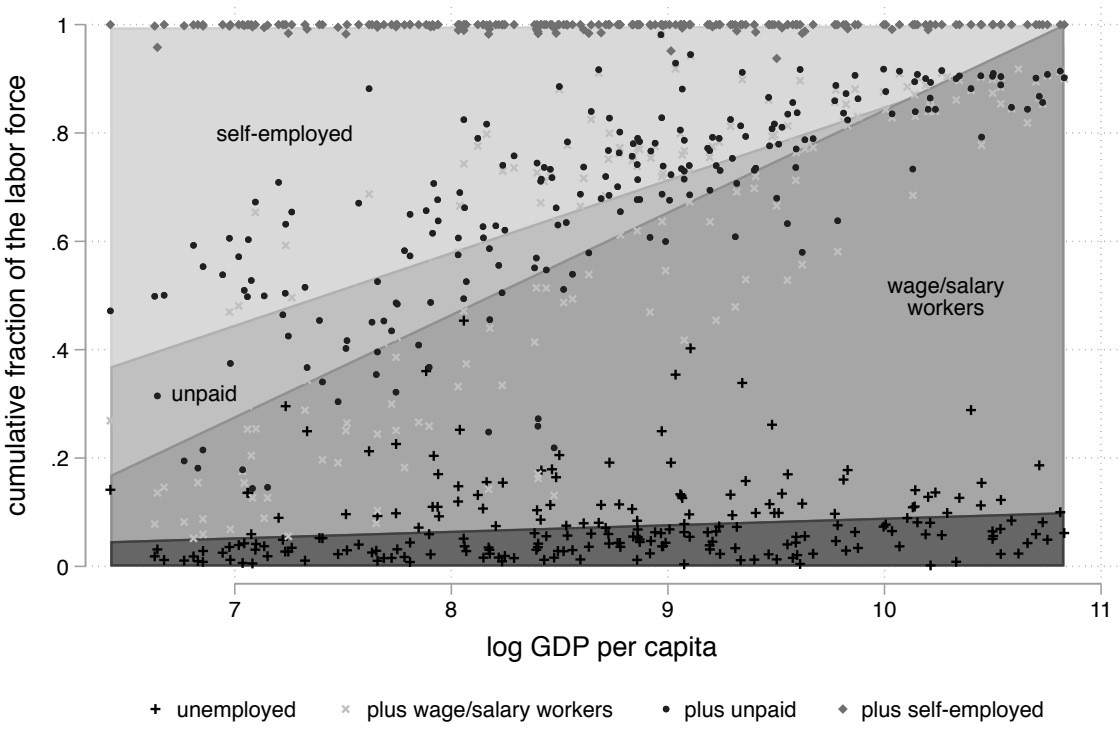


Figure 9: Composition of the labor force and development, national, incl. unpaid workers

Sources: See Figure 1.

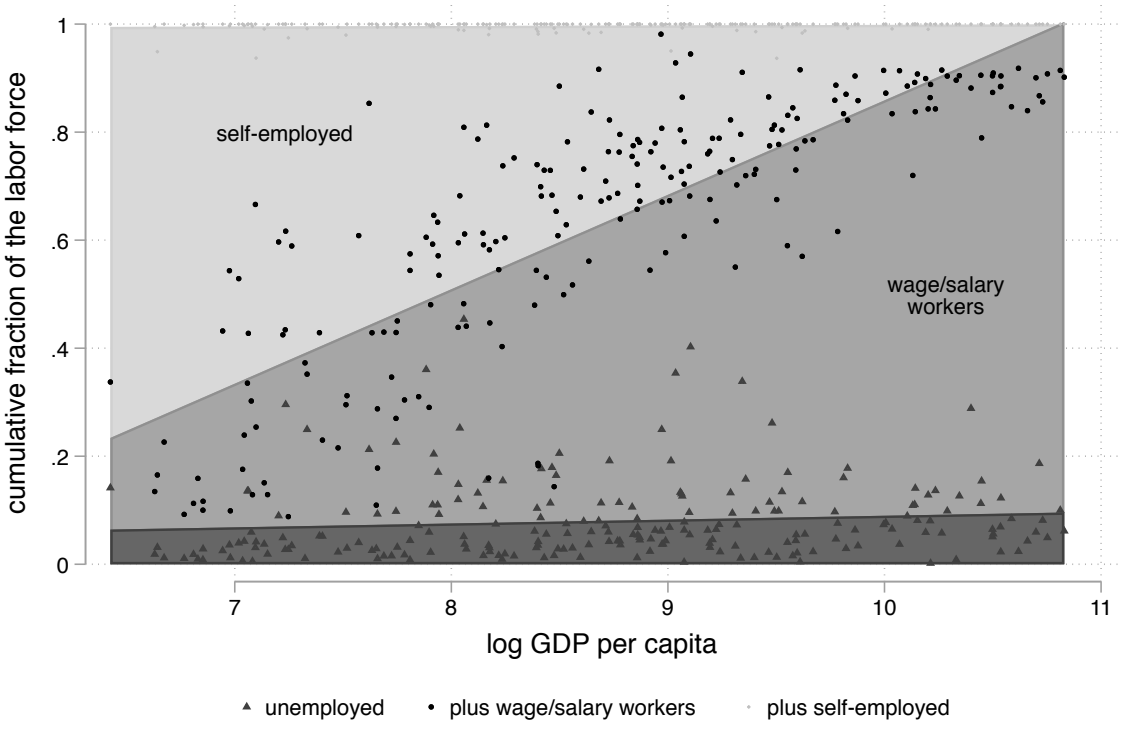


Figure 10: Composition of the labor force and development, national, excl. unpaid workers

Sources: See Figure 1.

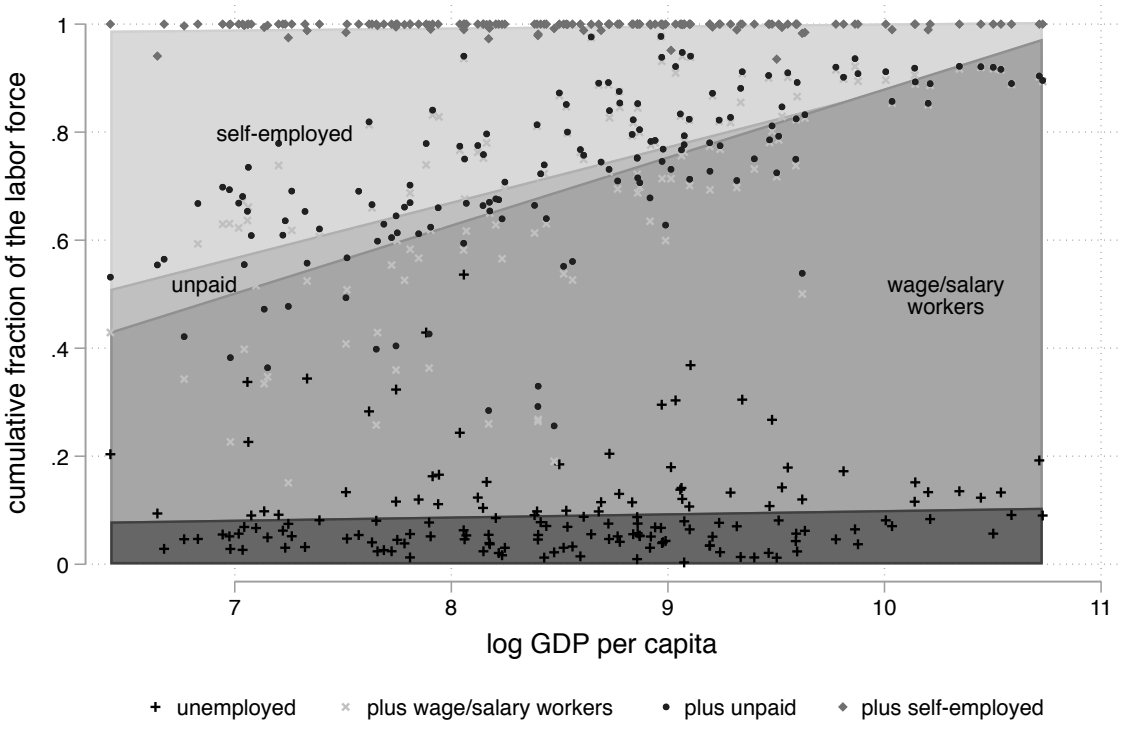


Figure 11: Composition of the labor force and development, incl. unpaid workers

Sources: See Figure 1.

Table 17: Composition of the labor force and development, pooled regressions

dependent variable:	self-employment rate	rate of wage employment	unemployment rate	UN ratio
Urban areas:				
log GDP per capita	-0.111***	0.112***	0.006	-0.022*
	(0.011)	(0.013)	(0.008)	(0.012)
R^2	0.433	0.422	0.005	0.037
observations	150	150	165	150
Entire country:				
log GDP per capita	-0.174***	0.168***	0.012**	-0.025**
	(0.012)	(0.012)	(0.005)	(0.010)
R^2	0.664	0.676	0.035	0.062
observations	214	214	235	214

Notes: The table shows regression coefficients from regressions of the dependent variable on log GDP per capita, pooling all observations. Constant not reported. Robust standard errors, clustered by country, in parentheses. * (**) [***] indicates $p < 0.1$ (< 0.05) [< 0.01]. Data sources as in Figure 1.

Table 18: Composition of the labor force and development, data from top comparability tier

dependent variable:	self-employment rate	rate of wage employment	unemployment rate	UN ratio
Urban areas:				
log GDP per capita	-0.145***	0.148***	0.002	-0.030**
	(0.023)	(0.023)	(0.008)	(0.013)
R^2	0.509	0.507	0.002	0.116
observations	93	93	101	93
countries	41	41	45	41
Entire country:				
log GDP per capita	-0.202***	0.189***	0.017***	-0.018
	(0.021)	(0.020)	(0.006)	(0.011)
R^2	0.656	0.639	0.135	0.054
observations	124	124	134	124
countries	50	50	55	50

Notes: The table shows regression coefficients from regressions of the dependent variable on log GDP per capita, using time averages of country data (between regression). Constant not reported. * (**) [***] indicates $p < 0.1$ (< 0.05) [< 0.01]. Data sources as in Figure 1.

Table 19: Composition of the labor force and development, ILO data

dependent variable:	self-employment rate	fraction own-account workers	fraction employers	unemployment rate	UN ratio
1995-2007:					
log GDP per capita	-0.086***	-0.092***	-0.007	-0.003	-0.021**
	(0.010)	(0.010)	(0.008)	(0.007)	(0.009)
R^2	0.490	0.528	0.011	0.004	0.168
observations	588	622	596	265	254
countries	80	83	83	36	31
All available years:					
log GDP per capita	-0.109***	-0.114***	0.001	0.014***	-0.023***
	(0.008)	(0.008)	(0.003)	(0.004)	(0.009)
R^2	0.641	0.663	0.000	0.138	0.127
observations	1241	1334	1255	598	548
countries	106	107	107	71	54
earliest sample year	1976	1960	1976	1960	1992
latest sample year	2014	2014	2014	2014	2014

Notes: The table shows regression coefficients from regressions of the dependent variable on log GDP per capita, using time averages of country data (between regression). Constant not reported. * (**) [***] indicates $p < 0.1$ (< 0.05) [< 0.01]. Data from the International Labour Organization (ILOSTAT).

Table 20: Composition of the labor force and development, role of unpaid workers

dependent variable:	fraction unpaid workers, urban	unemployment rate excluding unpaid workers, urban	fraction unpaid workers, entire country	unemployment rate excluding unpaid, entire country
log GDP per capita	-0.027***	-0.003	-0.061***	0.005
	(0.005)	(0.011)	(0.010)	(0.008)
R^2	0.340	0.002	0.347	0.007
observations	148	150	207	214
countries	58	58	66	68

Notes: The table shows regression coefficients from regressions of the dependent variable on log GDP per capita, using time averages of country data (between regression). Constant not reported. * (**) [***] indicates $p < 0.1$ (< 0.05) [< 0.01]. Data sources as in Figure 1. Results for a regression using pooled data are similar.

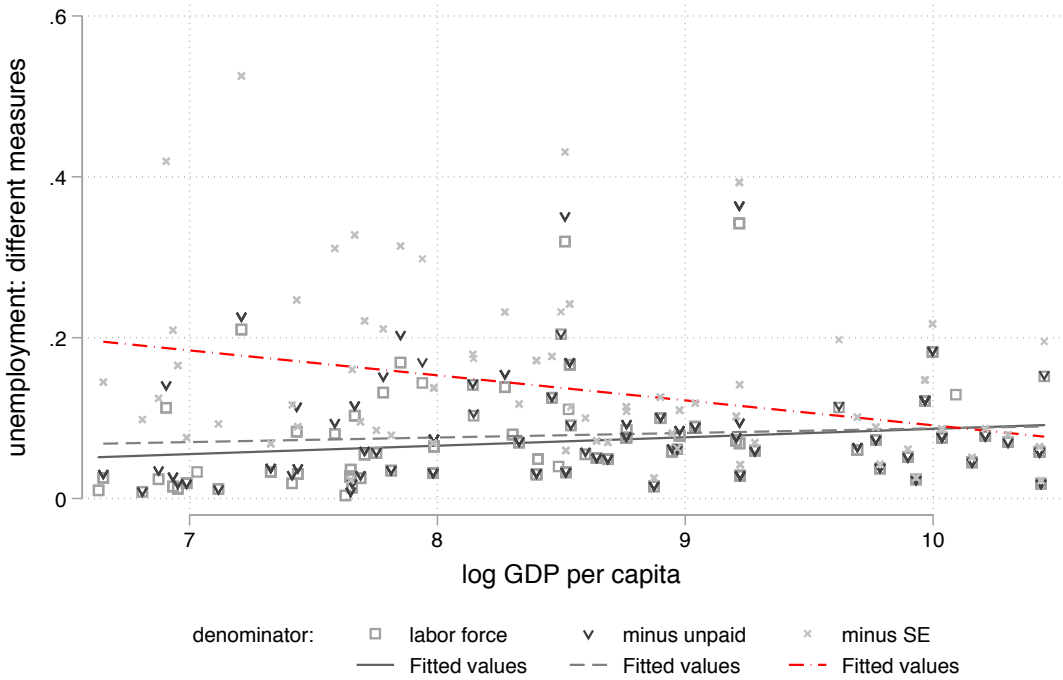


Figure 12: Different measures of unemployment and development

Notes: Data sources as in Figure 1. Data for the entire country. Regression outputs underlying the lines of best fit reported in Table 20.

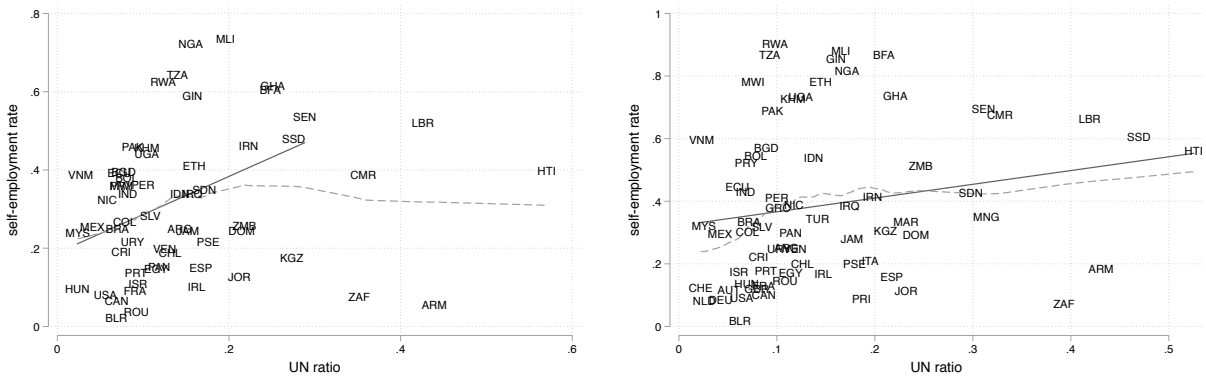


Figure 13: The self-employment rate versus the UN ratio $u/(u+n)$, urban (left) and overall (right), full range of the UN ratio

Notes: Dashed line: linear regression. Dotted line: Fit from locally weighted regressions (`lowess` command in Stata).

Table 21: Unemployment and development, participation rate and alternative measure of unemployment

dependent variable:	non-participation rate	narrow unemployment rate	UN ratio using narrow u rate
Urban areas:			
log GDP per capita	-0.028**	-0.008	-0.044***
	(0.012)	(0.011)	(0.014)
R^2	0.091	0.009	0.149
observations	150	150	150
countries	58	58	58
Entire country:			
log GDP per capita	-0.033***	0.002	-0.043***
	(0.011)	(0.009)	(0.011)
R^2	0.120	0.001	0.180
observations	214	214	214
countries	68	68	68

Notes: The table shows regression coefficients from regressions of the dependent variable on log GDP per capita, using time averages of country data (between regression). Constant not reported. Standard errors in parentheses. * (**) [***] indicates $p < 0.1$ (< 0.05) [< 0.01]. Data sources as in Figure 1.

Table 22: The relationship between self-employment and the UN ratio, controlling for GDP per capita, urban areas, pooled regressions

dependent variable:	self-employment rate	fraction own-account workers	fraction employers
UN ratio	0.542**	0.550**	0.026
	(0.222)	(0.217)	(0.033)
log GDP per capita	-0.112***	-0.125***	0.008***
	(0.010)	(0.011)	(0.003)
R^2	0.499	0.521	0.121
observations	136	126	126

Notes: The table shows regression coefficients from regressions of the dependent variable on the UN ratio and log GDP per capita, using pooled data. Constant not reported. Robust standard errors clustered at the country level in parentheses. * (**) [***] indicates $p < 0.1$ (< 0.05) [< 0.01]. Data sources as in Figure 1.

Table 23: The relationship between self-employment and the UN ratio, controlling for GDP per capita, urban areas, data from top comparability tier only

dependent variable:	self-employment rate	fraction own-account workers	fraction employers
UN ratio	0.692**	0.594*	0.066
	(0.315)	(0.343)	(0.062)
log GDP per capita	-0.132***	-0.149***	0.012**
	(0.023)	(0.028)	(0.005)
R^2	0.562	0.513	0.146
observations	90	83	83
countries	41	37	37

Notes: The table shows regression coefficients from regressions of the dependent variable on the UN ratio and log GDP per capita, using time averages of data (between regression). Constant not reported. Standard errors in parentheses. * (**) [***] indicates $p < 0.1$ (< 0.05) [< 0.01]. Data sources as in Figure 1.

Table 24: The relationship between self-employment and the UN ratio, controlling for GDP per capita, entire country

dependent variable:	self-employment rate	fraction own-account workers	fraction employers
Between regression:			
UN ratio	-0.067	-0.170	0.033
	(0.269)	(0.314)	(0.037)
log GDP per capita	-0.195***	-0.198***	0.010***
	(0.017)	(0.020)	(0.002)
R^2	0.684	0.633	0.242
observations	197	172	172
countries	64	59	59
Pooled regression:			
UN ratio	0.130	0.118	-0.006
	(0.193)	(0.208)	(0.026)
log GDP per capita	-0.175***	-0.191***	0.011***
	(0.013)	(0.015)	(0.002)
R^2	0.676	0.649	0.215
observations	197	172	172

Notes: The table shows regression coefficients from regressions of the dependent variable on the UN ratio and log GDP per capita, using time averages of data (between regression). Constant not reported. Standard errors in parentheses. * (**) [***] indicates $p < 0.1$ (< 0.05) [< 0.01]. Data sources as in Figure 1.

Table 25: The relationship between self-employment and the UN ratio, controlling for GDP per capita, entire country (ILO data)

dependent variable:	self-employment rate	fraction own-account workers	fraction employers
UN ratio	-0.194	-0.373	0.179**
	(0.350)	(0.318)	(0.075)
log GDP per capita	-0.098***	-0.102***	0.005
	(0.018)	(0.017)	(0.004)
R^2	0.534	0.591	0.169
observations	254	254	254
countries	31	31	31

Notes: The table shows regression coefficients from regressions of the dependent variable on the UN ratio and log GDP per capita, using ILO data for 1995 to 2007. The regressions use time averages of data (between regression). Constant not reported. Standard errors in parentheses. * (**) [***] indicates $p < 0.1$ (< 0.05) [< 0.01]. Results are virtually identical when years before 1995 are included.

Table 26: Calibration: model and data moments (8 countries and average, data values in parentheses)

country:	avg	ETH	USA	CAN	DEU	FRA	ITA	MEX	IDN
Targeted moments:									
Unemployment	0.180	0.440	0.256	0.062	0.086	0.062	0.398	0.091	0.044
outflow rate	(0.180)	(0.440)	(0.257)	(0.062)	(0.086)	(0.062)	(0.397)	(0.091)	(0.045)
Unemployment	0.106	0.051	0.069	0.107	0.130	0.152	0.042	0.058	0.237
rate	(0.106)	(0.051)	(0.069)	(0.107)	(0.129)	(0.152)	(0.042)	(0.058)	(0.237)
Casual	0.000	0.000	0.000	0.000	0.000	0.000	0.056	0.114	0.245
employment	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.056)	(0.114)	(0.240)
Fraction own-	0.149	0.049	0.069	0.046	0.040	0.157	0.221	0.311	0.288
account workers	(0.149)	(0.048)	(0.069)	(0.053)	(0.040)	(0.157)	(0.221)	(0.312)	(0.290)
Fraction	0.044	0.048	0.047	0.053	0.039	0.054	0.032	0.033	0.050
employers	(0.044)	(0.049)	(0.047)	(0.046)	(0.039)	(0.054)	(0.032)	(0.033)	(0.050)
Firm exit	0.109	0.110	0.105	0.060	0.090	0.085	0.140	0.140	0.142
rate (annual)	(0.109)	(0.110)	(0.105)	(0.060)	(0.090)	(0.085)	(0.140)	(0.140)	(0.142)
Firm size	0.740	0.846	0.876	0.830	0.923	0.816	0.755	0.316	0.871
target (see note)	(0.715)	(0.847)	(0.877)	(0.830)	(0.923)	(0.816)	(0.755)	(0.332)	(0.874)
Labor	0.670	0.670	0.670	0.670	0.670	0.670	0.673	0.670	0.671
income share	(0.67)	(0.67)	(0.67)	(0.67)	(0.67)	(0.67)	(0.67)	(0.67)	(0.67)
b/w	0.400	0.400	0.399	0.400	0.400	0.398	0.399	0.400	0.401
	(0.40)	(0.40)	(0.40)	(0.40)	(0.40)	(0.40)	(0.40)	(0.40)	(0.40)
Not targeted:									
UN ratio	0.128	0.057	0.077	0.117	0.139	0.185	0.055	0.086	0.320
	(0.129)	(0.056)	(0.077)	(0.117)	(0.139)	(0.185)	(0.055)	(0.086)	(0.320)
Entry rate h	0.077	0.018	0.014	0.005	0.004	0.009	0.072	0.439	0.014
Separation rate	0.026	0.025	0.020	0.007	0.013	0.012	0.052	0.027	0.046
Mean firm	5.175	10.298	8.617	10.091	12.627	4.730	3.731	2.577	2.235
employment									
Mean income relative to w for									
own-acct wkrs	1.03	1.16	1.36	1.04	1.22	1.29	3.37	1.04	1.10
employers	9.05	6.25	9.24	8.37	11.62	7.17	10.80	8.77	5.07
Business inc./Y	0.41	0.25	0.37	0.35	0.35	0.43	0.63	0.55	0.66

Notes: Countries are Ethiopia (ETH), United States (USA), Canada (CAN), Germany (DEU), France (FRA), Italy (ITA), Mexico (MEX), Indonesia (IDN). "avg" stands for the calibration targeting average values of data moments. Targeted model moments are in square brackets. The firm size target varies by country depending on data availability: For ETH, it is the share of firms with less than 10 employees; for MEX and IDN, it is the share of employment in firms with at least 20 employees; and for the remaining countries, it is the share of employment in firms with at least 10 employees.

Table 27: Calibration: parameter values (8 countries and average)

country:		avg	ETH	USA	CAN	DEU	FRA	ITA	MEX	IDN
externally calibrated:										
r	discount rate (annualized)					common: 0.04				
ϕ	retirement probability (annualized)					common: 1/40				
μ	matching function					common: 0.5				
γ	decreasing returns to scale					common: 0.85				
internally calibrated:										
k_f	entry cost	26.2	13.54	61	55.5	36	56.3	36	73.5	44
k_v	vacancy posting cost	30.1	69	10.4	24.1	285	106	144	12	66.7
η	worker bargaining power	0.225	0.432	0.12	0.158	0.19	0.239	0.295	0.207	0.364
b	utility flow of unemployment	0.235	0.188	0.196	0.205	0.26	0.198	0.208	0.177	0.204
λ_f, λ_s	firm exit rates (annualized)	0.0874	0.120	0.077	0.087	0.026	0.093	0.057	0.118	0.118
ξ	job destruction rate	0.018	0.03185	0.0194	0.0105	0.0001	0.00164	0.0066	0.012	0.014
σ_z	productivity dispersion	0.2	0.0224	0.162	0.18	0.27	0.155	0.11	0.022	0.1
ζ	relative self-employment productivity	0.6845	0.5191	0.661	0.72	0.535	0.743	0.704	1.55	1.078
δ	casual job probability	0	0.44	0	0	0	0	0	0.537	0.4

Notes: Countries are Ethiopia (ETH), United States (USA), Canada (CAN), Germany (DEU), France (FRA), Italy (ITA), Mexico (MEX), Indonesia (IDN). "avg" stands for the calibration targeting average values of data moments. The firm exit rate reported in this table differs from that reported in Table 26 since the latter also includes exits due to the owner's retirement.

B Proofs and derivations

B.1 Summary of model timing

The following summarizes the timing of events in this economy.

1. If individuals chose to enter, they pay the entry cost k_f and their productivity $z \sim f(z)$ is realized.
2. Depending on z, entrants decide whether
   (a) to keep the business and post vacancies to reach the optimal employment level,
   (b) to be self-employed, or
   (c) to exit and go to the unemployment pool.
3. Shocks (ϕ, λ_f, λ_s, ξ, δ, $\theta \cdot q(\theta)$) are realized.
4. Value functions are measured and occupational choices take place.
5. Production takes place and payoffs (w, b) are realized.

B.2 Detailed Derivation of Wage

As stated in the main part of the paper, workers and firms split the surplus according to workers' bargaining weight η. The total surplus is the sum of workers' and firms' surplus, explicit expressions of which are given below.

Worker's Surplus The value of employment is given by

$$W = w + \frac{1-s}{1+r}W + \frac{s-\phi}{1+r}U$$

Rewrite this to obtain $W - U$:

$$W - U = \frac{1+r}{r+s}w - \frac{r+\phi}{r+s}U$$

Firm's Surplus From equation (7),

$$F_f(n,z) = \frac{1+r}{(1+r)-(1-\phi)(1-\lambda_f)}\left(zn(z)^\gamma - n(z)w - \frac{k_v}{q(\theta)}n(z)(\xi + (1-\xi)\phi)\right) + \frac{(1-\phi)\lambda_f}{(1+r)-(1-\phi)(1-\lambda_f)}U. \tag{18}$$

Then the marginal value of hiring an additional worker the firm has just met, and keeping that worker until either the firm shuts down or some type of separation occurs, is given by

$$c_0 \left(y'(n) - w - n \cdot w'(n) \right),$$

where c_0 is derived as follows. From the firm's sequence problem, the marginal value of an additional worker is

$$\sum_{j=0}^{\infty} \left(\frac{(1-\phi)(1-\lambda_f)}{1+r} \right)^j \left[(1-\phi)(1-\xi) \right]^j \left(y'(n) - w - n \cdot w'(n) \right)$$

Let

$$c_0 \equiv \sum_{j=0}^{\infty} \left(\frac{(1-\phi)^2 (1-\lambda_f)(1-\xi)}{1+r} \right)^j = \frac{1+r}{(1+r) - (1-\phi)^2 (1-\lambda_f)(1-\xi)} = \frac{1+r}{r+s},$$

where $s \equiv 1 - (1-\phi)^2 (1-\lambda_f)(1-\xi)$.

Nash Bargaining The bargaining rule implies that the wage solves

$$(1-\eta)(W-U) = \eta c_0 \cdot \left(y'(n) - w - n \cdot w'(n) \right)$$

Using the expressions above, solving this for w yields the differential equation

$$w = (1-\eta) \frac{r+\phi}{1+r} U + \eta \left(y'(n) - n \cdot w'(n) \right). \tag{19}$$

At a firm's optimal employment, the solution to this equation (details below) is

$$w = \frac{r+\phi}{1+r} U + \frac{\eta}{1-\eta} \left[1 - \frac{(1-\phi)(1-\lambda_f)}{1+r} + \xi + (1-\xi)\phi \right] \cdot \frac{k_v}{q(\theta)}. \tag{20}$$

For this wage, a firm's optimal employment policy is

$$n(z) = (z\gamma)^{\frac{1}{1-\gamma}} \left\{ (\eta(\gamma-1)+1) \left[\left(1 - \frac{(1-\phi)(1-\lambda_f)}{1+r} + \xi + (1-\xi)\phi \right) \frac{k_v}{q} + w \right] \right\}^{\frac{-1}{1-\gamma}}. \tag{21}$$

Solution of the differential equation for w. Without the constant, the equation is

$$w'(n) + \frac{w}{\eta n} - \frac{y'(n)}{n} = 0. \tag{22}$$

The solution of the homogeneous equation

$$w'(n) + \frac{w}{\eta n} = 0$$

then is

$$w(n) = Cn^{-1/\eta}. \tag{23}$$

C is a function of integration that can be a function of n. So take the derivative of equation (23) with respect to n:

$$\frac{\partial w}{\partial n} = C'(n)\, n^{-1/\eta} - \frac{C}{\eta} n^{-1/\eta - 1}$$

Substituting this into (22) yields

$$C'(n) = y'(n)\, n^{1/\eta - 1}$$

Integrating this gives $C(n)$ as

$$C(n) = \int_0^n y'(z)\, z^{1/\eta - 1} dz + D$$

so the wage w is

$$w(n) = n^{-1/\eta} \int_0^n y'(z)\, z^{1/\eta - 1} dz + Dn^{-1/\eta}$$

The constant D can be dealt with assuming that the wage bill goes to zero as employment goes to zeros. This implies $D = 0$. The solution to equation (19) then is

$$w(n) = n^{-1/\eta} \int_0^n y'(z)\, z^{1/\eta - 1} dz + (1-\eta)\frac{r+\phi}{1+r} U$$

Integrating yields

$$w(n) = (1-\eta)\frac{r+\phi}{1+r} U + \frac{y'(n)}{\gamma - 1 + 1/\eta}. \tag{24}$$

The division in the last term here comes from the overhiring effect.

To obtain the wage at the firm's optimal constant level of employment (replacing any workers who leave), use the labor demand condition. To obtain this, equate the marginal value of having an additional employee for the firm's entire life, from (18), to the expected

hiring cost. This results in

$$y'(n) = w + n \cdot w'(n) + \frac{k_v}{q}\left[\frac{1+r-(1-\phi)(1-\lambda_f)}{1+r} + \xi + (1-\xi)\phi\right].$$

To simplify, take the derivative of (24) with respect to n, multiply by n, and replace the $n \cdot w'(n)$ term in the labor demand condition. This yields

$$y'(n) = w + \frac{z\gamma(\gamma-1)n^{\gamma-1}}{\gamma-1+1/\eta} + \left[\frac{1+r-(1-\phi)(1-\lambda_f)}{1+r} + \xi + (1-\xi)\phi\right]\frac{k_v}{q}$$

or

$$y'(n) = [\eta(\gamma-1)+1]\left\{w + \left[\frac{1+r-(1-\phi)(1-\lambda_f)}{1+r} + \xi + (1-\xi)\phi\right]\frac{k_v}{q}\right\}.$$

Solve this for n to obtain the labor demand condition in (21). Substituting this expression into (24) yields the wage at the optimal employment level given in equation (20).

C Data

In this section, I lay out how I compute durations and the distribution of employment status from IPUMS, UEUS and LFS data, and from Bigsten et al. (2007). I also thank the statistical offices that provided the data underlying IPUMS.

C.1 IPUMS data

IPUMS International data (see Minnesota Population Center 2017) is available at `https://international.ipums.org`. I use the variables EMPSTAT (employment status), CLASSWK (class of worker), URBAN (urban-rural status) and INDGEN (industry). In some cases, I also use age.

The variable EMPSTAT (employment status) takes the values 0 (not in universe), 1 (employed), 2 (unemployed), 3 (inactive), 9 (unknown/missing). More detailed 3-digit codes are also provided. The proportion missing is generally small. I code the value 3 as out of the labor force, and 1 and 2 as indicated. The labor force is the union of 1 and 2. My broad ("relaxed") measure of unemployment includes those who are unemployed because no work was available (code 230) and the inactive unemployed (240). (These categories are specified separately only for some countries.) For the narrow measure of unemployment, I exclude these two groups, where possible.

The variable CLASSWK (class of worker) is available for the employed. It takes the values 0 (not in universe), 1 (self-employed), 2 (wage/salary worker), 3 (unpaid worker), 4 (other), 9 (unknown/missing). More detailed 3-digit codes are also provided. I use them to distinguish own-account workers (120) and employers (110). Again, the proportion missing is small. I drop unpaid workers and "other".

The main analysis uses categories of CLASSWK and EMPSTAT as proportions of the labor force.

C.2 UEUS and LFS data for Ethiopia

I use the Urban Employment and Unemployment Surveys (UEUS) for 2012 and 2015, and the 2013 Labor Force Survey (LFS). Throughout, I use only data for Addis Ababa (ID101=14), and use weights (WGT_LB).

For the calibration, I use the distribution of employment status from the UEUS for 2012 (variable U311). I define the following groups: unemployed (23%), public sector worker (including government, government development organizations; 16%), private sector worker (14%), own-account worker (13%), employer (7%), domestic employee (8%), casual or temporary worker (13%), other (coops, unpaid family workers, "other", apprentices; 5%). I then ignore public sector employees and unpaid family workers (1.9% of employment). To further map the groups into model categories, I treat the sum of private sector workers, other, and half of casual or temporary worker as employees, and treat the other half of casual and temporary workers plus domestic workers as casual workers. This leaves us with 42% of private sector employees, 24% of casual workers, 24% of own-account workers, and 9% of employers. The implied unemployment rate is 24%.

In the UEUS for 2012 only, the self-employed provide a measure of "persons participating in the activities of their enterprise." To distinguish own-account workers and employers, I use this measure, not the reported own-account worker versus employer status.

To compute the unemployment outflow rate, I use the employment duration variable, U410. I drop observations with durations over 90 months. The data exhibit severe bunching, first at 0 and 6 months and then at each full year. I smooth this by assuming that a fraction of individuals reports a duration that is rounded downward to the closest year (or 6 months for durations between 6 and 11 months), with a propensity to round that can vary by year of duration. These assumptions generate a duration distribution similar to that in the data, for a common fixed (implied) unemployment outflow rate of 4.5%.

For employment duration, I use the employment duration variables LF319Y (years) and LF319M (months) from the LFS.

C.3 Employment status transitions

Table 7 shows a transition matrix over employment states for model and data. The data matrix is from Bigsten et al. (2007, Table 3, years 2000-2004). Their matrix contains seven employment states: self-employed, government worker, public enterprise worker, formal private sector worker, other private sector worker, unemployment, and out of the labor force. In line with the model, I ignore the second, third, and last groups. Since the model has no formal/informal distinction, I combine groups 4 and 5. I treat group 1 as applying to the union of own-account workers and employers.

C.4 Country codes and acknowledgements

I thank the statistical offices that provided the data underlying IPUMS:

National Institute of Statistics and Censuses, Argentina (ARG)

National Statistical Service, Armenia (ARM)

National Bureau of Statistics, Austria (AUT)

Bureau of Statistics, Bangladesh (BGD)

Ministry of Statistics and Analysis, Belarus (BLR)

National Institute of Statistics, Bolivia (BOL)

Institute of Geography and Statistics, Brazil (BRA)

National Institute of Statistics and Demography, Burkina Faso (BFA)

National Institute of Statistics, Cambodia (KHM)

Central Bureau of Census and Population Studies, Cameroon (CMR)

Statistics Canada, Canada (CAN)

National Institute of Statistics, Chile (CHL)

National Administrative Department of Statistics, Colombia (COL)

National Institute of Statistics and Censuses, Costa Rica (CRI)

National Statistics Office, Dominican Republic (DOM)

National Institute of Statistics and Censuses, Ecuador (ECU)

Central Agency for Public Mobilization and Statistics, Egypt (EGY)

Central Statistical Agency, Ethiopia (ETH)

National Institute of Statistics and Economic Studies, France (FRA)

Federal Statistical Office, Germany (DEU)

Ghana Statistical Services, Ghana (GHA)

National Statistical Office, Greece (GRC)

National Statistics Directorate, Guinea (GIN)

Institute of Statistics and Informatics, Haiti (HTI)

Central Statistical Office, Hungary (HUN)

Ministry of Statistics and Programme Implementation, India (IND)

Statistics Indonesia, Indonesia (IDN)

Statistical Center of Iran, Iran (IRN)

Central Statistical Office, Iraq (IRQ)

Central Statistics Office, Ireland (IRL)

Central Bureau of Statistics, Israel (ISR)

National Institute of Statistics, Italy (ITA)

Department of Statistics, Jordan (JOR)

National Statistical Committee, Kyrgyz Republic (KGZ)

National Statistical Office, Malawi (MWI)

Department of Statistics, Malaysia (MYS)

National Directorate of Statistics and Informatics, Mali (MLI)

National Institute of Statistics, Geography, and Informatics, Mexico (MEX)

High Commission of Planning, Morocco (MAR)

Statistics Netherlands, Netherlands (NLD)

National Institute of Statistics and Censuses, Nicaragua (NIC)

National Bureau of Statistics, Nigeria (NGA)

Statistics Division, Pakistan (PAK)

Census and Statistics Directorate, Panama (PAN)

General Directorate of Statistics, Surveys, and Censuses, Paraguay (PRY)

National Institute of Statistics and Informatics, Peru (PER)

National Institute of Statistics, Portugal (PRT)

National Institute of Statistics, Romania (ROU)

National Institute of Statistics, Rwanda (RWA)

National Agency of Statistics and Demography, Senegal (SEN)

Statistical Office, Slovenia (SLV)

Statistics South Africa, South Africa (ZAF)

National Institute of Statistics, Spain (ESP)

Central Bureau of Statistics, Sudan (SDN)

Federal Statistical Office, Switzerland (CHE)

National Bureau of Statistics, Tanzania (TZA)

Turkish Statistical Institute, Turkey (TUR)

Bureau of Statistics, Uganda (UGA)

Office of National Statistics, United Kingdom (GBR)

Bureau of the Census, United States (USA)

National Institute of Statistics, Uruguay (URY)

National Institute of Statistics, Venezuela (VEN)

General Statistics Office, Vietnam (VNM)

Central Statistical Office, Zambia (ZMB)