Wage employment, unemployment and self-employment across countries*

Markus Poschke†

McGill University, CIREQ and IZA

This version: March 2022. First version: Feb 2017

Abstract

Poor countries have low wage employment and high self-employment. This paper shows that self-employment increases with the ratio of unemployment to wage employment ratio. Cross-country differences in self-employment entry cannot account for this. Quantitative analysis of a heterogeneous-firm search and matching model with choice between job search and self-employment for eight rich and poor countries shows that variation in labor market frictions can explain almost the entire variation in not only unemployment, but also wage employment and self-employment. Moreover, labor market frictions reduce output not only by affecting employment, but also by pushing searchers into low-productivity own-account work.

Keywords: wage employment, unemployment, self-employment, labor market frictions, occupational choice, entrepreneurship, firm size, productivity

JEL codes: O11, E24, J64, L26

*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, Julieta Caunedo, Simon Franklin, Douglas Gollin, Christian Moser, Christopher Pissarides, Roberto Pinheiro, 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, the 2018 meetings of the European Economic Association, the IZA World Labor Conference 2018, the University of Toronto, Cornell University, the Federal Reserve Bank of Cleveland, the University of St. Gallen, Carleton University, the University of Miami, Florida International University and the University of Alberta for valuable comments and suggestions. I gratefully acknowledge financial support from the International Growth Centre (IGC) and from the Social Sciences and Humanities Research Council (SSHRC) of Canada, 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. A central distinguishing feature consists in their very low levels of wage employment. In Addis Ababa, the capital of Ethiopia, for example, only about half of the labor force is in wage employment, own-account workers and the self-employed account for more than a quarter of the labor force, and almost all wage employment is in firms with fewer than 10 workers.[1]

The employment structure in poor countries contrasts with that in rich countries, where wage employment is the norm, and most workers are employed in large firms. In the United States for example, own-account workers make up only about 5% of employment, and wage and salary workers account for about 85% of the labor force. About half of them work in firms with more than 500 employees (Hipple (2010), Census Business Dynamics Statistics).

These differences matter. Indeed, the creation of wage jobs has been identified as a key development challenge – it is the topic of the World Bank's 2013 and 2023 World Development Reports, and the employment rate is part of the United Nations Millennium Development Goals (World Bank 2012, United Nations 2010).[2] But why is wage employment so low, and self-employment so high, in developing countries?

The existing literature on the topic has mostly focussed on barriers to job creation and firm growth, the implications of regulation for firm size, and the effect of technology on the relative returns of wage work and self-employment.[3] In essence, the argument typically is that productivity or wages in wage employment are low in poor countries, while self-employment is comparatively unregulated and easily accessible. As a consequence, many workers enter self-employment. This type of argument implies a negative relationship between wage employment and self-employment across countries.

This paper proposes a different explanation. I argue that low levels of wage employment and high levels of self-employment cannot be understood without taking frictions in labor markets into account. This argument is motivated by the generally high levels of unemployment relative to wage employment in poor countries that I document in this paper. The proposed new mechanism is as follows: As labor market frictions make jobs in wage employment hard to find, they not only cause high unemployment relative to wage employment (few searchers are successful), but also promote self-employment, as an alternative to unattractive

[1]Data from Gindling and Newhouse (2012), World Bank (2012), and author's calculations using the Ethiopian Urban Employment and Unemployment Survey. Most of this article focusses on data for urban areas. Patterns at the level of the entire country are even starker.

[2]These references stress that for the purposes of this question, one should conceive of wage employment broadly, including both formal and informal employment. This paper takes the same approach.

[3]See e.g. Hsieh and Klenow (2014), Buera, Kaboski and Shin (2015); Restuccia and Rogerson (2008), Guner, Ventura and Xu (2008), Albrecht, Navarro and Vroman (2009); Gollin (2007) and Poschke (2018).

job search. Variation in labor market frictions across countries then implies both a negative relationship between wage employment and self-employment, and a positive relationship between the difficulty of job search (and thus the unemployment to wage employment ratio) and self-employment across countries.

The first contribution of this paper is to investigate these relationships among measures of labor force status, and to provide evidence for the new channel that is proposed here. I do so using harmonized census data provided by IPUMS International (Minnesota Population Center 2017), covering 58 countries ranging in income per capita from Ethiopia to the United States. This analysis reveals two relevant new facts. First, the ratio of unemployment to the sum of unemployment plus wage employment, $\tilde{u} \equiv u/(u+n)$ or the "UN ratio", is much higher in poor countries.[4] On average, it decreases 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 areas, self-employment is particularly high in countries with high $\tilde{u}$, even after controlling for GDP per capita. The relationship is quantitatively strong: an increase in $\tilde{u}$ by one percentage point is associated with an increase in the self-employment rate by around 0.7 percentage points.

The remainder of the paper uses two models to understand the sources of these patterns. The first is a simple flow model of the labor market, with fixed flow rates among the three states of wage employment, self-employment and unemployment. It allows for an accounting analysis that illustrates that the empirical patterns cannot be generated by cross-country differences in self-employment entry or exit rates alone, but that differences in job finding or destruction rates are required. The reason is that differences in self-employment entry rates cannot generate the correlation between self-employment and $\tilde{u}$ that is observed in the data, whereas differences in the job finding rate can do so.

To distinguish more precisely among several potential sources of the observed differences in wage employment, self-employment, and unemployment as well as their implications for productivity, I then build and quantitatively analyze a theoretical model. The literature on job search has scarcely addressed self-employment (see e.g. the review by Rogerson and Shimer (2011)), while the literature linking the firm size distribution and aggregate productivity has almost exclusively analyzed the allocation of employment across employer firms, and largely ignored both unemployment and self-employment – despite their importance in poor economies. (The few exceptions are discussed below.) The second contribution of this paper consists in filling this gap. To do so, I develop a theoretical framework that allows

[4]This occurs because poor countries have lower wage employment rates, while the unemployment rate does not vary systematically with income per capita (in line with the findings of Caselli (2005) for a more limited set of countries), implying higher unemployment relative to employment in poor countries.

linking wage employment, unemployment, self-employment, and productivity, and allows exploring their connections via counterfactual analysis.

My model builds on a version of the standard Diamond-Mortensen-Pissarides (DMP) search and matching model with firms that are heterogeneous in size and productivity, as in Elsby and Michaels (2013), augmented with a choice between job search on the one hand and entry into entrepreneurship on the other hand. The key assumption is that job search is subject to search and matching frictions, while entry into entrepreneurship is always possible at a cost.[5] Success, however, is uncertain, as an entrepreneur's productivity is only revealed after entry. This set of assumptions 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 Ethiopia (one of the poorest countries in the world) via Indonesia and Mexico all the way to some European economies and 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 wage employment, 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 wage employment and self-employment. Differences in labor market frictions explain almost all the variation in unemployment, wage employment and self-employment across the eight calibration economies. The model also accounts for at least a third of the positive relationship between self-employment and the UN 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, can generate the observed patterns in self-employment only at the cost of counterfactual variation in unemployment. Size-dependent distortions do not account for much of the variation in labor force status either.

The quantitative analysis leads to two further interesting findings. First, it reveals that

[5] This also presupposes that job search and self-employment are mutually exclusive activities, i.e., only one can be pursued at a time. This assumption is in line with the empirical evidence that self-employment tends to be a full time, persistent activity. See below for details.

while labor market frictions always reduce wage employment, they do so via higher unemployment when firm entry costs are high, as in rich economies, but via higher self-employment when firm entry is cheap, as in poor economies. Second, labor market frictions also affect aggregate output. Part of this comes simply from their effect on unemployment. This effect is largest in developed economies. 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. Labor market frictions thus cause misallocation of labor.

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 wage employment, 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.

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? Is screening hard, with the consequence that matches are experience goods and lasting matches may take time to form?[6] 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 pages takes a first stab at these questions.

[6] See Jovanovic (1984) and Pries and Rogerson (2005, 2022) on matches as experience goods. This interpretation is also in line with the finding by Blattman and Dercon (2018) that manufacturing firms in Ethiopia do not face a shortage of applicants to their vacancies, but experience high rates of quits and turnover.

Related literature. While existing work on unemployment and job search in developed economies is abundant, there are only a few papers studying poorer economies.[7] Among these, Albrecht et al. (2009), Margolis, Navarro and Robalino (2012), Narita (2020), Bradley (2016) and Galindo da Fonseca (2018) also allow for self-employment.[8] Yet, their focus is not on labor market frictions and self-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 these papers 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.

The most closely related theoretical analysis is in Rud and Trapeznikova (2021), who build a DMP-type model to study the effect of barriers to firm entry and labor market frictions on the wage distribution. By assuming that all workers who do not find wage employment engage in self-employment, their analysis equates own-account work to unemployment. As a consequence, it cannot tell apart the effects of entry barriers and labor market frictions.

There also is a small set of papers studying how labor market aggregates vary with income per capita across countries. The seminal paper by Gollin (2007) showed that the self-employment rate declines with income per capita across countries, and analyzed the relationship in a frictionless span of control model building on Lucas (1978). Bick, Fuchs-Schündeln and Lagakos (2018) document how hours worked vary with income per capita within and across countries. Bridgman, Duernecker and Herrendorf (2018) study variation in time spent on household production with GDP per capita.

Most closely related are two very recent working papers. Feng, Lagakos and Rauch (2018) study patterns in the unemployment rate by income per capita using data from IPUMS International and other sources for 55 countries. They find higher unemployment rates in richer countries among workers who have not completed secondary school. Their theoretical

[7]It is also true that little of the work on labor market search in developed countries considers self-employment. Two 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. In earlier work, Fonseca, Lopez-Garcia and Pissarides (2001) and Rissman (2003) analyzed the effect of entry barriers in the OECD and of unemployment insurance benefits, respectively.

[8]Zenou (2008), Ulyssea (2010), Bosch and Esteban-Pretel (2012), and Meghir, Narita and Robin (2015) consider the related but different problem of 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.

and quantitative analysis interprets this as a consequence of structural change. This aligns with the preponderance of agriculture, and agricultural self-employment, in poor countries. Donovan, Lu and Schoellman (2019) are the first to document patterns in labor market flows across a substantial number of countries using micro data. Their sample includes urban areas in 36 countries, ranging from Nicaragua and the Philippines (about the 30th percentile of the global distribution of income per capita across countries) to the United States. This paper makes an important contribution in terms of measurement. It also provides further empirical support to my argument: While most of the paper pools wage employment and self-employment, the appendix of the paper (Figure B1) shows that poor countries have lower flows from unemployment to wage employment and higher flows from unemployment to self-employment, in line with my calibration results.[9]

Finally, there is an emerging literature studying search behavior, labor market frictions and self-employment in developing economies at the micro level, using surveys and experiments. Several papers in this literature find support for the existence of various types of labor market frictions in the specific settings they study. Franklin (2018), Abebe, Caria and Ortiz-Ospina (2021) and Abebe, Caria, Fafchamps, Falco, Franklin and Quinn (forthcoming) all find that reducing search costs at the individual level improves job search outcomes in Addis Ababa. Bassi and Nansamba (2018) and Carranza, Garlick, Orkin and Rankin (2019) find that certifying worker skills affects labor market outcomes in urban areas of Uganda and in urban South Africa, respectively. Beam (2016) finds that job fairs improve employment outcomes by conveying information. Banerjee and Chiplunkar (2018) find that there is great scope for improving the process of matching graduates of an Indian vocational training institute to vacancies, even when it is already done by professionals. Blattman and Dercon (2018) show, again in Ethiopia, 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. All this work is highly complementary to this paper, and gives indications of the precise nature of frictions in urban labor markets in some specific poor countries.

The paper is organized as follows. The next section documents the joint relationship of wage employment, self-employment, unemployment and GDP per capita across countries. Section 3 contains a simple accounting analysis that identifies potential drivers of the cross-

[9]Rud and Trapeznikova (2021) also measure flows between self-employment and wage employment in six Sub-Saharan African countries. Their data, however, mostly covers rural areas, where most self-employment is in agriculture (p.45).

country patterns. Section 4 presents the economic model. Quantitative results are shown in Sections 5 to 7. Section 5 describes the calibration of the model economy using data from eight countries. Section 6 identifies the main quantitative determinants of cross-country differences in wage employment, unemployment and self-employment, and Section 7 analyzes the effects of labor market frictions on unemployment, self-employment and productivity in more detail. Section 8 concludes. Appendices contain additional figures and tables, and additional details on theory and numerical methods.

2 Wage employment, unemployment and self-employment across countries: Evidence

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

2.1 Data sources and measurement

My main source of data consists in the censuses available via the International Integrated Public Use Microdata Surveys (IPUMS International, Minnesota Population Center (2017)). IPUMS International provides access to micro data from almost 200 censuses collected in more than 60 economies since 1960. This data source has very broad coverage, both in terms of countries and in terms of individuals within each country. It allows computing measures of wage employment, self-employment and unemployment not only for the aggregate economy, but also for subgroups (like urban residents, young workers, etc.) for many countries. My main sample consists of urban residents of both sexes between the ages of 20 and 65.[10] Income per capita throughout is in 2011 US dollars, converted at PPP, from the Penn World Tables 9 (Feenstra, Inklaar and Timmer 2015), computed using the variables rgpde and pop.

My definitions of the states of wage employment, self-employment and unemployment follow those in the UN System of National Accounts (United Nations 2008). Employees, or the wage-employed, receive remuneration for their labor. The self-employed include both employers and own-account workers. "An unemployed person is one who is not an employee or self-employed but available for work and actively seeking work." (ibid., p.408.)

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

In the IPUMS Census data, individuals can be classified into these three categories using the harmonized EMPSTAT and CLASSWK variables. EMPSTAT (employment status) classifies individuals as employed (including both wage and self-employed), unemployed, or inactive. Typically, those who worked at least one hour in the reference period, including informal work or day labor, are considered employed. The union of the employed and the unemployed constitutes the labor force. In almost all censuses, CLASSWK (class of worker) further categorizes the employed as either self-employed, wage or salary workers, unpaid workers, or other, according to their main job. For the self-employed, most censuses distinguish employers and own-account workers.

These classifications mirror the UN definitions. The only concern regarding comparability comes from the fact that the reference period for job search used to classify a respondent as unemployed varies across censuses, and occasionally is not specified. Therefore, I group the censuses into quality tiers, in a way similar to Feng et al. (2018). 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.

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). Most importantly, in poor countries, many workers outside urban areas work in agriculture. The land distribution in these countries implies that this sector is dominated by self-employment on family farms (see e.g. Adamopoulos and Restuccia 2014), and the main occupational choice is self-employment in farming versus non-farming, with only a small role for wage employment (see e.g. Alvarez-Cuadrado, Amodio and Poschke 2019). 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. I report results for the entire country when it is informative.[11]

At the country level, the sample used in the analysis consists of 214 censuses covering 68 countries. The unemployment rate is available for an additional 21 censuses from 9 countries. For urban areas, the sample consists of 150 censuses covering urban areas in 58 countries.

[11] 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 employment is also much more limited than that reporting urban versus rural status, at 88 compared to 150.

The unemployment rate is available for an additional 15 censuses from 7 countries.

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 labor force status and development

Figure 1 depicts the prevalence of different types of labor force status in urban areas by country log income per capita. 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. In urban areas, they account for a very small share of the labor force even in the poorest countries.

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 the country income distribution very clear.

It is immediate from the figure that wage employment is much less common in poor countries. Wage employment rates range from about 40% of the labor force in urban areas of the poorest countries to over 80% in the richest ones. The self-employment rate, in contrast, is much higher in poor countries, echoing the well-known finding of Gollin (2007). Self-employment rates range from almost 50% 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.[12]

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 13 in the Appendix). The unemployment rate does not vary systematically with log income per capita, whereas the wage 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

[12]This mirrors findings in Caselli (2005) and Donovan et al. (2019).

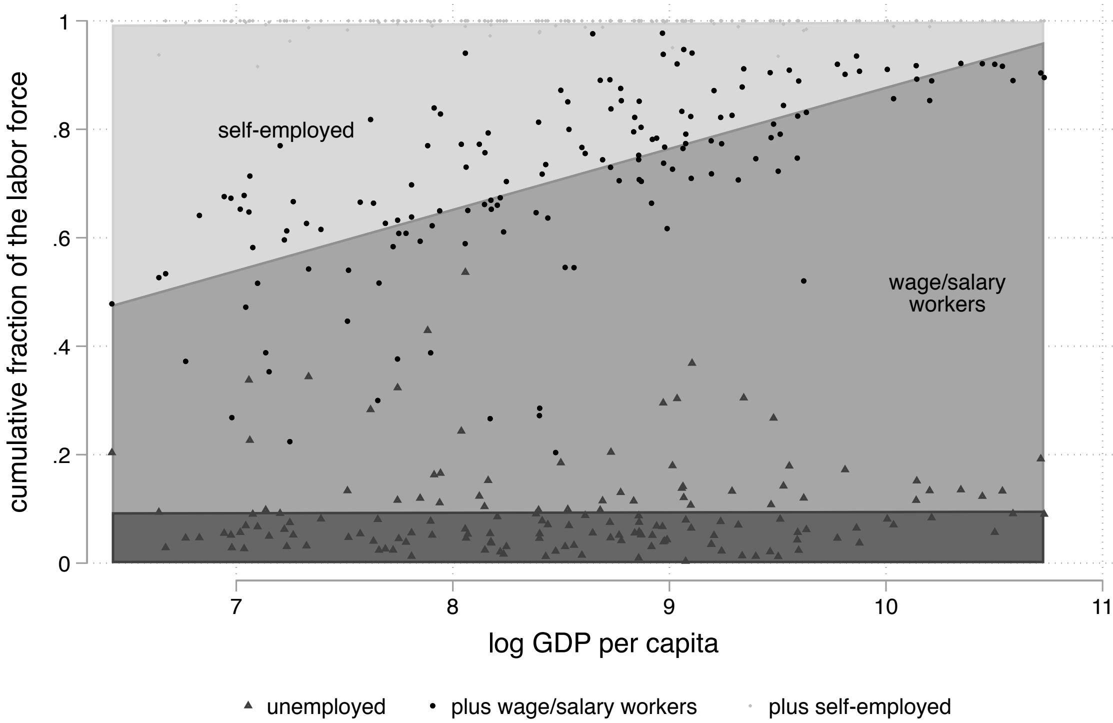


Figure 1: Composition of the labor force and development

Sources: GDP per capita: PWT 9.0. Employment status: IPUMS International. 150 censuses covering 58 countries over the years 1960 to 2011. Data for urban areas. Bottom area: unemployment rate.

wage employment rate increases by roughly the same amount. This translates into a decline in the self-employment rate, and an equivalent increase in the wage employment rate, by 9 percentage points every time income per capita doubles.

The middle panel of the table shows that regression results for the entire country are mostly similar, with even larger coefficients in absolute terms. Figure 7 shows results for the entire country graphically. The only difference compared to the urban results consists in the finding that for the entire country, the unemployment rate increases weakly with log GDP per capita. The bottom panel shows that this coefficient increases a bit further when using data for the entire country, but only for the countries in the top panel. (Coefficients for the other regressions remain very similar.) Hence, the difference between urban and national results cannot be attributed to differences in the sample. The coefficient for unemployment is close to that found by Feng et al. (2018) using 199 surveys and censuses from 55 countries.[13]

[13]Using data from 29 countries, Feng et al. (2018) find that unemployment also increases in log GDP per capita among urban workers with less than secondary education. In the larger sample used here (65 countries), this relationship is positive (with a coefficient of 0.018) but insignificant. It is negative (-0.011)

Table 1: Composition of the labor force and development

dependent variable:	wage employment rate	self-employment rate	unemployment rate	*UN* ratio
Urban areas:				
log GDP per capita	0.138	-0.132	0.003	-0.035
	(0.017)	(0.017)	(0.009)	(0.014)
R^2	0.543	0.507	0.002	0.099
observations	150	150	165	150
countries	58	58	65	58
Entire country, all countries:				
log GDP per capita	0.183	-0.187	0.012	-0.033
	(0.014)	(0.016)	(0.007)	(0.011)
R^2	0.718	0.670	0.041	0.121
observations	214	214	235	214
countries	68	68	77	68
Entire country, sample from top panel:				
log GDP per capita	0.204	-0.215	0.017	-0.035
	(0.018)	(0.020)	(0.009)	(0.014)
R^2	0.704	0.679	0.058	0.095
observations	150	150	150	150
countries	58	58	58	58

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. Data sources as in Figure 1. Results for a regression using pooled data are similar and are shown in Table 13.

Table 14 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

and insignificant for urban workers who completed secondary education or more. (More details available upon request.)

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. Data sources as in Figure 1. Results for a regression using pooled data are similar (not reported).

shows that the unemployment rate $u/(u+n+e)$ does not vary with income per capita in urban areas. (Let u denotes the unemployment rate, n the employment rate, and e the self-employment rate, as fractions of the labor force.) Yet, this invariance hides a systematic relationship: the denominator of the unemployment rate contains many wage employees and few self-employed in rich countries, but few wage employees and a large number of self-employed individuals in poor countries. That is, the reason why unemployment as a fraction of the labor force is not higher in poor countries despite low levels of wage employment consists in their high rates of self-employment.

In fact, there is a mechanical negative relationship between the self-employment rate and the unemployment rate: higher self-employment must reduce the unemployment rate, unless it arises from a one for one reduction in wage employment. These considerations imply that in a setting with significant self-employment, the unemployment rate, computed as a fraction of the labor force, captures the prevalence of unemployment, but does not accurately reflect the incidence of failed job search, i.e. how many people are searching for a job as an employee, but failing to find one.

An alternative measure of unemployment is the "*UN* ratio" $\tilde{u} \equiv u/(u+n)$. This is of course identical to the unemployment rate in a model without self-employment.[14] Since the *UN* ratio differs from the unemployment rate only in its denominator, it has a similar order

[14] In contrast to u, the stock accounting relationship between the *UN* ratio and the self-employment rate e is ambiguous and depends on what one assumes about the variable that adjusts with changes in e to ensure that $e+n+u=1$. If wage employment adjusts, the accounting relationship between $\tilde{u}$ and e is positive. If unemployment adjusts, it is negative. Sections 3 and 6 analyze the relationship between $\tilde{u}$ and e implied by cross-country differences in flow rates and structural parameters, respectively.

of magnitude. 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%).[15]

Since the unemployment rate does not vary systematically with GDP per capita, but poor countries have systematically lower wage employment, it is clear from Figure 1 that the *UN* ratio attains systematically higher values in these countries. This is corroborated by the regression coefficients in the last column 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.

Table 3 in the main text as well as Table 16 and Figure 8 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. The same is true for the fraction of the population that is not working (inactive plus unemployed). This implies that even if there may be some misclassification between unemployment and non-participation, the negative relationship between the *UN* ratio and GDP per capita appears very robust.[16] Finally, the relationships between the unemployment rate, the *UN* 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 15 in the Appendix shows that the relationships between the self-employment rate, the unemployment rate and GDP per capita are also similar in ILO data.

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, as captured by the higher *UN* rate. This failure to transform job seekers into employees could be due to limited hiring by firms, difficulties in search and matching, quick destruction of jobs, or any combination of these. All of these imply that job search is less

[15] The unemployment to wage employment rate u/n would have similar properties, and using it leads to similar results.

[16] The relationship established here differs from that in Bick et al. (2018), who find higher employment to population 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 countries with low participation rates from the IPUMS data. This is because these countries lack comparable hours data, which are the focus of the analysis in that paper.

Table 3: Unemployment and development, subsamples

dependent variable:	unemployment rate			*UN* ratio		
	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. Data sources as in Figure 1.

attractive in poorer countries, either because it is less likely to be successful, or because jobs, once found, do not last long.

2.3 Self-employment and unemployment

Less attractive job search could be expected to 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.

Figure 2 shows the bivariate relationship between the self-employment rate and the *UN* ratio, as a measure of the (un)attractiveness of search. 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 9 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.

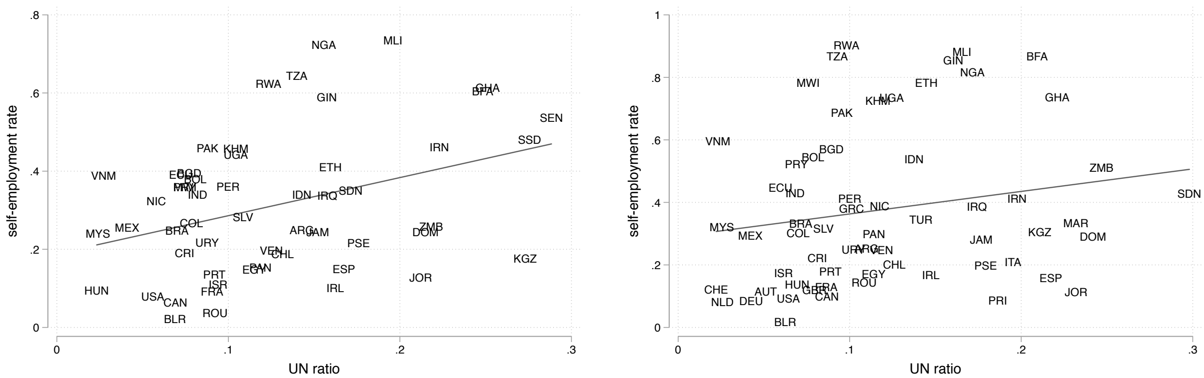


Figure 2: The self-employment rate versus the UN ratio $u/(u+n)$, urban (left) and overall (right)

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 9 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.

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 9. 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. An increase in 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 0.7 percentage points, due to an increase in the fraction of own-account workers by 0.8 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 17 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 18).

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 19 in the Appendix, and also Table 20 using ILO data). This is not surprising. When using data for the entire country, data for poor countries include many respondents in rural areas, where self-employment in small-scale agriculture is highly prevalent, and where there are few

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. Data sources as in Figure 1. Results for a regression using pooled data are similar (Table 17).

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 countries at different stages of development reveals three regularities: Labor markets in poor countries feature (1) systematically lower wage employment and higher self-employment rates, (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.

3 What drives differences in labor force status across countries? An accounting analysis

The previous section has documented very large differences in the composition of the labor force across countries. What can drive these differences? Before analyzing the data using a full economic model with optimal choices between the different labor market states, I perform a simple accounting analysis. The objective of this is to identify the differences in labor market *flow* rates across countries that are consistent with the observed patterns in *stocks.* This is useful, because in a dynamic model, the individual choices modelled explicitly in the next section map directly into flows, and therefore differences in flow rates can easily be linked to fundamentals that could be driving them.[17]

So, consider a labor market where individuals can be in any of the following three states: unemployment (U), self-employment (SE), or wage employment (N). In this section, I do not differentiate between own-account workers and employers, and treat them all as self-employed. Every period, individuals can transition across employment states at rates summarized in the matrix shown in Table 5: The unemployed enter self-employment at a rate h or, conditional on not doing so, find a job with probability f, the employed lose their job with probability s, and the self-employed close their firm and transition into unemployment with probability λ. For tractability, the analysis in the main text abstracts from flows from N to SE. Appendix B shows that results are similar when this flow is permitted. I also set the flow from self-employment to wage employment to zero. (See the next section for a discussion.)

Table 5: Accounting analysis: flow rates across labor market states

from/to	U	N	SE
U	$(1-h)(1-f)$	$(1-h)f$	h
N	s	$1-s$	0
SE	λ	0	$1-\lambda$

The table shows per period flow rates from the states in rows to those in columns.

These flows imply the following steady state stocks for the three labor market states:

$$\begin{aligned} u &= \frac{s}{s+(1-h)f+hs/\lambda} & \qquad e &= \frac{h}{\lambda}u = \frac{hs/\lambda}{s+(1-h)f+hs/\lambda} \\ n &= \frac{(1-h)f}{s+(1-h)f+hs/\lambda} & \qquad \tilde{u} &\equiv \frac{u}{u+n} = \frac{s}{s+(1-h)f} \end{aligned} \tag{1}$$

where u (e) [n] denotes the unemployment (self-employment) [wage employment] rate. Each state increases in its own inflow rate, decreases in its own outflow rate, and $u+n+e=1$. The equation for u looks very similar to the Beveridge curve well-known from the analysis of labor markets without self-employment. The presence of self-employment leads to an additional term in the denominator of the unemployment rate. The expression for the UN

[17] For a smaller number of countries, Donovan et al. (2019) document labor market flows directly. The findings of this section are similar when using these flows, as discussed below.

ratio, or $\tilde{u}$, comes close to the standard expression for the Beveridge curve, $s/(s+f)$. The only difference is the $(1-h)$ term in the denominator, which captures that larger flows from unemployment to self-employment (higher h) imply smaller flows from unemployment to wage employment.

Cross-country differences in any of the flow rates can generate differences in labor market outcomes. For example, high unemployment could be due to a low job finding rate or a high separation rate. The unemployment rate also depends on h. The self-employment rate depends on the balance between entry (at rate h) and exit (at rate λ), and on the size of the pool of origin, u.

The previous section provided evidence not only of dispersion in stocks, but also showed a significant and sizeable positive relationship between the self-employment rate e and the *UN* ratio $\tilde{u}$, with a regression coefficient of 0.7 (Table 4). This piece of information is informative about the type of variation in flow rates required to match cross country data. In particular, the remainder of this section shows that despite the dependence of both the self-employment rate and the unemployment rate on the entry rate into self-employment (h), variation in h alone generates a relationship between e and $\tilde{u}$ far from that observed in the data.

Cross-country variation only in the entry rate. Consider first a scenario where only the self-employment entry rate h varies across countries. Then, the model equivalents of the coefficients of the regressions of e on u and $\tilde{u}$, respectively, are given by

$$\left.\frac{de}{du}\right|_{\text{vary only } h} = \frac{s+f}{\lambda f - s} \tag{2}$$

$$\left.\frac{de}{d\tilde{u}}\right|_{\text{vary only } h} = \frac{(1-e)^2}{\lambda}\frac{s+f}{f}. \tag{3}$$

In this case, there should be a positive relationship between the self-employment rate and the *UN* ratio. The relationship between the self-employment and the unemployment rate is negative if individuals who enter self-employment stay out of unemployment longer than those who remain in unemployment and search for a job ($1/\lambda > f/s$), as is typical in the data.

What is the size of these model-implied relationships? Given that λ is generally close to 1% at a monthly frequency, so that λf is negligible compared to s, de/du is approximately $-(s+f)/s$. This is minus the inverse of the steady state unemployment rate in a model without self-employment, and therefore it is generally on the order of minus 5 to minus 33 (for unemployment rates ranging from 3 to 20%).[18] This large number of course reflects

[18] These numbers are similar when computed using the expression in flow rates and data from Donovan

that if only self-employment entry differs across countries, differences in the entry rate affect self-employment much more strongly and directly than unemployment, and therefore de/du is much larger than 1. This is even more pronounced for $de/d\tilde{u}$. Reflecting the fact that changes in h hardly affect the UN ratio, this is on the order of 50 to 300.

The equivalent of Figure 2 from this simple accounting model for the case where countries differ only in the self-employment entry rate h would thus feature a near-vertical line of best fit – implying essentially no relationship between the self-employment rate and the UN ratio. Clearly, this does not even come close to the regression coefficient of 0.7 found in the data. Hence, variation only in self-employment entry – due for example to differences in the cost of entry or the regulatory burden – cannot account for the cross-country data.

Cross-country variation only in the job finding rate. If instead only f varies across countries, the model equivalents of the coefficients of the regressions of e on u and $\tilde{u}$, respectively, are given by

$$\left.\frac{de}{du}\right|_{\text{vary only } f} = \frac{h}{\lambda} = \frac{e}{u} \tag{4}$$

$$\left.\frac{de}{d\tilde{u}}\right|_{\text{vary only } f} = (1-e)^2 \frac{de}{du}. \tag{5}$$

With self-employment rates between 10 and 40% and unemployment rates between 5 and 20%, these objects take values of 2 to 3.3 and 0.7 to 2.7, respectively. As a result, variation only in f comes much closer to matching the empirical relationship between the self-employment rate and the UN ratio. However, it predicts a counterfactual positive relationship between the self-employment rate and the unemployment rate.

Differences in s only have the same implications as differences in f only. Differences in λ only lead to variation in e but not $\tilde{u}$, and thus do not help account for data patterns.

This analysis assumed that flows across labor market states are exogenous and independent of each other. In the model presented in the next section, both the self-employment entry rate h and the job finding rate f will be endogenous objects and functions of fundamentals, like the strength of labor market frictions or the ease of entry, which can in turn vary across countries. One can already anticipate their joint variation: Anything that makes self-employment entry easier will tend to raise h. If some of these new firms hire workers, f could in turn increase. As a result, if the most important variation across countries was in factors primarily driving self-employment entry, h and f should be positively correlated.

et al. (2019). The same is true for all other model predictions reported in the remainder of this section.

This would exacerbate the problems with the model $de/d\tilde{u}$ illustrated above. In turn, anything that reduces the job finding rate f should raise the self-employment entry rate h, as some job seekers find it more attractive to start a firm rather than engage in now lengthier search for a job. As a result, if countries mostly vary in factors determining job finding rates, like labor market frictions, h and f should be negatively correlated. Such a situation would allow for values of de/du and $de/d\tilde{u}$ more in line with the values observed in the cross-country data.

To summarize, the accounting analysis shows that variation in both the self-employment entry rate and the job finding or destruction rate is required to account for the joint variation of labor market states observed in the cross-country data. This conclusion is in line with evidence from Donovan et al. (2019), who find both higher entry rates into self-employment and lower job finding rates in poor countries. The theoretical and quantitative analysis in the remainder of the paper goes beyond flow rates and takes a step to determine which fundamentals drive the observed patterns.

4 A model of frictional labor markets with endogenous entry into self-employment and entrepreneurship

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 a version of the Diamond-Mortensen-Pissarides (DMP) model of random search and matching in labor markets with firms that differ in size and productivity. Compared to a standard DMP model, 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 in their productivity, so that some entrants become own-account workers, while others become 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 wage employment, 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.[19] 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.

4.1 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: 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

[19] 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.

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 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.[20]

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

$$e_s = \frac{(1-\delta)h(1-\phi)p_s}{\tilde{\lambda}_s}u \tag{6}$$

and

$$e_f = \frac{(1-\delta)h(1-\phi)p_f}{\tilde{\lambda}_f}u, \tag{7}$$

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*

[20] 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 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 (2018) 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.

curve (MBC)

$$u = \frac{(1 - e_f - e_s)s + e_f\tilde{\lambda}_f + e_s\tilde{\lambda}_s}{s + (1 - \delta)(1 - h)(1 - \phi)\theta q + (1 - \delta)(1 - \phi)h(p_f + p_s)}. \tag{8}$$

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

$$u = \frac{s}{s + (1 - \delta)(1 - h)(1 - \phi)\theta q + (1 - \delta)h(1 - \phi)(p_f + p_s)s/\tilde{\lambda}_f}. \tag{9}$$

Like equation (1) in Section 3, this expression is similar 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 not identical 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 (8), and results in the final fraction in the denominator in equation (9). If the flow rate into unemployment is lower for entrepreneurs than for employees, then a larger entrepreneurship rate tends to reduce unemployment. Finally, the measure of employees follows as

$$n = 1 - u - e_s - e_f. \tag{10}$$

Who can search? One model assumption 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 (in the same month). Nor can they transition directly from self-employment to wage employment.

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 in terms of both time and money (Abebe et al. 2021, Carranza et al. 2019).

Abebe et al. (forthcoming) 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. The self-employed in contrast report working an average of 50 hours per week, similar to employees. Self-employment thus truly appears to be a distinct activity from job search. 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 (Abebe et al. 2021). This does not require the worker to have capital.

Nevertheless, there are some flows from self-employment to wage work in the data. Donovan et al. (2019, Figure B1) document quarterly flow rates of about 10% for a broad set of countries. (These rates are lower in some rich countries, namely Great Britain, Cyprus, Iceland, Lithuania and Malta.) In principle, on the job search by the self-employed could be allowed for in the model to accommodate these flows. I abstract from it mostly because it would be hard to discipline this aspect of the model in several of the countries used in the calibration, in particular the poorest ones. Moreover, the model does allow for indirect transitions between self-employment and wage employment over time. A quantitative evaluation of the model predictions in this dimension in Section 5 shows that the restriction of job search to the unemployed does not appear to be too restrictive.

4.2 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 firm's 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

$$F_s(z) = \zeta z + \frac{(1-\phi)(1-\lambda_s)}{1+r} F_s(z) + \frac{(1-\phi)\lambda_s}{1+r} U \tag{11}$$

$$F_f(z) = zn(z)^\gamma - wn(z) - \frac{k_v}{q}\hat{\xi} n(z) + \frac{(1-\phi)(1-\lambda_f)}{1+r} F_f(z) + \frac{(1-\phi)\lambda_f}{1+r} U \tag{12}$$

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. These departures occur at a rate $\hat{\xi} \equiv \xi + (1-\xi)\phi$.

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 $G(z)$.[21] 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 3.) It is clear that the value of unemployment, U, is independent of z. It is also clear from equation (11) 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.[22] (The linear hiring costs arising from labor market frictions imply that it is optimal for firms to move to optimal employment directly upon entry.) Continuation values as a function of z are as depicted in Figure 3. 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

$$\tilde{g}(z) = \frac{g(z)}{1 - G(z_f)}, \quad z \geq z_f, \tag{13}$$

[21]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.

[22]Convexity reflects the ability of employers to leverage their productivity z by hiring workers accordingly.

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

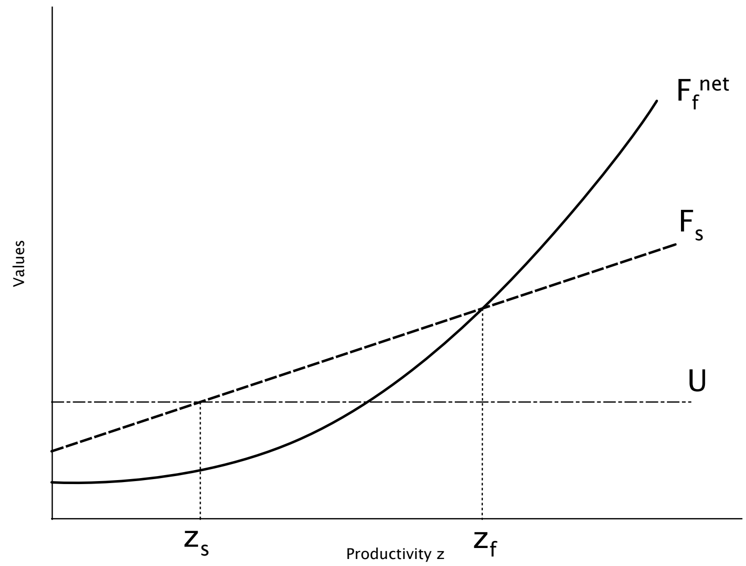


Figure 3: The values of unemployment (U), self-employment (F_s), and the value of being an employer net of hiring costs at entry ($F_f^{\text{net}}(z) = F_f(z) - n(z)k_v/q$), with associated productivity cutoffs

Combining these possibilities, the value of entry is given by

$$Q = \frac{1-\phi}{1+r}\left[-k_f + \int \max\left(F_f(z) - \frac{k_v}{q(\theta)}n(z),\ F_s(z),\ U\right)\mathrm{d}G(z)\right] \tag{14}$$

I now turn to workers and the unemployed.

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

$$W = w + \frac{1-s}{1+r}W + \frac{s-\phi}{1+r}U. \tag{15}$$

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

$$S = b + \frac{1-\phi}{1+r}\left[\theta q W + (1-\theta q) U\right] \tag{16}$$

$$\underline{U} = b + \frac{1-\phi}{1+r} U. \tag{17}$$

With occupational choice, the value of unemployment is given by

$$U = \delta \underline{U} + (1-\delta)\max\left(S, Q\right). \tag{18}$$

Since workers are ex ante identical, $S = Q$ in an equilibrium with entry. If this holds, an endogenous fraction h of the unemployed start a firm. In the following, I focus on such an equilibrium.[23]

4.3 Wage determination and vacancy posting

Upon matching, a firm and a worker bargain over the wage. Like Cahuc, Marque and Wasmer (2008) 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.[24] Wages are bargained upon hiring, and remain constant thereafter. Then it can be shown (see Appendix C.2 for a detailed derivation) that

$$w = \frac{r+\phi}{1+r} U + \frac{\eta}{1-\eta}\left[1 - \frac{(1-\phi)(1-\lambda_f)}{1+r} + \hat{\xi}\right] \cdot \frac{k_v}{q\left(\theta\right)}. \tag{19}$$

Three remarks are in order. First, the wage curve given by equation (19) 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.

[23]An equilibrium with only own-account workers could arise if the relative productivity of own-account workers is high. I abstract from this equilibrium for lack of empirical relevance for urban labor markets.

[24]See Stole and Zwiebel (1996) and Bruegemann, Gautier and Menzio (2019) for the game-theoretic foundations of this assumption.

A firm's optimal employment is given by

$$n\left(z\right)=\left(z\gamma\right)^{\frac{1}{1-\gamma}}\left\{\left(\eta(\gamma-1)+1\right)\left[\left(1-\frac{(1-\phi)(1-\lambda_f)}{1+r}+\hat{\xi}\right)\frac{k_v}{q}+w\right]\right\}^{\frac{-1}{1-\gamma}}. \qquad (20)$$

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 $\hat{\xi}$ per period, and thus need to post $\hat{\xi}n(z)/q$ 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 (7), new entrants account for a fraction $\tilde{\lambda}_f$ of employers. As a result, total vacancies in a stationary equilibrium of this economy are given by

$$v=\frac{\tilde{\lambda}_f+(1-\tilde{\lambda}_f)\hat{\xi}}{q}\,e_f\int n(z)\tilde{g}(z)\mathrm{d}z. \qquad (21)$$

4.4 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 (11) to (12) and (14) to (18),
2. households are indifferent between occupational choices: $Q = S$,
3. wages fulfill equation (19),
4. the equilibrium distributions are generated by household choices and are stationary, according to equations (6) to (10) and (13),
5. firms post vacancies optimally (equations (20) and (21)), 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.

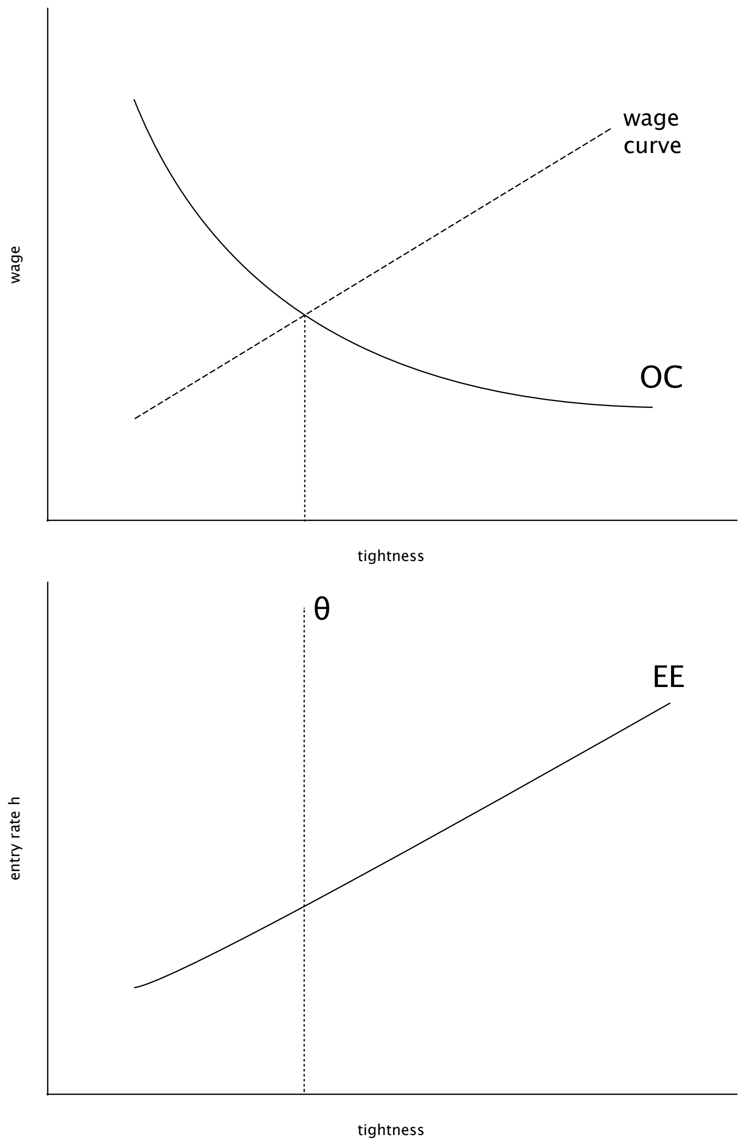


Figure 4: A sketch of equilibrium determination

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 (7), (9) and (21).

Figure 4 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 (19), is familiar 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 (21) and thus market tightness depend on the entry rate, which enters equation (21) via equations (7) and (9) (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 costlier 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 4. 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. 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 from the MBC 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 and self-employment both decline. To quantify all effects, I next turn to an empirically guided examination of the quantitative properties of the model.

5 Calibration

In the remaining sections of the paper, I analyze the quantitative properties of the model, and assess its ability to account for the joint cross-country variation in wage employment, 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 wage employment, 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.

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.[25] 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. These wide ranges reflect the variation of these variables in the full cross-country data.

The calibration requires setting parameter values for 16 parameters for each country. Two of these allow for normalizations. I then proceed by calibrating five parameters externally, i.e. setting them to values commonly used in the literature. The remaining nine parameters, which include all parameters related to labor market frictions as well as firm entry costs, are calibrated internally to match a set of informative country-specific data moments. As a benchmark for the analysis below, I also calibrate the model to an "average" economy,

[25] 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. (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.)

described by average values of all target moments.

Externally set parameters and normalizations. 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 each model economy can be normalized. 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.

Moments for the internal calibration. The remaining nine parameters are calibrated internally to match a set of nine target moments. The parameters are the entry cost k_f, the vacancy posting cost k_v, workers' bargaining power η, the utility flow of unemployment b, the firm exit rate λ_f, the match separation rate ξ, the standard deviation of the (log) productivity distribution, σ_z, the relative productivity of own-account workers ζ and the probability of having to take a casual job, δ. Denote this set of parameters by Γ.

Calibrating these parameters requires using statistics on the structure of employment, on some flows between different labor force statuses, and on the firm size distribution. I use the following nine moments: 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. While the wage employment rate is not targeted directly, it is implied by the unemployment and self-employment rates.

The values for these data moments in each country are shown in Table 21. I next describe their sources, beginning with unemployment outflow rates. 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 D. 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 latest available census for each country. For Ethiopia, they are taken from the UEUS. For the US, they are computed using information from Hipple (2010).[27]

Information on the concentration of employment is from Poschke (2018) for most countries, from Berry, Rodriguez and Sandee (2002) for Indonesia, from Bartelsman, Haltiwanger and Scarpetta (2004) for Mexico, and from the UEUS for Ethiopia. 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.[28] The firm exit rate is from Bartelsman et al. (2004) for most countries, and from Bigsten, Mengistae and Shimeles (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

[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. For the US, where evidence for duration dependence is strong, there is a larger discrepancy, and I use the figure from Shimer (2012). Alternative sources for unemployment outflow rates for some countries are Elsby et al. (2013, EHS) and Donovan et al. (2019, DLS). They report the following rates: USA 0.565 (EHS), 0.3 (DLS), Germany 0.06 (EHS), France 0.077 (EHS), 0.15 (DLS), Italy 0.043 (EHS), Mexico 0.43 (DLS). Overall, these rates are close to those used in the calibration (shown in Table 21).

[27] For the fraction of employers in Ethiopia, there is a discrepancy between the Census and the UEUS figure, which contains more detailed information on firm employment. I target the average of the two values.

[28] The measure that is available differs slightly by country. It is the share of firms with less than 10 employees for Ethiopia, and the share of employment in firms with at least 20 employees in Mexico and Indonesia/at least 10 employees in all other countries. I compute the average value of this moment for the "average" country calibration using the model-predicted share of employment in firms with at least 10 employees for Mexico, Indonesia and Ethiopia.

Indonesia and Mexico, where it is reported as a type of employment. 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.[29]

Note that the calibration does not use direct information on job destruction rates, since it is not available for the poorest countries. Where possible, calibrated job destruction rates are compared to their data counterparts below.

Calibration loss function and identification. I calibrate the model by finding, for each country i, the set of parameters that minimizes the equally weighted sum of squared differences between model and data moments, $\mathcal{C}_i(\Gamma)$. Let this set of parameters be $\hat{\Gamma}_i$.

In the following, I discuss heuristically why the selected data moments are informative for the parameters in Γ.[30] First, moments related to unemployment: Given a productivity level of the matching function and wages, the vacancy posting cost determines employers' hiring efforts (equation 20), and thus the job finding rate. 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 the firm exit rate, the level of the unemployment rate then identifies the job destruction rate ξ from the MBC.

A second set of moments relates to self-employment and entrepreneurship. Here, I set 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. Own-account workers account for the remainder of the self-employed. Their fraction of employment ranges from 4% in Germany to over 30% 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 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

[29] For assessing the importance of labor market frictions, assuming common b/w is conservative. The most plausible alternative is that b/w is lower in poor countries, since they do not provide unemployment insurance benefits. In this case, even larger labor market frictions, either in the form of higher k_v or higher job destruction ξ, would be required to match their unemployment rates.

[30] Targets have to be matched jointly by setting all parameters. Nevertheless, each parameter affects some targets more strongly than others. This discussion focuses on these relationships.

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.

Finally, given unemployment, the rate of casual employment in an economy directly identifies δ. The labor income share is informative about workers' bargaining power η, and the ratio b/w pins down the flow value of unemployment b.

Calibration results. To save space, I do not report all calibration results and parameters in the main text – see Tables 21 and 22 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.

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	0.538
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:		
Wage employment	0.505	0.504
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	

Table 6 shows the model fit for Ethiopia. It is overall very close. The table also shows model predictions for some non-targeted moments. The entrepreneurial entry rate from unemployment is 1.4% 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.

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 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 D for details. This is the most recent flow information for urban Ethiopia. Note that it combines 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). The fact that the model produces realistic transition rates between entrepreneurship and wage employment implies that the restriction that only allows this flow to be indirect appears not too limiting.[31]

[31] The implied annual transition rate from self-employment to wage employment is 1.8%, slightly above that found by Rud and Trapeznikova (2021) for the entire country. 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,

Table 8: 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 21 for details.)

Table 8 compares model moments and parameters for Ethiopia to those for the US and for the average economy. Target moments are almost identical to model moments and are shown in Table 21. Table 8 shows a subset of five targeted moments and five parameters, to stress five salient differences across the countries. (Tables 21 and 22 contain all other moments and parameters, as well as those for the other economies.) First, vacancy posting costs (relative to productivity) are very high in Ethiopia, and very low in the US. This is the reason for the low unemployment outflow rate in Ethiopia, and the first reason for its the high unemployment rate. Second, the job destruction rate is high in Ethiopia relative to the US. This is the second reason for the high unemployment rate in Ethiopia. Third, the entry cost (relative to productivity) is low in Ethiopia, and high in the US. This contributes to the higher self-employment rate in Ethiopia. It is also in line with the findings of Bollard, Klenow and Li (2016). 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

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.

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 employment in large firms.[32]

It should be noted here that the presence of 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 as a low level of σ_z. This is clear if one models SDDs as productivity-specific taxes on firm revenue.[33] A popular specification assumes that firm revenue is taxed at a rate τ such that $1-\tau(z) = (z/\bar{z})^{-\nu}, \nu \geq 0$. (See e.g. Buera and Fattal-Jaef (2016).) The parameter ν controls the "progressivity" of SDDs, and the constant $\bar{z}$, together with ν, the average level. With this tax function, an employer's optimal labor demand is proportional to $z^{\frac{1-\nu}{1-\gamma}}$. From this, it is clear that the allocative consequences of a reduction in the standard deviation of $\log z$ can be replicated exactly by an increase in ν. As a consequence, ν and σ_z cannot be identified separately in the calibration. Therefore, 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. 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.

Non-targeted moments. 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 several flows in the model in addition to those for Ethiopia shown in Table 7 that were discussed above. First, the total job separation rate 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 OECD members in the set of calibration countries. The correlation between model-implied separation rates and empirical ones is above 0.9. Flows from wage employment to unemployment implied by the model are also close to Donovan et al.'s (2019) measures for France and the US, but exceed them for Mexico.

Second, self-employment entry rates from unemployment can be compared to those reported by Donovan et al. (2019, Figure B1) for the countries present in both papers: France,

[32]Table 22 shows wide variation in parameters across the calibration countries. Calibrated vacancy posting costs are lowest in the US and Mexico, reflecting their high unemployment outflow rates, and high in Ethiopia, Indonesia, and continental Europe, reflecting lower outflow rates. Calibrated job destruction rates are low in continental Europe, reflecing levels of unemployment that are relatively low given the low outflow rates. Entry costs mostly fall between those for the US and Ethiopia.

[33]The modeling device of size- or productivity-specific taxes can capture both factors like a higher burden of taxes and regulation 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 (2021) and Grobovšek (2020)). Financial frictions also constrain more productive firms more, at least for a given amount of assets (see e.g. Cagetti and De Nardi (2006) and Buera (2009)).

the United States and Mexico. The calibration is very close to its data counterparts, with flow rates of slightly more than one percent per quarter for France and around 3 percent for the US in both model and data, and 7 percent (model) versus 10 percent (data) for Mexico.

6 Which factors drive cross-country differences in wage employment, unemployment and self-employment?

Which factors drive the cross-country patterns in wage employment, self-employment and unemployment shown in Section 2? To answer this question, I conduct two decomposition exercises. In both cases, the starting point is the average country calibration, with parameters $\hat{\Gamma}_{\text{avg}}$.[34] By construction, this does not minimize the calibration loss functions $\mathcal{C}_i(\Gamma)$ for the other countries, but it provides a common benchmark.

Both exercises then use joint variation across model outcomes in reaction to parameter changes to assess the relative importance of different parameters. The first exercise evaluates, for each parameter and country i, how $\mathcal{C}_i$ changes when the parameter is chosen to match the country's wage employment rate. The second exercise takes the opposite perspective: it evaluates which individual parameter, or small set of parameters, contributes most to minimizing $\mathcal{C}_i$ when the other parameters are kept as in $\hat{\Gamma}_i$.

Both exercises also indicate what theories are plausible candidates for explaining the cross-country variation in labor market outcomes. For instance, a large explanatory power (large reduction in $\mathcal{C}$) due to the cost of entry would support theories emphasizing variation in the cost of entry, whereas an important role of the labor market parameters would support theories stressing labor market frictions. While the model abstracts from financial frictions, an important role for entry costs or the relative productivity of employer firms would indicate their importance.

6.1 Counterfactual 1: Fitting wage employment shares

For the first exercise, I find, for each country and separately for each internally calibrated parameter, the value of the parameter that lets the model match the country's wage employment rate from the data, while keeping all the remaining parameters as in $\hat{\Gamma}_{\text{avg}}$. For country i and parameter j, let this set of parameters be $\tilde{\Gamma}^{i,j}_{\text{avg}}$. This parameter change may improve or worsen the fit of the model in other dimensions. I therefore evaluate how the calibration loss function $\mathcal{C}_i$ changes as each parameter is used, in turn, to match each country's wage em-

[34] A similar approach would be to start from the calibration for a particular country, e.g. the US.

ployment rate. This exercise gives a first indication as to which parameters might plausibly account for variation in wage employment rates, and which have counterfactual implications in other dimensions.

Results from this exercise are reported in Table 9. The first column shows the improvement in the fit of the model for each parameter, measured as

$$1 - \frac{\sum_i \mathcal{C}_i(\tilde{\Gamma}^{i,j}_{\text{avg}})}{\sum_i \mathcal{C}_i(\hat{\Gamma}_{\text{avg}})}. \tag{22}$$

This measure is one for a perfect fit to the calibration targets in all countries, between zero and one for an improvement relative to the fit of the model when $\hat{\Gamma}_{\text{avg}}$ is used in all countries, and negative if fitting wage employment rates worsens the fit in other dimensions. Subsequent columns show the improvement in the model's explanatory power in terms of individual variables of interest. For each variable m, this is computed as

$$1 - \frac{\sum_i [m(\tilde{\Gamma}^{i,j}_{\text{avg}}) - \bar{m}_i]^2}{\sum_i [m(\hat{\Gamma}_{\text{avg}}) - \bar{m}_i]^2}, \tag{23}$$

where $m(\Gamma)$ is the model-implied value of the moment for parameters Γ, and $\bar{m}_i$ is the data moment for country i.

Table 9: Improvement in explanatory power of the model when only one parameter is country-specific – set to match each country's wage employment rate

	Improvement in fit of			
	all calibration targets	unemployment outflow rate	unemployment rate	self-employment rate
Country-specific parameter:				
k_f	-0.259	-0.109	-0.413	0.663
k_v	0.224	0.162	0.230	0.903
η	-0.525	-0.400	-0.870	0.312
b	-2.173	-0.288	-0.826	0.837
λ_F	-0.135	0.053	0.179	0.234
ξ	0.113	0.024	0.345	0.686
σ_z	-0.117	0.115	-0.059	0.590

Notes: The first column reports the reduction in the sum of the calibration loss statistic for all eight countries when one parameter is chosen to match each country's rate of wage employment, relative to the sum of loss statistics when the parameters for the calibration for the average target are used in all countries (equation 22). The subsequent columns report the reduction in the sum of squared deviations between model predictions and data values for the indicated statistics (equation 23).

The first column shows that the calibration fit only improves when the vacancy posting cost k_v or the match destruction rate ξ are used to match the wage employment rate. When k_v is set to match each country's wage employment rate, the model fit noticeably improves, with a decline in the sum of calibration loss statistics across the eight countries of 22%. The last three columns show why fitting wage employment rates using k_v improves the model fit: apart from the perfect fit to each country's wage employment rate, it leads to a closer fit of the unemployment outflow rate, the unemployment rate, and the self-employment rate.

It is generally the case that matching each country's wage employment rate also results in a closer match to the self-employment rates. The same cannot be said for the unemployment rate. For example, matching each country's wage employment rate by changing the entry cost k_f is feasible. However, doing so requires very high entry costs in countries with high wage employment. These high entry costs depress self-employment and promote wage employment – but at the cost of also reducing the number of employers, and thus raising unemployment. Hence the worse fit of the unemployment rate for this scenario. Matching wage employment rates with country-specific k_v instead requires reducing k_v to achieve high wage employment. This also results in lower self-employment and a lower UN ratio, in line with the data.

6.2 Counterfactual 2: Which parameters drive the overall fit?

In a second approach to gauging the relative importance of the different model parameters, I assess their relevance for the overall model fit in the calibration. To measure this, I again start from the average country calibration, and then recalibrate the model for each country, keeping all parameters as in $\hat{\Gamma}_{\text{avg}}$, except for one or a combination of few parameters. That is, for each country i and parameter (combination) j, I find the set of parameters $\hat{\Gamma}^{i,j}_{\text{avg}}$ that minimizes $\mathcal{C}_i(\Gamma)$. I then compute the improvement in the overall model fit, measured as in equation (22), as well as that in various outcomes of interest, measured as in equation (23).

Results for this exercise are shown in Table 10. The first column shows the improvement in model fit when one, two, or three parameters are country-specific. By construction, letting all internally calibrated parameters adjust would allow the model to fit all countries perfectly, implying an improvement measure of one. The remaining columns show the improvement in the model's explanatory power in terms of individual variables of interest.

It is 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 the case with common parameters $\hat{\Gamma}_{\text{avg}}$. Additionally allowing for the job destruction rate ξ

Table 10: Improvement in explanatory power of the model when only a subset of parameters is country-specific – set to improve each country's calibration

	Improvement in fit of					
	all calibration targets	wage employment rate	unemployment outflow rate	unemployment rate	UN ratio	self-employment rate
One country-specific parameter:						
k_f	0.173	0.687	0.099	-0.075	0.143	0.701
k_v	0.438	0.418	0.715	0.306	0.370	0.105
η	0.118	-0.142	0.209	0.213	0.117	-0.141
b	0.124	0.034	0.167	0.003	-0.013	0.224
λ_f	0.065	0.426	0.001	0.100	0.202	0.315
ξ	0.190	0.879	0.021	0.284	0.413	0.883
σ_z	0.159	0.730	0.079	0.019	0.204	0.591
ζ	0.138	0.623	-0.017	-0.113	0.003	0.915
Two country-specific parameters:						
k_v, ξ	0.708	0.748	0.939	0.191	0.336	0.808
Three country-specific parameters:						
k_v, b, ξ	0.915	0.787	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 (equation 22). The subsequent columns report the reduction in the sum of squared deviations between model predictions and data values for the indicated statistics (equation 23).

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.

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.[35] 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_f.

To a lesser extent, the same is true for productivity dispersion σ_z. Country-specific dispersion achieves an improvement in the calibration targets jointly of not much more than 1/8. It does lead to a much improved match of wage employment rates to the data. This is because a reduction in dispersion, because of selection, implies smaller employer firms and thus reduces wage employment. However, the counterpart to this is a strong increase in self-employment, with the implication that the unemployment rate hardly changes, and UN rates change little. This finding is important since, as mentioned above, changes in σ_z are isomorphic to changes in size-dependent distortions, parameterized in a popular way. SDDs thus do not appear to be a primary driver of variation in labor force status across the calibration countries.

Figure 5 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 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

[35] This exercise is similar to the one in Feng et al. (2018), who analyze the effect of an increase in productivity in a "modern" sector of employer firms relative to that in a "traditional" sector of own-account workers.

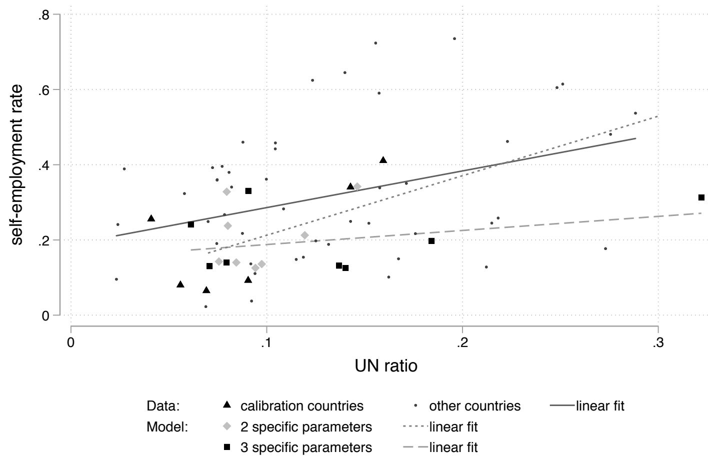


Figure 5: Self-employment and the *UN* ratio: data and model outcomes

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. Regression coefficients: 0.881 with 2 specific parameters, 0.374 with 3 specific parameters.

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 0.881 for the case with 2 specific parameters, and 0.374 with 3 specific parameters.

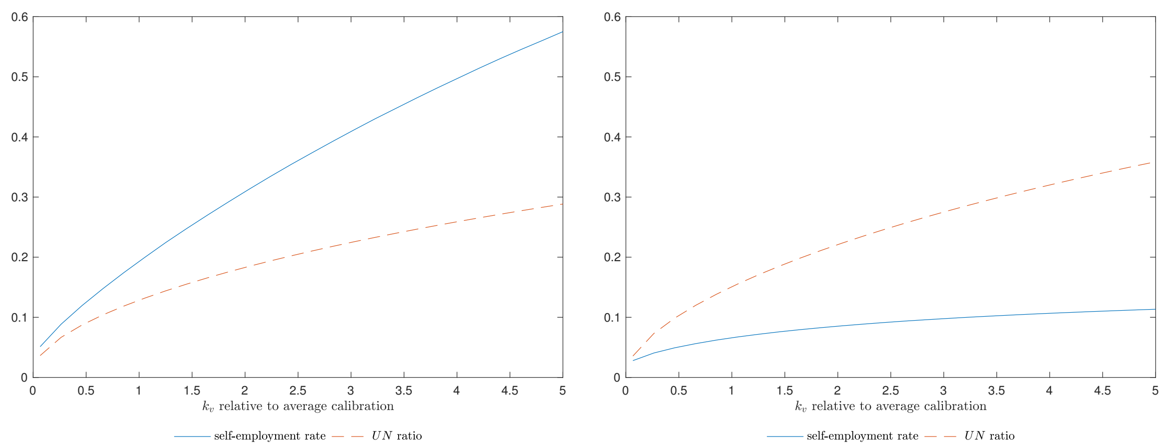
It is clear from the lines of best fit that the model outcomes are qualitatively in line with the data. Quantitatively, the model captures at least one third of the strength of the relationship between the self-employment rate and the *UN* ratio in the data.

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 wage employment, self-employment and the *UN* ratio individually – implying variation in the wage employment and self-employment rates across countries of more than 20 percentage points –, but also the bivariate relationship of self-employment and the *UN* ratio. 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 and wage employment.

7 Labor market frictions, employment status, and productivity

Having shown the importance of labor market frictions in accounting for cross-country differences in labor market outcomes, I illustrate their effects in more detail in this section. 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?

Figure 6 shows the effect of changes in k_v on the self-employment rate and the UN ratio. It is clear from Figure 6a 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.



(a) Low k_f (from average country calibration)

(b) High k_f (from US calibration)

Figure 6: 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 8 and 22 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.

Which margin reacts more strongly depends on parameters, in particular the cost of establishing new firms, as is clear from comparing the two panels of Figure 6. This shows that a reduction in k_v attracts fewer people from self-employment into job search when entry is costly (k_f is high), as in the right panel. 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 rate only falls by 1.5 percentage points as k_v is reduced by half. This pattern holds not only when starting from the average country calibration (shown here), but also when starting at the calibration for the US (see Figure 10).

Table 11 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 workers declines partly because more entrants decide to become employers, but even more because the outside option of search becomes more valuable, as shown by an increase of 8% in the lowest level of productivity at which own-account work is optimal, z_s.

The reduction in vacancy posting costs leads to an increase in aggregate output. This effect is shown in Table 12. 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 almost 9 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 12 differ mainly in 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 unem-

Table 11: The effect of reducing 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 8 and 22. In the second column, k_f takes on the value from the calibration for the US, as also seen in Figure 6b.

ployment. This is natural, given the small changes in self-employment in these economies, shown in Table 11. But in economies with low entry costs and high self-employment, changes in the self-employment rate can account for 30% (average economy) up to 70% (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.

While data limitations prevent a full analysis of the explanatory power of differences in labor market frictions in terms of cross-country output differences, these results suggest that differences in frictions have a sizable effect. As shown, reducing k_v in the model economy for Ethiopia by half results in an output increase by almost 9%. Such an increase in output would reduce the ratio of US to Ethiopian GDP per capita by 8%.

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

Table 12: 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	8.8	1.4
aggregate output net of k_v	7.7	10.0	3.9	6.1
aggregate output net of k_v and k_f	10.6	10.8	16.7	7.0
output of employer firms/employee	-1.5	-0.5	-3.8	-0.7
counterfactual output:				
group sizes as in benchmark	0.0	0.1	-0.9	0.0
only u changes	2.9	5.1	2.6	1.7
only self-employment rates change	1.2	0.0	6.5	-0.4
all group sizes change (average group output as in benchmark)	4.1	5.1	9.2	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 8 and 22. In the second column, k_f takes on the value from the calibration for the US, as also seen in Figure 6b. 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.

is particularly large in economies with strong labor market frictions and low entry costs.

8 Conclusion

The distribution of employment states varies strongly with income per capita. Labor markets in poor countries are characterized not only by lower levels of wage employment and higher levels of self-employment, but also by more unemployment relative to wage employment (a high *UN* ratio). 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 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 low wage employment and 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 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 and screening hard, leading both to a high destruction rate and a high cost of creating a lasting match. This appears to be consistent with the evidence documented by Blattman and Dercon (2018) for Ethiopia. Further analysis for more different settings could prove valuable.

Second, there is a variety of experimental work that has identified the presence of labor market frictions in specific settings. Extensions of the theory that allow relating it directly to this line of work also appear promising.

References

Abebe, G., Caria, S., Fafchamps, M., Falco, P., Franklin, S. and Quinn, S. (forthcoming), 'Anonymity or Distance? Job Search and Labour Market Exclusion in a Growing African City', *Review of Economic Studies* .

Abebe, G., Caria, S. and Ortiz-Ospina, E. (2021), 'The Selection of Talent: Experimental and Structural Evidence from Ethiopia', *American Economic Review* **111**, 1757–1806.

Adamopoulos, T. and Restuccia, D. (2014), 'The size distribution of farms and international productivity differences', *American Economic Review* **104**(6), 1667–97.

Akcigit, U., Alp, H. and Peters, M. (2021), 'Lack of Selection and Limits to Delegation: Firm Dynamics in Developing Countries ', *American Economic Review* **111**(1), 231–275.

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.

Alvarez-Cuadrado, F., Amodio, F. and Poschke, M. (2019), 'Selection and absolute advantage in farming and entrepreneurship', *CEPR Discussion Paper* .

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

Banerjee, A. and Chiplunkar, G. (2018), 'How important are matching frictions in the labour market? experimental & non-experimental evidence from a large indian firm', *Working Paper, Yale University* .

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', *Working Paper, University of Southern California* .

Beam, E. A. (2016), 'Do job fairs matter? experimental evidence on the impact of job-fair attendance', *Journal of Development Economics* **120**, 32–40.

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. (2018), 'The impacts of industrial and entrepreneurial work on income and health: Experimental evidence from ethiopia', *American Economic Journal: Applied Economics* **10**, 1–38.

Bollard, A., Klenow, P. and Li, H. (2016), 'Entry costs rise with development', *mimeo, Stanford University* .

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.

Bridgman, B., Duernecker, G. and Herrendorf, B. (2018), 'Structural transformation, marketization, and household production around the world', *Journal of Development Economics* **133**, 102–126.

Bruegemann, B., Gautier, P. and Menzio, G. (2019), 'Intra Firm Bargaining and Shapley Values', *Review of Economic Studies* **86**(2), 564–592.

Buera, F. and Fattal-Jaef, R. (2016), 'The dynamics of development: Entrepreneurship, innovation, and reallocation', *mimeo, Washington University in St. Louis* .

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

Buera, F. J., Kaboski, J. P. and Shin, Y. (2015), 'Entrepreneurship and Financial Frictions: A Macrodevelopment Perspective', *Annual Review of Economics* **7**.

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. (2008), 'A Theory of Wages and Labor Demand with Intrafirm Bargaining and Matching Frictions', *International Economic Review* **48**(3), 943–972.

Carranza, E., Garlick, R., Orkin, K. and Rankin, N. (2019), 'Job search and hiring with two-sided limited information about workseekers' skills'.

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

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. (2019), 'Labor Market Flows and Development', *mimeo, Federal Reserve Bank of Minneapolis* .

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. (2018), '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. (2018), 'Location, search costs and youth unemployment: Experimental evidence from transport subsidies', *The Economic Journal* **614**, 2353–2379.

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

Gindling, T. H. and Newhouse, D. (2012), 'Self-employment in the developing world', *World Bank Policy Research Working Paper* (6201).

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. (2020), 'Managerial Delegation, Law Enforcement, and Aggregate Productivity', *Review of Economic Studies* **87**, 2256–2289.

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. (2014), 'The Life Cycle of Plants in India and Mexico', *Quarterly Journal of Economics* **129**(3), 1035–1084.

Jovanovic, B. (1984), 'Matching, turnover, and unemployment', *Journal of Political Economy* **92**(1), 108–122.

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. (2020), 'Self Employment in Developing Countries: a Search-Equilibrium Approach', *Review of Economic Dynamics* **35**, 1–34.

Poschke, M. (2018), 'The firm size distribution across countries and skill-biased change in entrepreneurial technology', *American Economic Journal: Macroeconomics* **10**, 1–41.

Pries, M. J. and Rogerson, R. (2022), 'Declining worker turnover: the role of short duration employment spells', *American Economic Journal: Macroeconomics* **14**, 260–300.

Pries, M. and Rogerson, R. (2005), 'Hiring policies, labor market institutions, and labor market flows', *Journal of Political Economy* **113**(4), 811–839.

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

Rissman, E. R. (2003), 'Self-Employment as an Alternative to Unemployment', *Federal Reserve Bank of Chicago Working Paper* **2003-34**.

Rogerson, R. and Shimer, R. (2011), Search in macroeconomic models of the labor market, *in* 'Handbook of Labor Economics', Vol. 4, Elsevier, Amsterdam, pp. 619–700.

Rud, J. P. and Trapeznikova, I. (2021), 'Job creation and wages in least developed countries: Evidence from sub-saharan africa', *The Economic Journal* **131**, 1331–1364.

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.

United Nations (2008), *System of National Accounts 2008*, United Nations, New York, NY.

United Nations (2010), *We Can End Poverty 2015: Millennium Development Goals*, United Nations, New York, NY.

World Bank (2012), *World Development Report 2013: Jobs*, World Bank, Washington, DC.

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

Appendix – For Online Publication

A Additional Tables and Figures

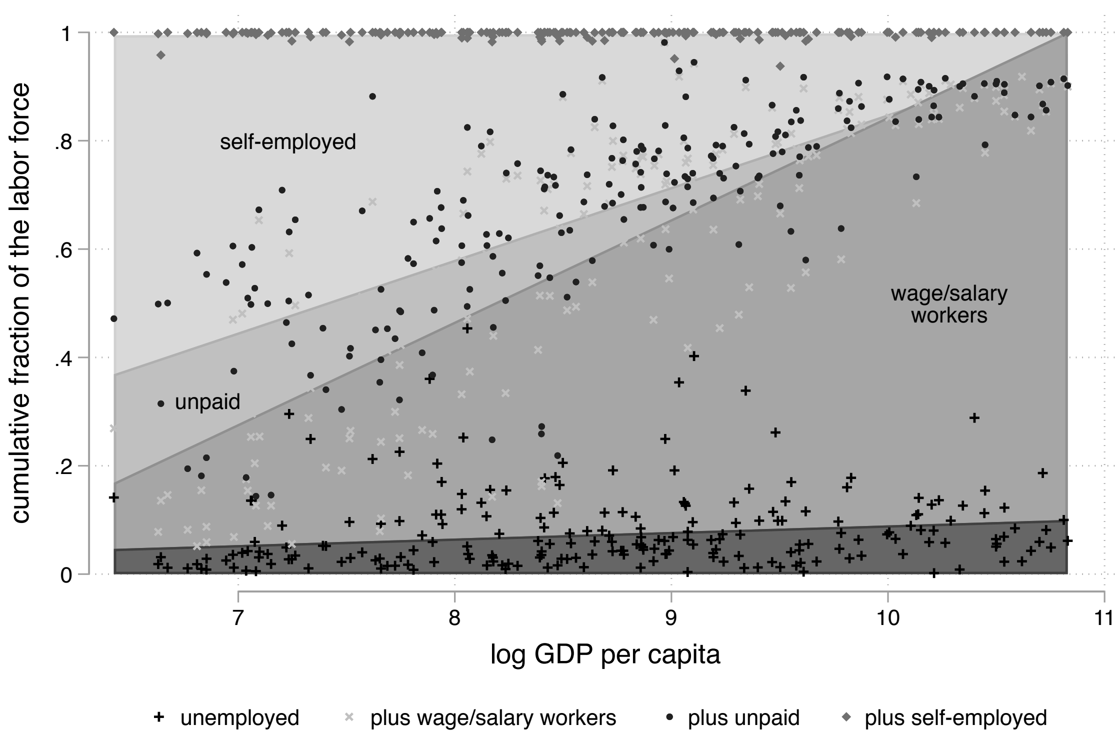


Figure 7: Composition of the labor force and development, national, incl. unpaid workers

Sources: See Figure 1.

Table 13: 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. Data sources as in Figure 1.

Table 14: 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. Data sources as in Figure 1.

Table 15: Composition of the labor force and development, ILO data

dependent variable:	self-employment rate	fraction own-account workers	fraction employers	unemployment rate	UN ratio
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.
Data from the International Labour Organization (ILOSTAT).

Table 16: Unemployment and development, participation rate and alternative measure of unemployment

dependent variable:	non-participation rate	fraction not working	narrow unemployment rate	UN ratio using narrow u rate
Urban areas:				
log GDP per capita	-0.028	-0.027	-0.008	-0.044
	(0.012)	(0.013)	(0.011)	(0.014)
R^2	0.091	0.075	0.009	0.149
observations	150	150	150	150
countries	58	58	58	58
Entire country:				
log GDP per capita	-0.033	-0.027	0.002	-0.043
	(0.011)	(0.012)	(0.009)	(0.011)
R^2	0.120	0.070	0.001	0.180
observations	214	214	214	214
countries	68	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). (Results for a pooled regression are similar.) Constant not reported. Standard errors in parentheses. Data sources as in Figure 1.

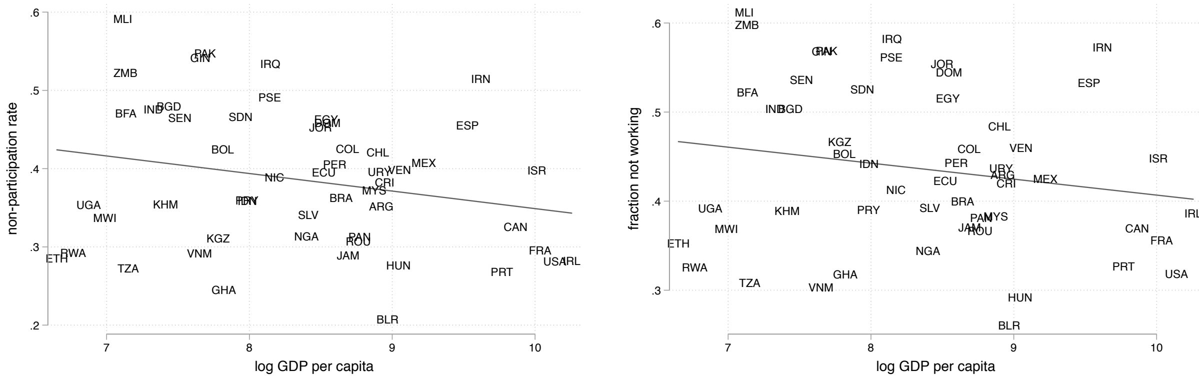


Figure 8: Non-participation rate and fraction of the population not working versus log GDP per capita

Notes: Data for urban areas. For each country, the time average is shown. Regression outputs underlying the lines of best fit reported in Table 16.

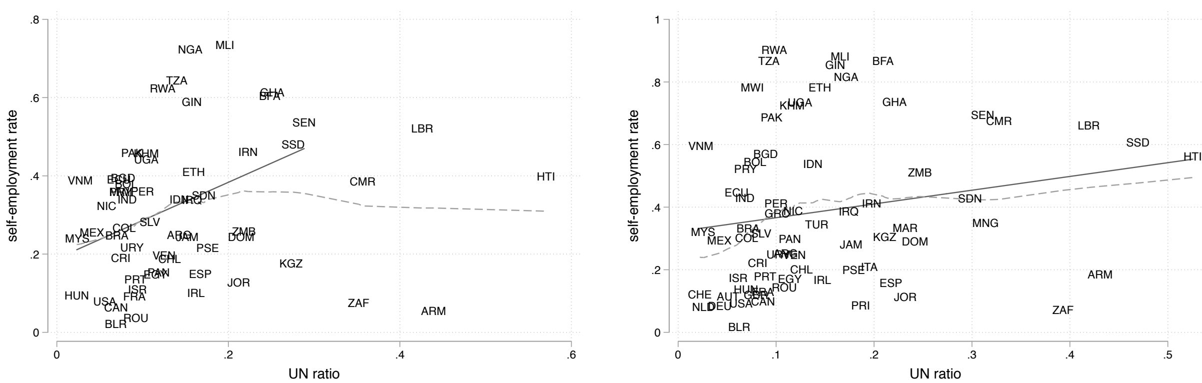


Figure 9: The self-employment rate versus the UN ratio $u/(u+n)$, urban (left) and overall (right), full range of the UN ratio

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

Table 17: 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. Data sources as in Figure 1.

Table 18: 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. Data sources as in Figure 1.

Table 19: The relationship between self-employment and the *UN* ratio, controlling for GDP per capita, entire country

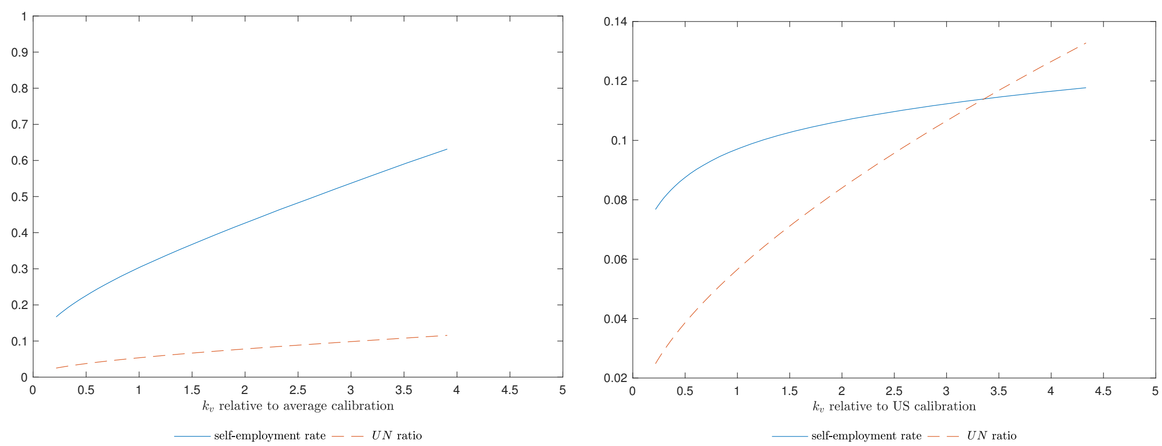
dependent variable:	self-employment rate	fraction own-account workers	fraction employers
Between regression:			
UN ratio	-0.067	-0.170	0.033
	(0.269)	(0.314)	(0.037)
log GDP per capita	-0.195	-0.198	0.010
	(0.017)	(0.020)	(0.002)
R^2	0.684	0.633	0.242
observations	197	172	172
countries	64	59	59
Pooled 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. Data sources as in Figure 1.

Table 20: 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. Results are virtually identical when years before 1995 are included.



(a) Low k_f (b) High k_f

Figure 10: 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 8 and 22 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.

Table 21: Calibration: model and data moments (8 countries and average, data values in parentheses)

country:	avg	USA	CAN	DEU	FRA	ITA	MEX	IDN	ETH
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 employment	5.175	10.298	8.617	10.091	12.627	4.730	3.731	2.577	2.235
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 22: 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 21 since the latter also includes exits due to the owner's retirement.

B Details on the accounting model

B.1 Self-employment entry from unemployment only

The flows among the three states u, n and e are given in Table 5. They result in the steady state stocks for u, n and e given in (1). Comparative statics of these stocks with respect to the flow rates are:

$$\frac{\partial u}{\partial h} = -\frac{u^2}{s}\left(\frac{s}{\lambda} - f\right) < 0 \quad \text{if } s/f > \lambda$$
$$\frac{\partial u}{\partial f} = -u^2\frac{1-h}{s} < 0$$
$$\frac{\partial u}{\partial s} = u^2\frac{(1-h)f}{s^2}$$
$$\frac{\partial u}{\partial \lambda} = u^2\frac{h}{\lambda^2}$$
$$\frac{\partial e}{\partial h} = \frac{\lambda e^2}{h^2}\frac{s+f}{s} = \frac{u^2}{\lambda}\frac{s+f}{s}$$
$$\frac{\partial e}{\partial f} = -\lambda e^2\frac{(1-h)}{hs} = -u^2\frac{h}{s\lambda}(1-h)$$
$$\frac{\partial e}{\partial s} = \frac{1-h}{h}\frac{\lambda f e^2}{s^2} = u^2(1-h)\frac{h}{\lambda}\frac{f}{s^2}$$
$$\frac{\partial e}{\partial \lambda} = -e^2\frac{s+(1-h)f}{hs} = -u^2\frac{h}{\lambda^2}\frac{s+(1-h)f}{s}$$
$$\frac{\partial \tilde{u}}{\partial h} = \tilde{u}^2\frac{f}{s} = \frac{u^2}{(1-e)^2}\frac{f}{s}$$
$$\frac{\partial \tilde{u}}{\partial f} = -\tilde{u}^2\frac{1-h}{s} = -\frac{u^2}{(1-e)^2}\frac{1-h}{s}$$
$$\frac{\partial \tilde{u}}{\partial s} = \tilde{u}^2\frac{(1-h)f}{s^2}$$
$$\frac{\partial \tilde{u}}{\partial \lambda} = 0.$$

A key observation is that in the data, e and $\tilde{u}$ are positively correlated. This could be generated by variation in h, f or s, with

$$\begin{aligned}
\left.\frac{de}{d\tilde{u}}\right|_{\text{vary only } h} &= \frac{(1-e)^2}{\lambda}\frac{s+f}{f} \\
\left.\frac{de}{d\tilde{u}}\right|_{\text{vary only } f} &= (1-e)^2\frac{h}{\lambda} \\
\left.\frac{de}{d\tilde{u}}\right|_{\text{vary only } s} &= (1-e)^2\frac{h}{\lambda}.
\end{aligned}$$

As discussed in the main text, the first of these expressions takes on very large values in practice, whereas the other two have plausible values in the range from 1 to 1.5.

Variation in h, f or s also implies

$$\begin{aligned}
\left.\frac{de}{du}\right|_{\text{vary only } h} &= \frac{s+f}{\lambda f - s} \\
\left.\frac{de}{du}\right|_{\text{vary only } f} &= \frac{h}{\lambda} \\
\left.\frac{de}{du}\right|_{\text{vary only } s} &= \frac{h}{\lambda}.
\end{aligned}$$

As discussed in the main text, the first of these expressions takes on values around minus 7-20 in practice, whereas the other two have plausible values in the range from 1.5 to 2.

B.2 Self-employment entry from both employment and unemployment

Now consider the case where the employed also enter self-employment, at a per period rate of $\hat{g} \equiv gh$. This implies that the stocks are

$$\begin{aligned}
u &= \frac{\lambda(gh+s)}{(1-h)f(gh+\lambda)+gh(h+\lambda)+s(h+\lambda)} \\
e &= \frac{(1-h)fgh+h(gh+s)}{(1-h)f(gh+\lambda)+gh(h+\lambda)+s(h+\lambda)} \\
n &= \frac{(1-h)f\lambda}{(1-h)f(gh+\lambda)+gh(h+\lambda)+s(h+\lambda)} \\
\tilde{u} &\equiv \frac{u}{u+n} = \frac{gh+s}{(1-h)f+gh+s}.
\end{aligned}$$

(gh now is an employment outflow and shows up accordingly.)

Here, the derivatives look more complicated, but one can show that comparative statics imply

$$
\begin{array}{llll}
 & \dfrac{\partial e}{\partial h} > 0 \qquad sign\left(\dfrac{\partial e}{\partial f}\right) = sign(g-h) & \dfrac{\partial e}{\partial s} > 0 & \dfrac{\partial e}{\partial \lambda} < 0 \\
sign\left(\dfrac{\partial u}{\partial h}\right) = sign(\lambda - s/f) & \dfrac{\partial u}{\partial f} < 0 & \dfrac{\partial u}{\partial s} > 0 & \dfrac{\partial u}{\partial \lambda} > 0 \\
\dfrac{\partial \tilde{u}}{\partial h} > 0 & \dfrac{\partial \tilde{u}}{\partial f} < 0 & \dfrac{\partial \tilde{u}}{\partial s} > 0 & \dfrac{\partial \tilde{u}}{\partial \lambda} = 0.
\end{array}
$$

C Proofs and derivations

C.1 Summary of model timing

The following summarizes the timing of events within a period 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 $(\phi,\ \lambda_f,\ \lambda_s,\ \xi,\ \delta,\ \theta \cdot q(\theta))$ are realized.
4. Value functions are measured and occupational choices take place.
5. Production takes place and payoffs $(w,\ b)$ are realized.

C.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 $W-U$:

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

Firm's Surplus From equation (12),

$$F_f(n,z)=\frac{1+r}{(1+r)-(1-\phi)(1-\lambda_f)}\left(zn(z)^{\gamma}-n(z)w-\frac{k_v}{q(\theta)}n(z)(\xi+(1-\xi)\phi)\right) \tag{24}$$
$$+\frac{(1-\phi)\lambda_f}{(1+r)-(1-\phi)(1-\lambda_f)}U.$$

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

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

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

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

Let

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

where $s\equiv1-(1-\phi)^2(1-\lambda_f)(1-\xi)$.

Nash Bargaining The bargaining rule implies that the wage solves

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

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

$$w=(1-\eta)\frac{r+\phi}{1+r}U+\eta\left(y'\left(n\right)-n\cdot w'\left(n\right)\right). \tag{25}$$

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

$$w = \frac{r+\phi}{1+r}U + \frac{\eta}{1-\eta}\left[1 - \frac{(1-\phi)(1-\lambda_f)}{1+r} + \xi + (1-\xi)\phi\right] \cdot \frac{k_v}{q(\theta)}. \tag{26}$$

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

$$n(z) = (z\gamma)^{\frac{1}{1-\gamma}} \left\{(\eta(\gamma-1)+1)\left[\left(1 - \frac{(1-\phi)(1-\lambda_f)}{1+r} + \xi + (1-\xi)\phi\right)\frac{k_v}{q} + w\right]\right\}^{\frac{-1}{1-\gamma}}. \tag{27}$$

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

$$w'(n) + \frac{w}{\eta n} - \frac{y'(n)}{n} = 0. \tag{28}$$

The solution of the homogeneous equation

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

then is

$$w(n) = Cn^{-1/\eta}. \tag{29}$$

C is a function of integration that can be a function of n. So take the derivative of equation (29) with respect to n:

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

Substituting this into (28) 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 (25) 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

$$w(n) = (1-\eta) \frac{r+\phi}{1+r} U + \frac{y'(n)}{\gamma - 1 + 1/\eta}. \tag{30}$$

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 (24), 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 (30) 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 (27). Substituting this expression into (30) yields the wage at the optimal employment level given in equation (26).

D 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.

D.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 AGE, and use the provided weights.

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 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 used in some tables, 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.

D.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%.

D.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.

D.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)