

Deconstructing the Missing Middle:

Informality and Growth of Firms in Sub-Saharan Africa^{*}

Kaleb Abreha[†] Xavier Cirera[†] Elwyn Davies[†] Roberto N. Fattal-Jaef[‡]
Hibert Maemir[†]

October 3, 2022

Abstract This paper uses a novel firm-level dataset and finds a "missing middle" in the employment-based firm size distribution in four sub-Saharan African countries. This "missing middle" hinges on the inclusion of informal producers, and it is not explained by state or foreign owned firms at the top of the size distribution, nor does it emerge from the size distribution of entrants. We reconcile these empirical results with a model of firm dynamics with endogenous informality, and show that calibrated values of entry barriers and productivity-dependent idiosyncratic distortions generate a "missing middle" that is consistent with its underlying drivers in the data.

JEL Classification: L11, O11, O14, O17, O55

Keywords: Missing middle, size distribution of firms, informality, market distortion, establishment censuses, manufacturing, sub-Saharan Africa

^{*} The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank and its affiliated organizations or those of the Executive Directors of the World Bank or the governments they represent. All errors are ours. We would like to thank the infoDev Trust Fund for its financial support.

[†] The World Bank, 1818 H Street, NW Washington, DC 20433, USA. Corresponding email addresses: kabreha1@worldbank.org (K. Abreha); xcirera@worldbank.org (X. Cirera); edavies@worldbank.org (E. Davies); rfattaljaef@worldbank.org (R. Fattal-Jaef), and hmaemir@worldbank.org (H. Maemir).

1 Introduction

In the quest to identify the drivers of the wide gaps in total factor productivity across countries, a large and growing literature has emerged focusing on cross-country differences in the size distribution of manufacturing firms. A rooted conjecture is that the size distribution of firms in developing countries exhibits a "missing middle", reflecting market frictions and policy distortions that discourage production at an intermediate scale; thereby reducing aggregate productivity. However, the literature evaluating this hypothesis has not yet been conclusive. Tybout (2000) and Tybout (2014) document a lower employment share accounted for by middle-sized firms in low-income countries, while Hsieh and Olken (2014) argue against the missing middle providing an adequate characterization of the firm-level data in these countries.¹

In this paper, we revisit the missing middle hypothesis through the lens of a set of comprehensive firm-level databases from four Sub-Saharan African economies, a region where some of the underlying frictions and distortions are arguably most likely at play (see for example Cirera, Fattal-Jaef, and Maemir (2019)). One challenge when characterizing the "true" distribution of firms in developing countries is the large prevalence of informality. To this end, we compile detailed and comprehensive censuses of establishments including informal producers in four sub-Saharan African countries, namely Burkina Faso, Cameroon, Ghana, and Rwanda. A virtue of the data is that it provides information on the establishment's registration with various government agencies and its bookkeeping practices. We leverage this information to construct common definitions of informality across databases and to assess the sensibility of the results to alternative definitions. A limitation of this data, on the other hand, is the lack of financial information about the establishment, restricting our measurement of size to the number of employees and forbidding any measurement of productivity at the firm level.² To calibrate distortions in the quantitative model, we use Ghana's census of industrial firms covering the universe of formal establishments.

We observe a missing middle in the employment-based size distribution of firms and proceed to deconstruct it into its potential drivers. Our main finding is that the missing middle emerges only under the inclusion of informal establishments in the economy. Restricting the analysis to formal firms leads to a lower average size than in advanced economies but a shape of the size distribution that resembles qualitatively that in the U.S. A first conclusion to be drawn from this result is that imperfect coverage of the informal sector may lead to missing the missing middle. The second conclusion, to be reinforced below through other elements of the deconstruction, is that it is informative of the mechanisms through which distortions manifest in the economy. In particular, it points to distortions that reduce average formal-sector size while concentrating employment in small-sized businesses through the proliferation of an informal sector rather than through a reallocation from the middle to the bottom of the formal sector's size distribution.

A second step in the deconstruction of the missing middle is to assess whether it is observable in the size distribution of entrants. A plausible conjecture regarding firm demographics in developing

¹ We provide further details on the specific findings from these authors in section 2

² The data also prevents us also from decomposing formal and informal employment within a formal firm, an intensive margin that was highlighted as important in (Ulyssea, 2020) and (Ulyssea, 2018). Notice, however, that this distinction does not affect the observation of a missing middle, but may interfere with the inference of it not being present in the formal employment's size distribution

countries is that it is characterized by little growth and turnover of firms, a feature that would motivate the missing middle to emerge at entry. Our evidence refutes this conjecture, as it shows a size distribution of entrants where the employment share of the largest firms is significantly lower than in the whole economy's distribution. Rather, our data points to a growth of the right tail and a depletion of the middle as an outcome of growth and churning in the formal sector over the course of the life cycle. Inspecting the size distribution of firms across various cohorts, we find that the formal firm size distribution reallocates employment from the bottom to the top of the distribution as the cohort ages, whereas the informal firm size distribution remains rather stable. The missing middle, then, is an outcome of stagnant informal businesses and growing formal producers conditional on survival.³

The last two steps in the deconstruction involve assessing the role of government and foreign-owned enterprises in accounting for the employment share at the top. Also motivated by the conjecture of little growth and turnover of firms in less developed countries, a plausible explanation for the employment concentration among large firms is that it is accounted for by the government and foreign-owned firms. Able to circumvent domestic distortions hampering firm growth, government and foreign ownership stand as prominent large employers in the economy. Our data, however, attributes a weak role to this channel. Even conditioning on domestic and private firms, a missing middle emerges in the data.

The deconstruction of the missing middle leads us to study the role of entry barriers and productivity-dependent idiosyncratic distortions in explaining the missing middle and its driving forces. To this end, we extend a standard model of firm dynamics and feed it with calibrated values of entry barriers and idiosyncratic distortions from the Ghanaian firm-level data. Our calibration strategy of distortions follows the approach in [Fattal-Jaef \(2022\)](#). The productivity-elasticity of idiosyncratic distortions is calibrated as the regression coefficient from projecting between $\log TFP_{R?}$ and $\log TFP_Q$, defined as in [Hsieh and Klenow \(2009\)](#), and the entry barrier is inferred residually so that the distorted economy's average formal firm size replicates the Ghanaian economy.

Equipped with estimates of the distortions, we propose a model of firm dynamics with endogenous informality to characterize their implications for the firm-size distribution and aggregate outcomes. The model treats formal and informal goods as imperfectly substitutable. Consumers confront a nested constant elasticity of substitution demand system combining baskets of formal and informal varieties. There is endogenous entry to the formal and informal sectors. Upon entry, formal producers are endowed with an endogenous innovation technology that shapes firm growth over the life cycle. The informal sector's productivity evolves exogenously. Both supply and demand side forces interact in shaping the economy's response to the distortions. When these are introduced, the entry to the formal sector declines, formal firm growth flattens, and formal varieties become relatively more expensive. As a result, consumers switch expenditures from formal to informal varieties, motivating a rise in firm entry into the informal sector. We quantify these mechanisms when both distortions interact and when they are introduced in isolation

We find that an empirically plausible missing middle emerges when there are both entry barriers

³ Our inference of the life-cycle properties of the firm-size distribution is limited by the cross-sectional nature of our data, which forbids us from disentangling the role of firm growth from the role of survival in explaining the rise of the right tail of the formal firm's size distribution.

ers and idiosyncratic distortions at play. Entry barriers polarize employment in smaller and larger firms, but the allocation of employment is excessively concentrated at the top. This is because, while fostering informality, entering formal firms confronts less competition and therefore increases their average size. Idiosyncratic distortions can also generate a rise in the informal sector, and thus reallocate employment towards small informal producers. However, the productivity-dependent nature of idiosyncratic distortions reallocates formal employment from large to middle-size firms, a force that counteracts the emergence of the missing middle. When they interact, the limitations of each distortion are attenuated and a missing middle comparable to the ones in our data emerges. Specifically, we find that these distortions can account for three-quarters of the observed informal employment in Ghana and reduce aggregate TFP by 12%.

The rest of the paper is organized as follows. Section 2 briefly reviews the debate about the existence of the missing middle. Section 3 describes the data used in the analysis, and section 4 discusses the findings. Section 5 describes the model and section 6 the calibration and quantification exercise. The final section concludes.

2 The Missing Middle Debate and other Related Literature

Our work lies at the intersection of the literature debating the existence of a missing middle in the size distribution of firms in developing countries and the body of work documenting properties of informal production and quantifying their aggregate implications. We begin discussing our placement in the missing middle literature and then turn to provide our contribution to the broader literature on distortions, informality, and cross-country differences in productivity.

2.1 The Missing Middle Debate

Our motivation for assessing if a country's firm size distribution features a missing middle lies in that it conveys information about the nature of the underlying distortions in the economy. This diagnostic value of the missing middle becomes particularly useful when a lack of detailed firm-level data precludes any firm-level characterization of the economy besides the measurement of firm size. As is the case in the Sub-Saharan African countries we cover, where we only count with comprehensive information on the employment composition of the firms but have no information about their balance sheet, any direct inference of distortions based on this information is precluded. However, since the size distribution is part of an economy's equilibrium, it is shaped by underlying drivers of resource allocation. While not providing a unique mapping to the exact combination of distortions at play, we argue that by combining a theory of the firm size distribution with candidate distortions we can assess the ability of these distortions to generate a firm size distribution as in the data. In this paper, we propose entry barriers and idiosyncratic distortions as plausible distortions.

Besides the conceptual motivation, our work is also inspired by the conflicting results in the missing middle literature. [Tybout \(2000\)](#) presents the distribution of employment shares by firm size for 19 countries and documents a spike in the 1-4 size category which drops off quickly in the middle size category of 10-49 before rising again for the largest firms. [Tybout \(2014\)](#) proposes a parametric assessment of the existence of a missing middle comparing the share of employment in

medium-sized firms in the data against the implied share from a Pareto distribution fitted to the U.S. firm size distribution. The author shows that a missing middle emerges also under this definition.

In an influential reassessment of the missing middle hypothesis, [Hsieh and Olken \(2014\)](#) argue that the theories of the missing middle are about firms rather than employment, and hence postulate the unweighted firm size distribution as the relevant metric. They show that the firm size distribution is unimodal and that a missing middle emerges only under the restrictive assumptions of an employment-weighted distribution with size classes grouped into wide bins.

As stated earlier, our view is that documenting the pervasiveness of a missing middle in the employment-based size distribution is of informational value to uncover distortions hampering aggregate productivity. Moreover, besides being more likely to portray a missing middle, [Hopenhayn \(2014\)](#) shows that focusing on employment-weighted distributions is also the conceptually sound approach for diagnosing distortions and assessing their aggregate effects. Against this backdrop, our contribution to the missing middle debate is to document the pervasiveness of the missing middle in the employment-based size distribution of four Sub-Saharan African economies and highlight its dependence on the consideration of informal producers.

2.2 Other Related Literature

The second part of our paper, where we propose a quantitative model of firm dynamics with an informal sector, touches upon a large literature on distortions, informality, and aggregate productivity.

Building on the pioneering model of the informal sector proposed in [Rauch \(1991\)](#), a large quantitative literature emerged studying the role of varying types of distortions in explaining the prevalence of informality and in accounting for the aggregate productivity differences across countries. Salient examples in this area are [Antunes and Cavalcanti \(2007\)](#), [D’Erasmus and Moscoso Boedo \(2012\)](#), [Ordoñez \(2014\)](#), [Lopez-Martin \(2019\)](#), [Alvarez and Ruane \(2019\)](#), and [Ulyssea \(2018\)](#) among others.

We provide three points of departure from this literature. First, we model informal and formal consumption goods baskets as differentiated products through a CES demand system. This channel brings the reallocation of expenditure across types of goods as a new and realistic margin of adjustment, complementing the supply side forces covered in the existing work.⁴ Second, our quantitative analysis is disciplined by seeking to establish a tight connection with the missing middle in the firm size distribution. While the existing work appeals to statistics of the firm size distribution of formal and informal firms to calibrate parameter values, the focus is on aggregate effects, rather than seeking to account for the shape of the distribution. Last, we feed broad measures of entry and allocative wedges in the economy identified jointly from the firm-level data and the model’s equilibrium ([Fattal-Jaef, 2022](#)). While less informative than studying how a specific distortion within a class of distortion shapes informality and aggregate productivity, as is the case in most of the literature, our approach subjects the entire class of distortion to a higher hurdle, namely its ability to

⁴ Empirical studies documenting the importance of differentiating formal from informal expenditure are [Bachas, Gadenne, and Jensen \(2020\)](#) and [Atkin, Faber, and Gonzalez-Navarro \(2018\)](#)

generate a missing middle in the size distribution.⁵⁶

3 Data

Firm-Level Data in Sub-Saharan Africa

The lack of suitable data constitutes one of the main challenges associated with testing the missing middle conjecture. Even though the information required for this type of exercise is extremely basic—the power of looking at the size distribution to assess broader macroeconomic patterns precisely lies in its simplicity and low data requirement—the availability of such information proves to be an extremely binding constraint in developing countries. Such data availability is particularly challenging for most sub-Saharan African countries. Additionally, in view of the pervasiveness of informal economic activities in these countries, it is necessary that revisit the question of a missing middle should use census data of both formal and informal firms across the entire spectrum of sizes.

Fortunately, over recent years, a number of sub-Saharan countries have engaged in establishment censuses as part of wider programs aimed at informing economic policies. Crucially for our purpose, the census is by definition taken of all establishments, including unregistered ones, so the ensuing data is representative of the true size distribution. These censuses are typically nothing more than business registers and essentially include contact and location information, legal form, records of business registration, tax and other administrative data, type of economic activity, and the number of workers, which is the crucial element for our purpose.

For this paper, we have managed to go directly to the data sources and obtain firsthand establishment censuses from the statistical agencies of select sub-Saharan countries, namely Burkina Faso, Cameroon, Ghana, and Rwanda. The selection of these countries is primarily because their census datasets feature a number of characteristics that arguably ascertain their representativeness of the true size distribution. On the whole, these data are consistent with [Lewis \(1954\)](#) widely held view of developing economies as characterized by the coexistence of a few capital-intensive industries and a mass of inefficient businesses (“tiny islands of capitalist employment surrounded by a vast sea of subsistence workers”). Specifically, small establishments (fewer than ten workers) represent the overwhelming majority of the manufacturing sector in our datasets; for comparison, small businesses account for less than half the manufacturing establishments in the U.S. Moreover, our datasets do not present any obvious censoring or threshold for inclusion in the census. To the best of our knowledge, our contribution is the first that encompasses the entire manufacturing sectors—including informal establishments—of sub-Saharan countries in a comparative fashion.⁷

⁵ To fix ideas, idiosyncratic distortions are a class of distortions causing allocative inefficiency. Financial frictions are a specific distortion within this class. Similarly, business registration costs are a specific example within the class of entry barriers.

⁶ [Alvarez and Ruane \(2019\)](#) also consider idiosyncratic distortions alongside other specific frictions in the context of Mexico. In our model, however, these distortions are strongly increasing with the physical productivity of the firms, an elasticity which we estimate from the firm-level Ghanaian data. Moreover, the productivity elasticity of distortions interacts with the endogenous innovation decisions of firms, absent in their work. We show that endowing firms with an endogenous life-cycle constitutes an important mechanism through which distortions generate a missing middle in the distribution.

⁷ Our dataset is more comprehensive by enumerating all establishments. For example, [Hsieh and Olken \(2014\)](#) have census establishments with 20 or more employees, but they have a sample of those with 20 or fewer workers in the case

Summary of firm-level data from select countries in sub-Saharan Africa⁸

Burkina Faso. The data comes from the census conducted in 2016 by Institut National de la Sstatistique et de la Démographie (INSD). It covers establishments mainly in the non-agricultural sector and contains records such as an establishment's industry classification, age, employment, accounting practices, state ownership, and foreign equity. Establishments are classified into formal and informal depending on whether they have a registration number (with mobile trade and mobile credit, tax identification number) and use SYSCOA as an accounting standard.⁹

Cameroon. The data is obtained from the first Recensement Général des Entreprises (RGE) conducted in 2009 by the Institut National de la Statistique (INS). The RGE covers all businesses of the modern sector (that is, excluding non-commercial agriculture) with a physical location, specifically excluding the street trading industry (hawkers, peddlers, and others), the (car and bike) taxi industry, and the callbox industry. It provides detailed information such as registration with administrative agencies, tax payment, accounting practices, age, employment size, state ownership, and foreign equity. In the census, formal establishments are considered formal if they prepare a Statistical and Fiscal Declaration (DSF) or those that do not prepare a DSF but which keep an operating account and a partial account of balance sheets.¹⁰

Ghana. The data comes from the 2014 Integrated Business Establishment Survey (IBES) conducted by the Ghana Statistical Service. The IBES is an economic census of all business establishments (registered and unregistered) in the non-agricultural sectors of the economy. It has detailed information about registration with administrative agencies, tax payment, accounting practices, age, employment size, state ownership, and foreign equity. The data contain records of an establishment's registration status with government agencies, and formality is defined based on registration status with Registrar General's Department (RGD) and keeping formal accounts.¹¹

Rwanda. The source of the data is the 2014 Establishment Census (EC) conducted by the National Institute of Statistics of Rwanda (NISR). The EC constitutes a virtually exhaustive business registry as surveys every establishment that has a fixed location and operates an economic activity. The ensuing microdata is comprehensive and detailed with respect to key establishment characteristics such as location, registration and legal status, age, ownership and management, industrial classification, and employment size. The census covers both formal and informal businesses. An establishment is considered to operate in the formal sector if it is registered at Rwanda Revenue Authority (RRA) and maintains regular operational accounts.¹²

of Indonesian manufacturing. Similarly, for Indian manufacturing, their dataset comprises a complete enumeration of formal establishments with 100 or more workers and a sample of formal establishments with fewer than 100 workers and informal ones. The exception is the Mexican manufacturing data, which comprises all establishments with fixed locations.

⁸ Note that we interchangeably use terms firm, establishment, and plant unless otherwise indicated.

⁹ National Institute of Statistics and Demography (INSD). [Analysis report of the XVII Industrial and Commercial Census](#). November 2018.

¹⁰ National Institute of Statistics of Cameroon (INS). [General Census of Enterprises In 2009](#). September 2010.

¹¹ Ghana Statistical Service. [Integrated Business Establishment Survey: Summary Report](#). September 2015.

¹² National Institute of Statistics of Rwanda (NISR). [The Establishment Census, 2014](#). June 2015.

Table 1 summarizes the key features of the census data from these sub-Saharan African countries. We now turn to the description of each of the datasets. In Burkina Faso, 99,261 establishments hire about 256,000 workers in 2015. The corresponding figures are 88,144 establishments and 430,000 workers in 2008 (Cameroon); 638,234 establishments and 3.4 million workers in 2013 (Ghana), and 154,236 establishments and 493,000 workers in 2013 (Rwanda). Establishments with foreign equity or state ownership are rare in number and employ a small number of workers. Notable exceptions are the employment shares of foreign-owned establishments in Cameroon (17 percent), and state-owned enterprises in Ghana (15 percent) and Rwanda (14 percent).

Manufacturing production involves a small fraction of the establishments and the workforce in the economy ranging from 7 percent (Ghana) to 17 percent (Burkina Faso) and 8 percent (Rwanda) to 21 percent (Burkina Faso), respectively. It also involves a limited number of foreign-owned establishments, but they play a bigger role as sources of manufacturing employment. In Burkina Faso, they constitute about 1 percent of the establishments and 7 percent of the employment in manufacturing. The corresponding figures are 5 and 25 percent in Cameroon, 1 and 7 percent in Ghana, and 1 and 17 percent in Rwanda. Likewise, state-owned establishments are less than 1 percent of manufacturing establishments. However, they contribute a small fraction to manufacturing employment; about 1 percent in Burkina Faso, Cameroon, and Ghana, and 4 percent in Rwanda.

Table 1: Description of the establishment census data

Indicator	Coverage	Burkina Faso (2015)				Cameroon (2008)				Ghana (2013)				Rwanda (2013)						
		Total	Foreign-owned	State-owned	%	Total	Foreign-owned	State-owned	%	Total	Foreign-owned	State-owned	%	Total	Foreign-owned	State-owned	%			
Establishment	All	99,261	1,177	1.19	0.03	88,144	6,385	7.24	185	0.21	638,234	8,827	1.38	25,536	4.00	154,236	10,510	0.98	1,577	1.02
	Manufacturing	16,705	170	1.02	0.02	10,922	519	4.75	1	0.01	99,437	711	0.72	164	0.16	10,742	152	1.42	8	0.07
Employment	All	255,908	20,774	8.12	2.41	429,758	73,464	17.09	9,590	2.23	3,383,206	136,265	4.03	497,078	14.69	493,302	28,920	5.86	67,546	13.69
	Manufacturing	53,116	3,981	7.49	0.75	82,502	20,742	25.14	352	0.43	437,316	31,272	7.15	5,657	1.29	39,708	6,883	17.33	1,667	4.20

Source: Establishment censuses obtained from the statistical agencies of the select countries; see section 3. Note: Establishments with missing employment are excluded.

Table 2: Summary statistics on establishment census data

Country	All sectors										Manufacturing											
	Establishment					Employment size (# of workers)					Establishment					Employment size (# of workers)						
	Formal	Formal	Informal	Mean	p50	Formal	Formal	Informal	Mean	p50	Formal	Formal	Informal	Mean	p50	Formal	Formal	Informal	Mean	p50		
Burkina Faso (2015)	9.56	31.81	2	1	15	3	27.48	2	2	36	6	3.89	27.48	2	2	36	6	3.89	27.48	2	2	36
Cameroon (2008)	7.78	46.11	3	2	29	6	58.65	3	2	84	14	5.26	58.65	3	2	84	14	5.26	58.65	3	2	84
Ghana (2013)	32.62	50.63	4	2	8	2	50.07	3	2	8	3	28.12	50.07	3	2	8	3	28.12	50.07	3	2	8
Rwanda (2013)	23.69	64.18	2	1	9	2	59.11	2	1	10	2	21.47	59.11	2	1	10	2	21.47	59.11	2	1	10

Source: Establishment censuses obtained from the statistical agencies of the select countries; see section 3. Note: Establishments with missing employment are excluded.

4 The size distribution in sub-Saharan Africa

Our exploration has generated a few stylized facts about the patterns and drivers of the size distribution of manufacturing plants in Burkina Faso, Cameroon, Ghana, and Rwanda. Whenever possible, we compare them to the well-known size distribution of firms in the U.S. for perspective.

In our sample of countries, the size of an establishment is approximated by the number of persons engaged in the production of its products or services. It comprises both permanent and temporary workers, but there is no distinction between full- and part-time workers. For our purpose, establishments are categorized into six size classes: 1-4, 5-9, 10-19, 20-99, and 100 or more workers.¹³

The size distribution of manufacturing plants is characterized by a missing middle

Fact 1. *Medium-sized manufacturing plants account for a relatively low employment share*

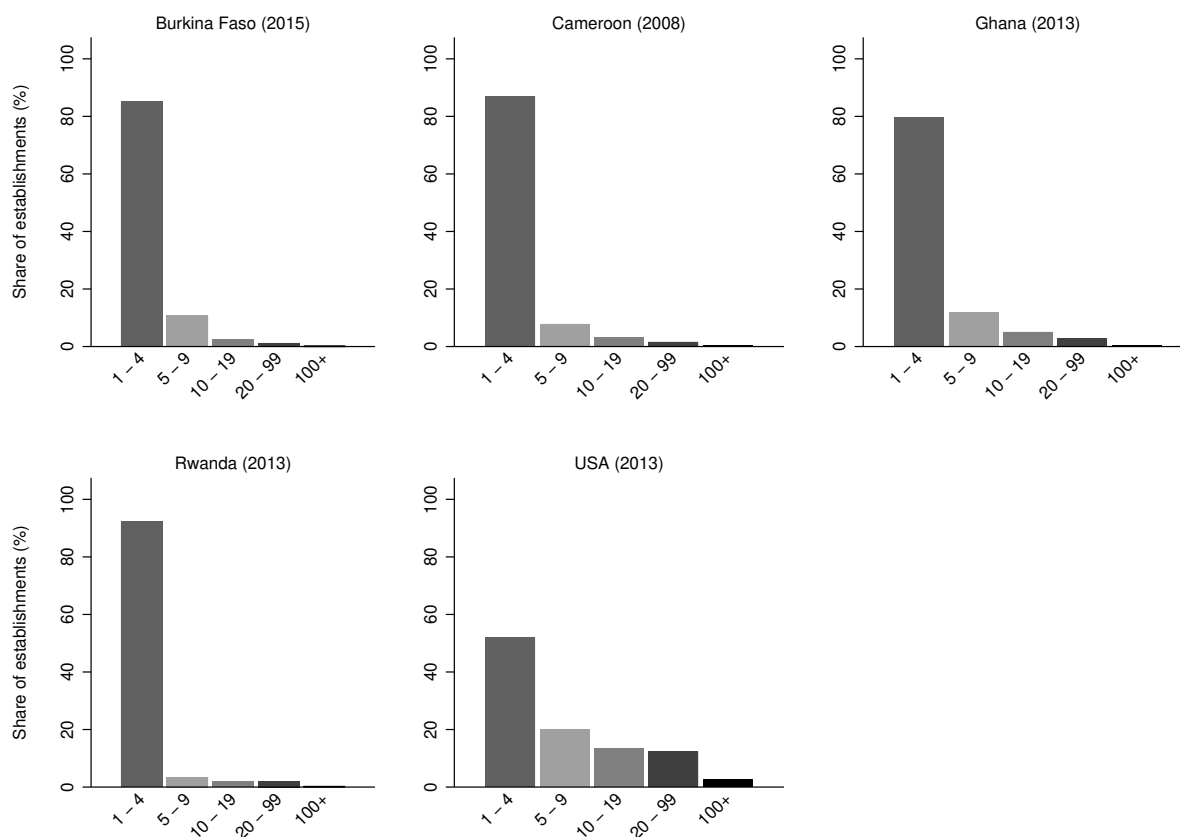
Figure 1 plots the distribution of establishments across the five size classes. It is shown that small establishments dominate the size distribution. Specifically, establishments of size 1-4 workers constitute more than 75 percent of all the establishments in the economy. Although a similar pattern is observed in the U.S., the concentration in the small size categories is not to the same extent as observed in our sample countries. Interestingly, the number of establishments monotonically decreases when moving up to larger size classes, which represents a sharp decline compared with the pattern in the U.S. Figure 2 also shows the preponderance of manufacturing establishments in smaller size classes, featuring the same structural characteristics for manufacturing plants in Mexico, Indonesia, and India (Hsieh & Olken, 2014). However, this sharply contrasts with the observation in the U.S. where the distribution is more even. The monotonic and sharp decline in the density of establishments with larger size classes highlights a rather low number of establishments in the intermediate and large range of the production scale.

By contrast, figure 3 and 4 reveal a missing middle in the employment distribution. Most of the workers are employed in either small or large establishments, and employment in the middle size categories is well below the share of employment in the other size categories. For the U.S., the monotonic increase in the distribution of employment shares across size classes confirms a major difference in the structure of manufacturing between developing and developed countries.¹⁴

¹³ The appendix provides results based on different size cut-offs. The key findings still hold.

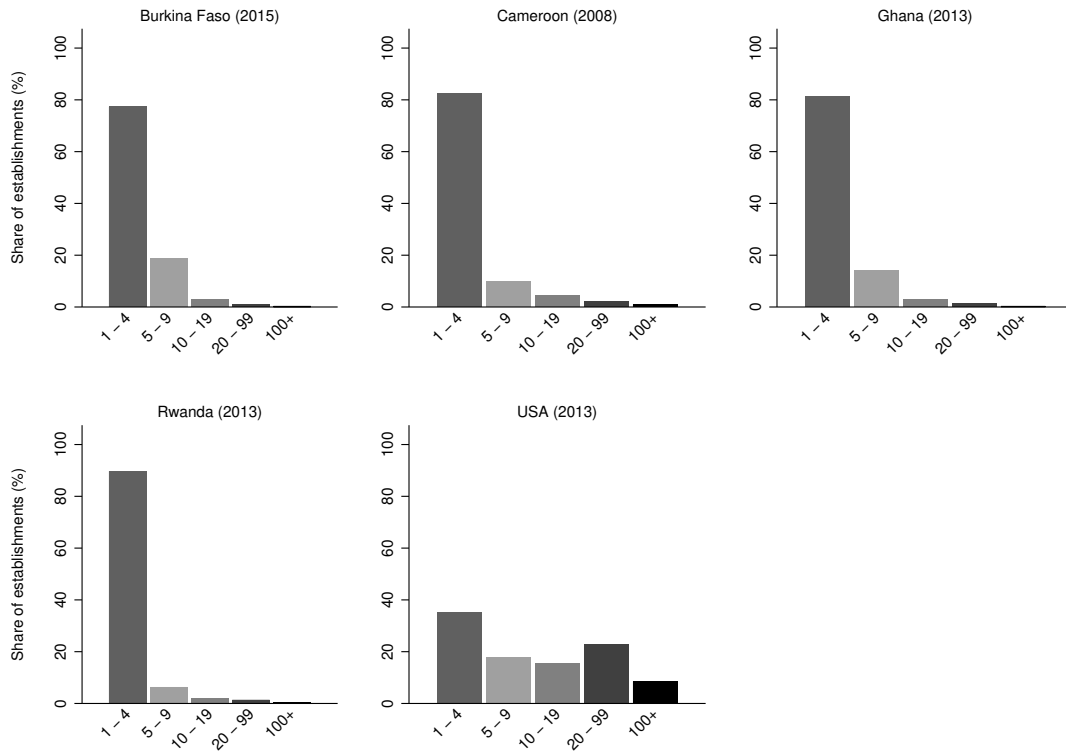
¹⁴ Since the focus is on manufacturing production and the pattern of the establishment size distribution is clearly known, the forthcoming discussion on the size distribution exclusively considers the employment distribution. Henceforth, results on the establishment distribution and other sectors of the economy are relegated to the appendix and are only referred to whenever necessary.

Figure 1: Establishment distribution: Economy-wide



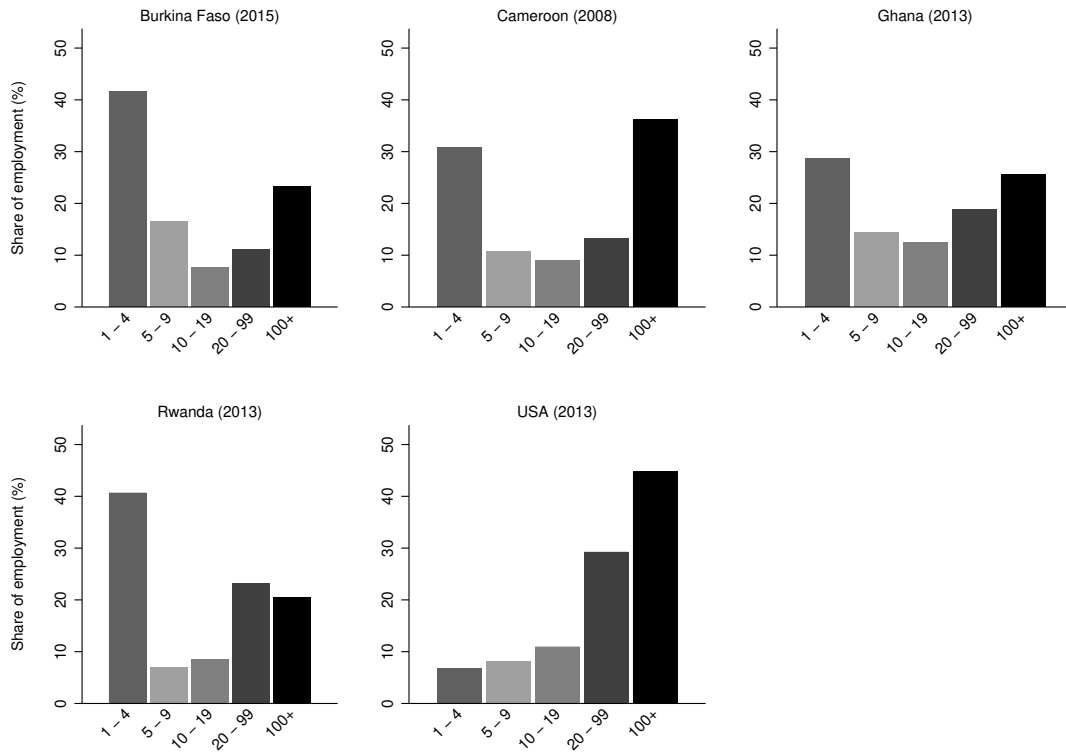
Source: Establishment censuses obtained from the statistical agencies of the select countries; see section 3. For the U.S., the data comes from the 2019 Business Dynamics Statistics (BDS) dataset. *Note:* The figure is constructed based on data for establishments in all sectors irrespective of their state or foreign ownership status. Establishments with missing employment data are excluded.

Figure 2: Establishment distribution: Manufacturing



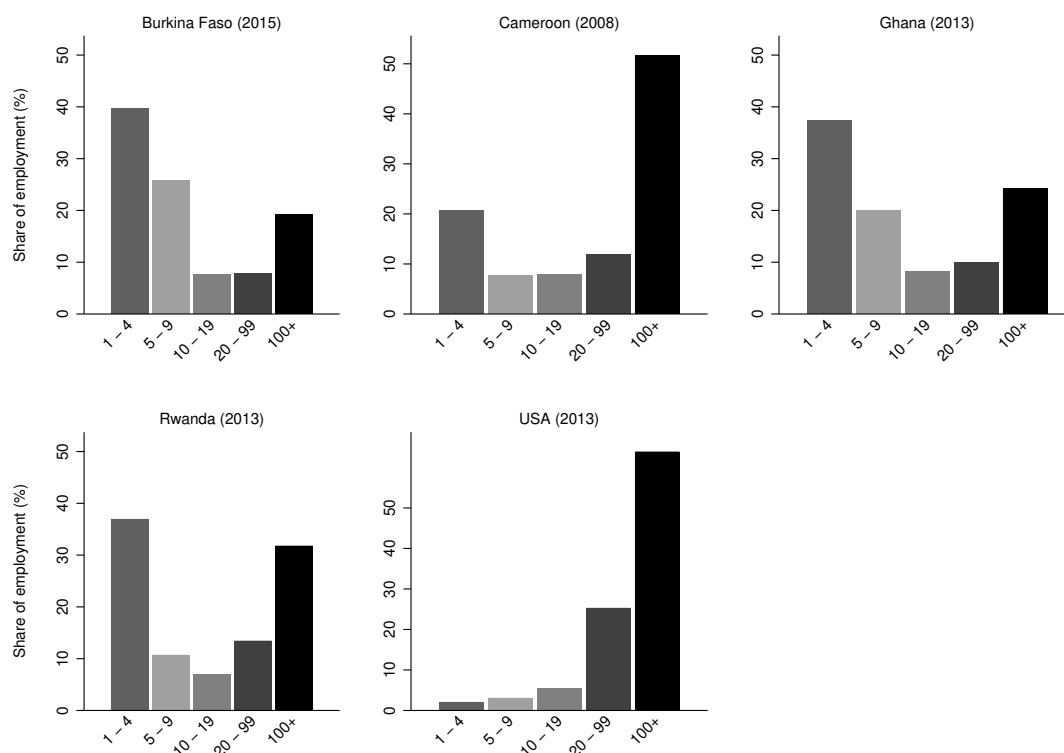
Source: Establishment censuses obtained from the statistical agencies of the select countries; see section 3. For the U.S., the data comes from the 2019 Business Dynamics Statistics (BDS) dataset. *Note:* The figure is constructed based on data for establishments in the manufacturing sector irrespective of their state or foreign ownership status. Establishments with missing employment data are excluded.

Figure 3: Employment distribution: Economy-wide



Source: Establishment censuses obtained from the statistical agencies of the select countries; see section 3. For the U.S., the data comes from the 2019 Business Dynamics Statistics (BDS) dataset. *Note:* The figure is constructed based on data for establishments in all sectors irrespective of their state or foreign ownership status. Establishments with missing employment data are excluded.

Figure 4: Employment distribution: Manufacturing

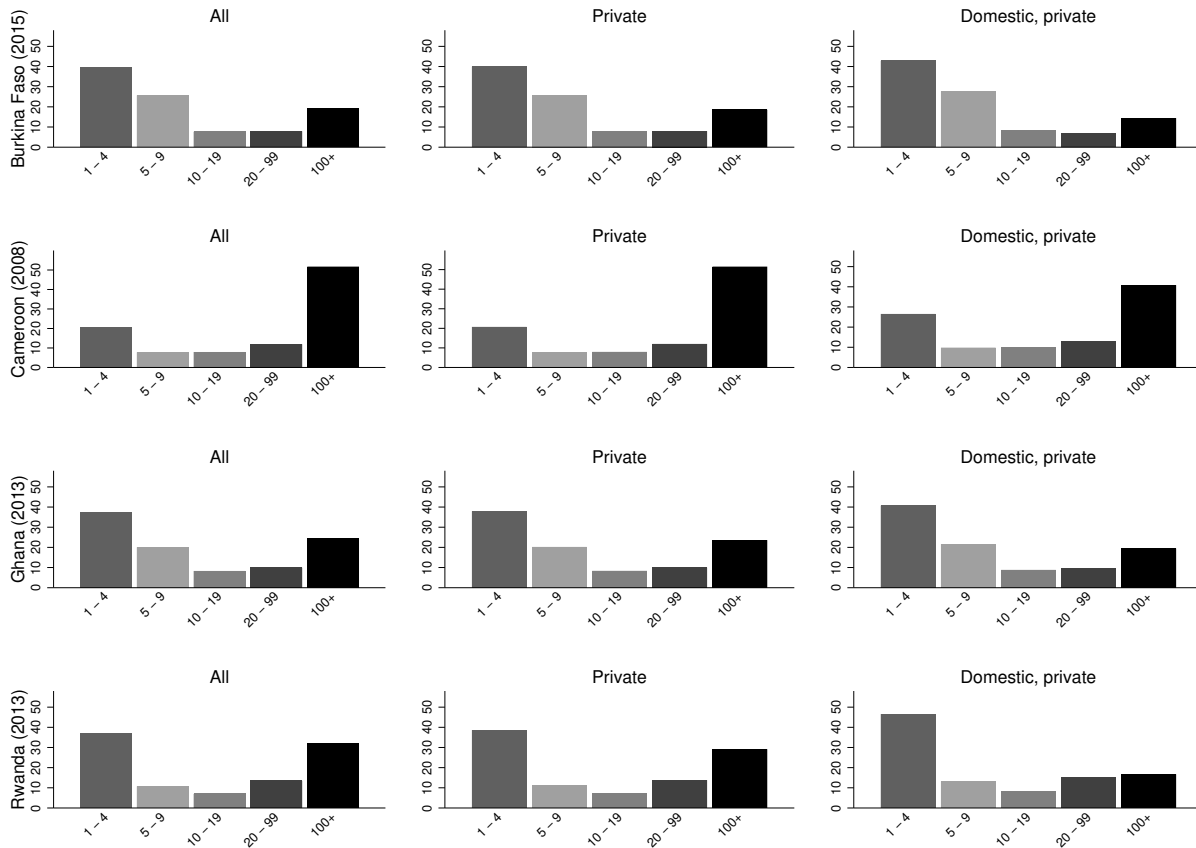


Source: Establishment censuses obtained from the statistical agencies of the select countries; see section 3. For the U.S., the data comes from the 2019 Business Dynamics Statistics (BDS) dataset. *Note:* The figure is constructed based on data for establishments in the manufacturing sector irrespective of their state or foreign ownership status. Establishments with missing employment data are excluded.

Furthermore, figure 5 shows that most of the employment is concentrated in the lower and uppermost tails. Apart from Cameroon, where most of the workforce is employed in establishments with at least 100 workers, the employment share of the smallest size category (1-4 workers) is the highest for the manufacturing sectors of Burkina Faso, Ghana, and Rwanda. At the same time, the next largest employment share belongs to the uppermost tail of the distribution, the size class of 100 or more workers.

An in-depth assessment of the size distribution in manufacturing uncovers that the feature of the missing middle is not attributed to few yet large foreign-owned establishments or state-owned enterprises. This observed pattern hold for private (those that are not state-owned), domestic (in which local citizens are principal owners), or domestic, private manufacturing establishments. Moreover, the missing middle phenomenon does not emerge from a specific way we slice the data, rather it reflects the inherent feature of the size distribution in the sample countries. As a robustness check, we plot the size distribution based on broader size categories. The key takeaways remain the same (see figures A.3, A.4, and A.5 in the appendix).

Figure 5: Employment distribution by state and foreign ownership: Manufacturing



Source: Establishment censuses obtained from the statistical agencies of the select countries; see section 3. Note: The figure is constructed based on data for establishments in the manufacturing sector. Establishments with missing employment data are excluded.

Table 3: Difference Between Actual vs Predicted Employment Shares (based on Pareto distribution)

	κ	Size class – Number of workers				
		1 – 4	5 – 9	10 – 19	20 – 99	100+
Panel a - Narrow size classes						
Burkina Faso (2015)	1.383	-0.013	0.150	-0.013	-0.066	0.014
Cameroon (2008)	1.149	0.021	0.012	0.015	-0.015	0.011
Ghana (2013)	1.336	0.007	0.098	-0.006	-0.053	0.023
Rwanda (2013)	1.299	0.047	0.012	-0.015	-0.017	0.039
Panel b - Broad size classes						
		1 – 9	10 – 49	50+		
Burkina Faso (2015)	1.454	0.027	-0.052	0.044		
Cameroon (2008)	1.143	0.015	-0.011	0.008		
Ghana (2013)	1.366	0.028	-0.042	0.032		
Rwanda (2013)	1.294	0.022	-0.024	0.019		

Source: Establishment censuses obtained from the statistical agencies of the select countries; see section 3. *Note:* The Pareto shape parameter κ is estimated based on minimizing Euclidean distance between observed and theoretical shares.

Also when we compare the distribution to that of a Pareto distribution, often taken as the firm size distribution in an undistorted economy (see e.g., (E. G. J. Luttmer, 2007)), we see that there is an underrepresentation of medium-sized firms in aggregate employment. This is the metric used by ?. Assuming that the efficient cumulative size distribution is described by $F(\ell) = 1 - \ell^{-\zeta}$ where ℓ is firm size as measured by the number of worker and ζ is the country-specific shape parameter that accounts for cross-country variation the scale of production, it is straightforward to show that the efficient share of manufacturing labor force employed in firms of size range $\ell_i \leq \ell < \ell_j$ is given by $\hat{s}(\ell_i, \ell_j | \zeta) = \ell_i^{1-\zeta} - \ell_j^{1-\zeta}$. For each of the bins laid out above, we calibrate ζ by country based on the minimization of the Euclidean distance between observed shares and theoretical shares given by the equation above (Axtell 2001; Di Giovanni, Levchenko, and Ranciere 2011; Tybout 2014). We then compute the theoretical shares \hat{s} based on the estimated shape coefficients $\hat{\zeta}$, and compare them to the actual shares. Table 3 reports the estimation results and shows that that the middle category is underrepresented, irrespective of the binning.

The preponderance of informal firms is the main driver of the missing middle

Fact 2. *The preponderance of informal firms is the main driver of the missing middle*

Roughly speaking, a formal establishment is a business entity with a registration record at an administrative agency such as a business registry or tax authority. For the baseline analysis, registration status at the tax authority is used as an indicator of formality. We choose this indicator because of its availability and comparability across the sample countries. The only exception is Burkina Faso, for which we use the information on whether an establishment applies the SYSCOA accounting standard to define formality.

Not surprisingly, we find that formal establishments constitute a very small percentage of the manufacturing plants in sub-Saharan Africa (table 2). In Burkina Faso, they constitute about 4 percent of the establishments and hire about 27 percent of the workforce in manufacturing. The respective figures for Cameroon, Ghana, and Rwanda are 5 and 59 percent; 28 and 50 percent, and 21 and 59 percent, respectively. These results indicate that a significant number of the manufacturing establishments and their workforce operates in the informal sector.¹⁵

As shown in figure 6, informal establishments are behind the observation of the missing middle in the employment distribution. Informal establishments are primary sources of employment in the lower ranges of the production scale, but their relevance sharply diminishes as the scale of production goes up. In view of the practical difficulty of defining informality, we plot the employment distribution using alternative criteria for classifying establishments into formal and informal. We see that driver of the missing middle is robust to alternative definitions of informality.

It is necessary to emphasize that the pervasiveness of informality in these countries is not a trivial matter. To begin with, there exist sizable growth and productivity differences between formal and informal establishments, having a serious implication for job creation and standard of living. Another concern is the issue of narrow tax bases that government would undermine a country's capacity to finance the provision of public services. Further, informality adversely affects welfare simply because of low wages, poor working conditions, and limited benefits to workers. On the

¹⁵ See tables A.1–A.4 in the appendix for summary using alternative definitions of formality.

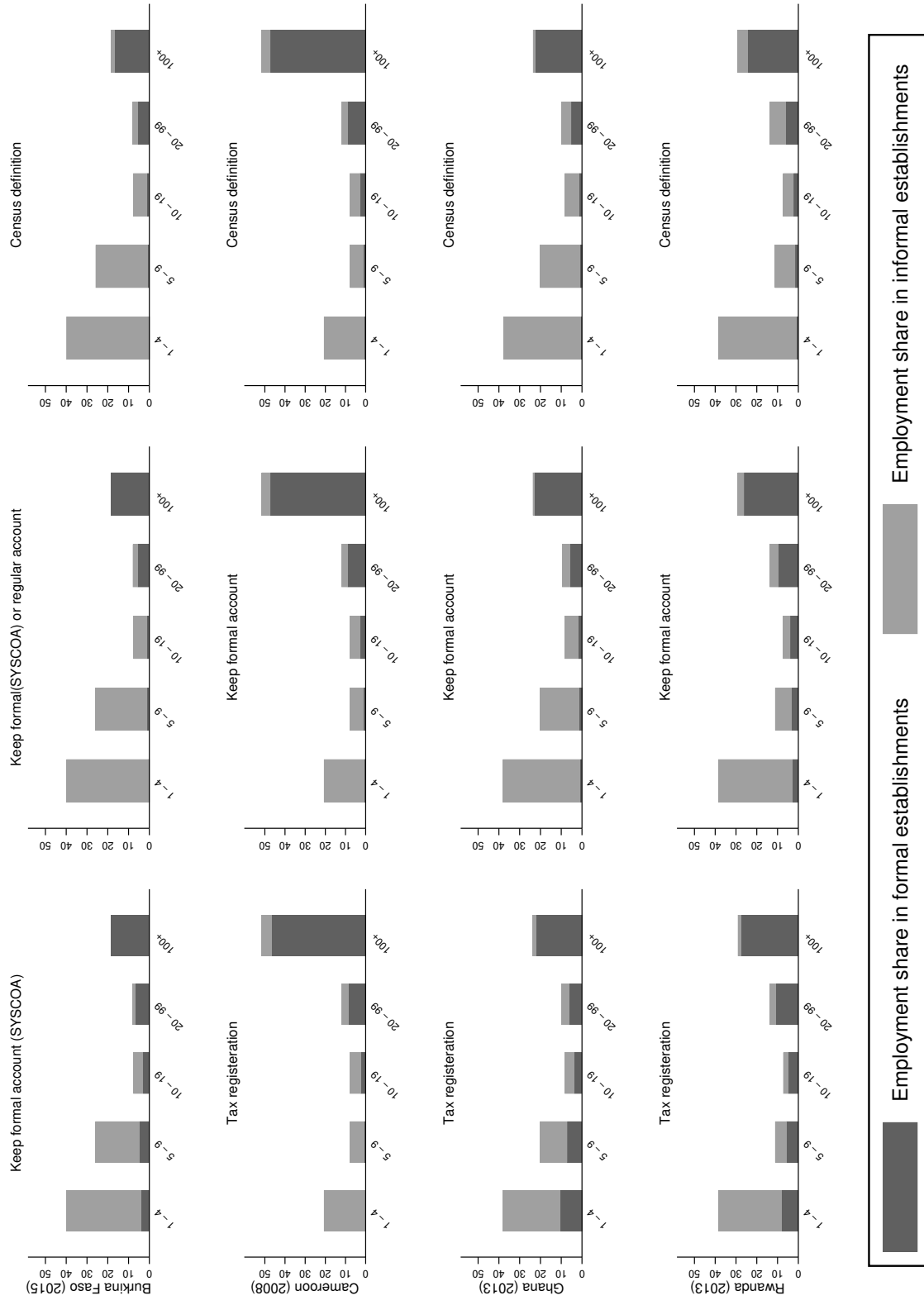
other hand, informality reflects a bad business environment such as unreliable, low-quality public services as well as weak enforcement of rules and regulations that make informality attractive.

The size distribution of formal firms mirrors the distribution in developed economies

Fact 3. *The size distribution of formal firms mirrors the distribution in developed economies*

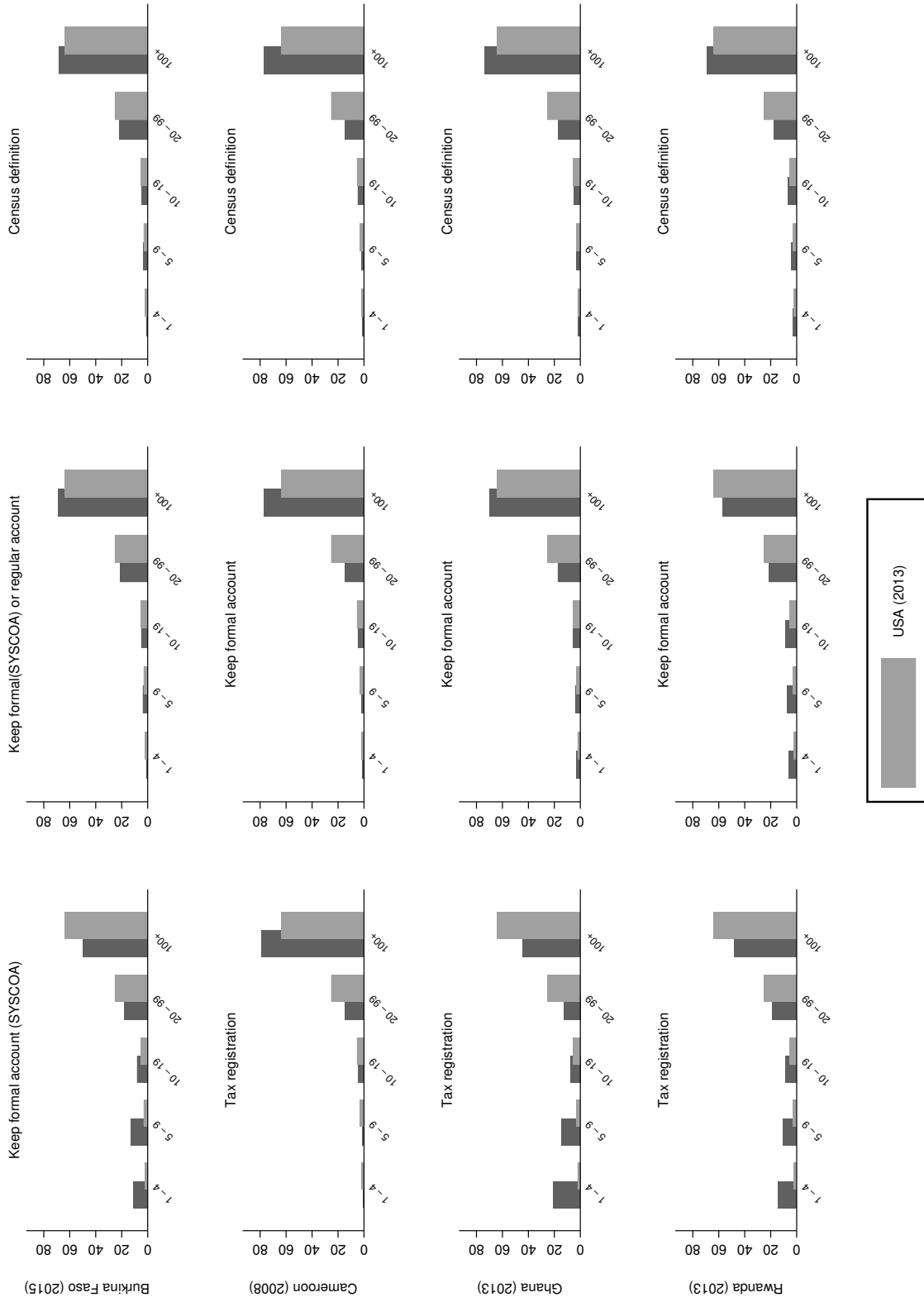
Figure 7 reveals that formal manufacturing has a strikingly different pattern compared with the rest of manufacturing; it features a high concentration of employment in the upper tail of the size distribution. Most importantly, this is very comparable to the size distribution of developed countries. It appears that the population of formal firms in sub-Saharan African countries seems to obey the same distribution law as in developed economies even if the business environment in sub-Saharan Africa is characterized by a multitude of market distortions and policy uncertainties. Again, this phenomenon holds no matter the choice of criteria used to define formality.

Figure 6: Formal and informal employment distribution: Manufacturing



Source: Establishment censuses obtained from the statistical agencies of the select countries; see section 3. *Note:* The figure is constructed based on data for establishments in the manufacturing sector. Establishments with missing employment data and state-owned are excluded. Census definition of formality slightly differs across the statistical offices of the sample countries. In Burkina Faso, an establishment is considered formal if it has a registration number (with mobile trade and mobile credit, tax identification number) and uses SYSCOA as an accounting standard. In Cameroon, formal establishments are the ones that prepare a Statistical and Fiscal Declaration (DSF) or those that do not prepare DSF but keep an operating account and a partial account of balance sheets. In Ghana, formal establishments are defined as those that are registered with the Registrar General's Department (RGD) and keep formal accounts. In Rwanda, an establishment is considered to operate in the formal sector if it is registered at Rwanda Revenue Authority (RRA) and maintains regular operational accounts.

Figure 7: Formal employment distribution compared with the U.S.: Manufacturing



Source: Establishment censuses obtained from the statistical agencies of the select countries; see section 3. *Note:* The figure is constructed based on data for establishments in the manufacturing sector. Establishments with missing employment data and state-owned and state-owned are excluded. Census definition of formality slightly differs across the statistical offices of the sample countries. In Burkina Faso, an establishment is considered formal if it has a registration number (with mobile trade and mobile credit, tax identification number) and uses SYSCOA as an accounting standard. In Cameroon, formal establishments are the ones that prepare a Statistical and Fiscal Declaration (DSF) or those that do not prepare DSF but keep an operating account and a partial account of balance sheets. In Ghana, formal establishments are defined as those that are registered with the Registrar General's Department (RGD) and keep formal accounts. In Rwanda, an establishment is considered to operate in the formal sector if it is registered at Rwanda Revenue Authority (RRA) and maintains regular operational accounts.

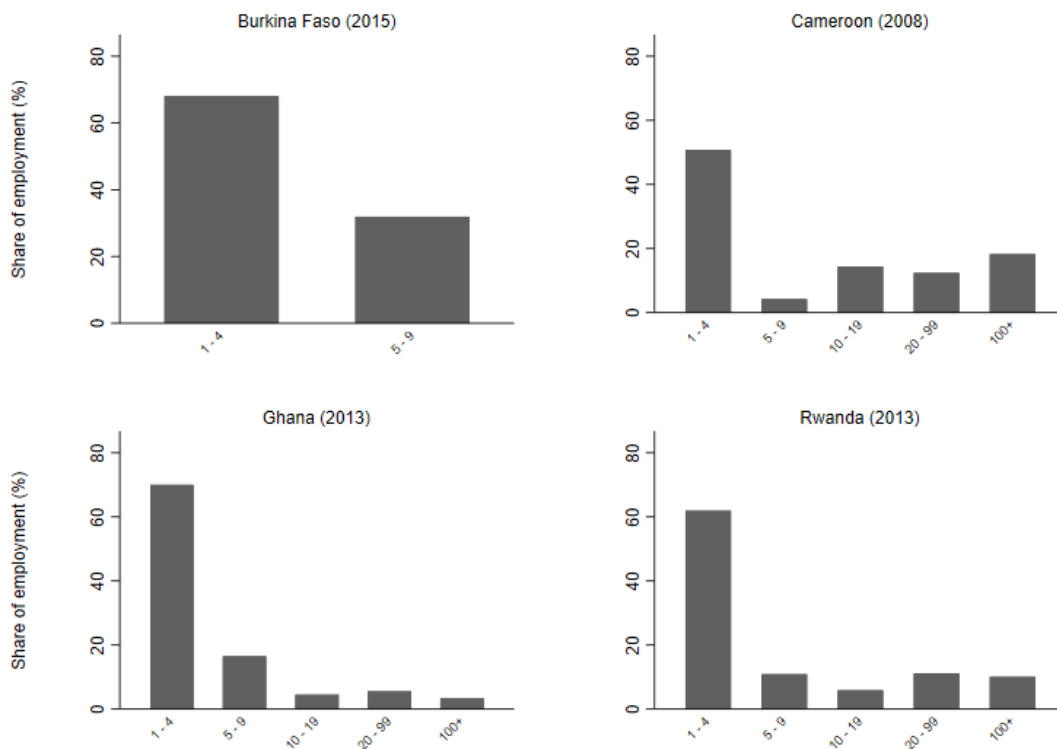
The missing middle does not arise from entry

Fact 4. *The missing middle does not arise at entry*

The last finding from our exploration is the fact that the missing middle pattern does not seem to arise from the size distribution of entrants. As a corollary, this suggests that the missing middle has to emerge as the outcome of the life cycle of manufacturing firms. Figure 8 plots the the size distribution of new cohort of entrants. We see that employment in the new cohort of establishments is concentrated in the smallest class and distributed more evenly across the remaining size classes.

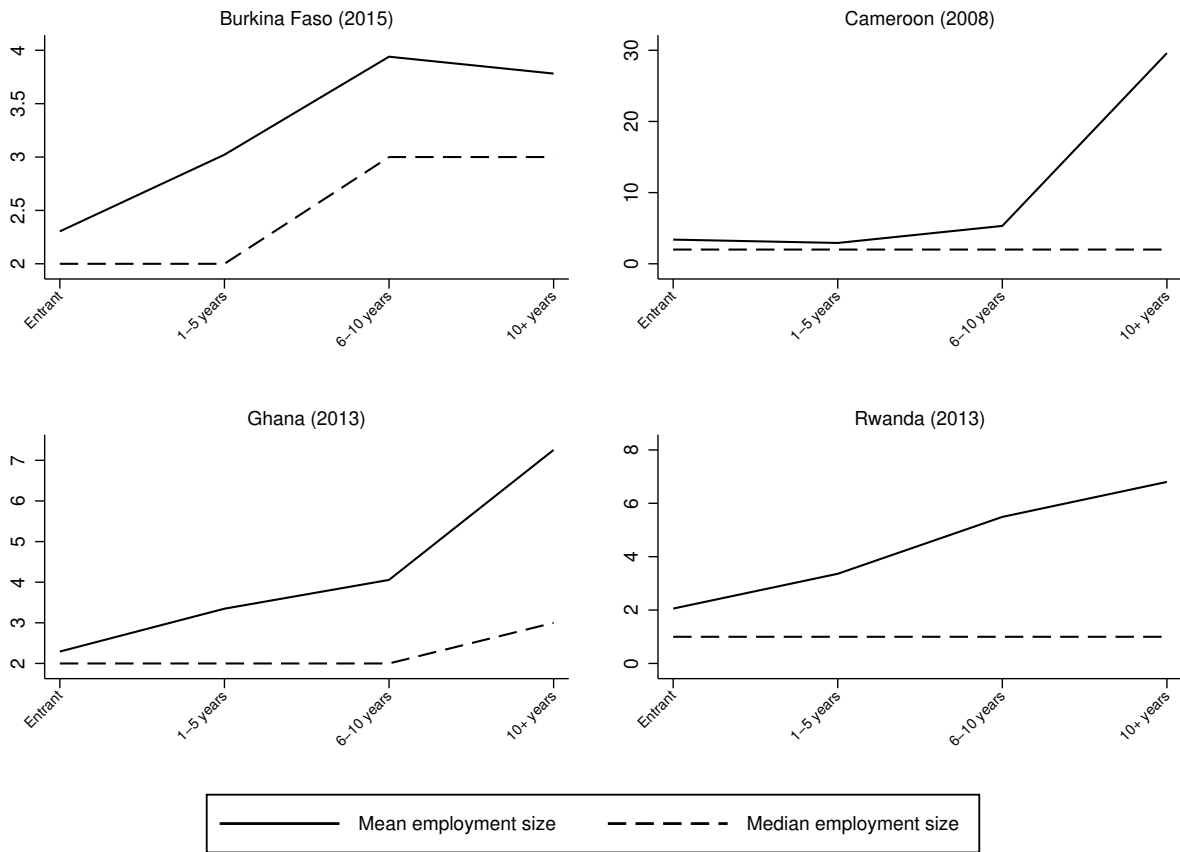
Another factor possibly explaining the missing middle is the pattern of growth along the life cycle of establishments. That is, variations in growth along with selection can drive employment distribution. For this purpose, establishments are divided into six age cohorts: entrant, 1-5, 6-10, 11-15, 1-20, and above 20 years. Figure 9 shows that entrants are the smallest compared with other age cohorts. There also exist a few larger establishments, as evidenced by the mean size being exceedingly higher than the median size for each age cohort, and the difference is even more magnified among older age cohorts. Two things are worth noting. The first one is the small difference between the mean and median size for Burkina Faso. Second, the median establishment does not see a significant increase in its size over its life cycle.

Figure 8: Employment distribution among entrants: Manufacturing



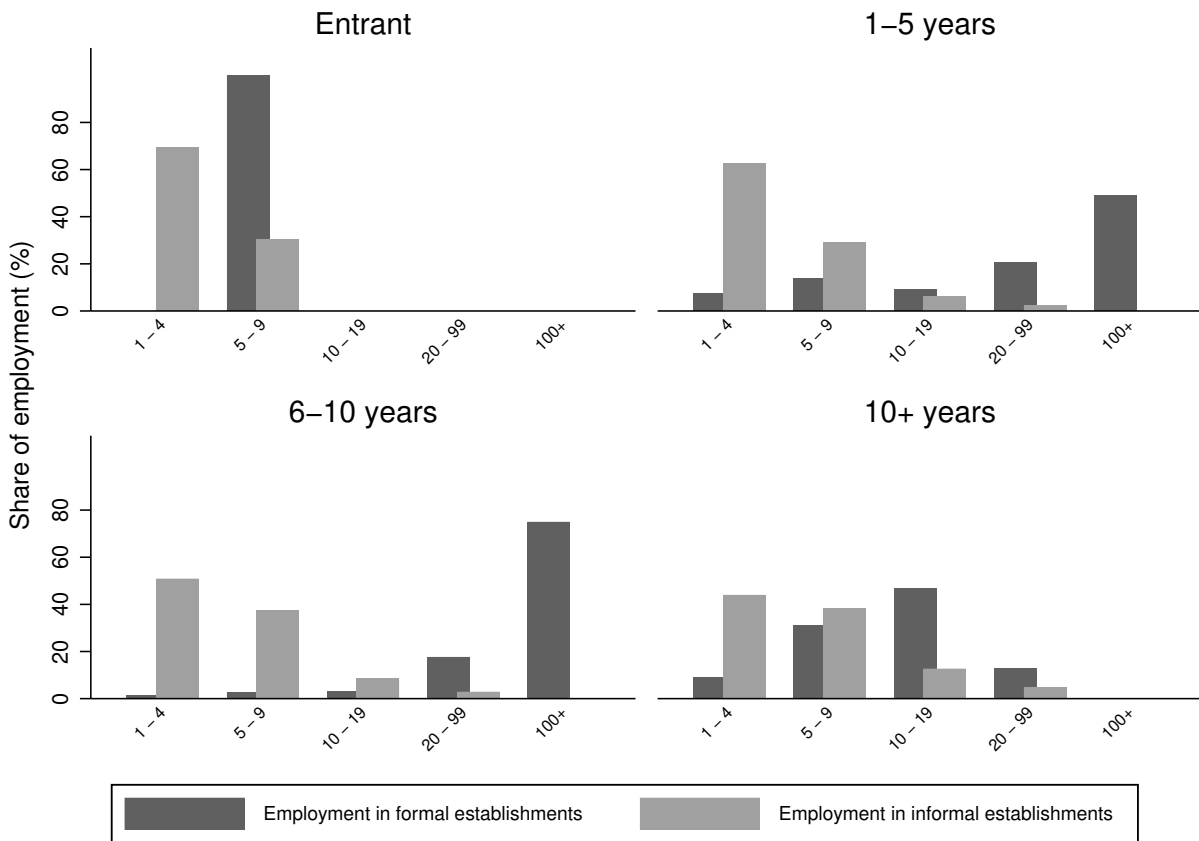
Source: Establishment censuses obtained from the statistical agencies of the select countries; see section 3. *Note:* The figure is constructed based on data for establishments in the manufacturing sector. Establishments with missing employment or age data and state-owned are excluded.

Figure 9: Size of establishments by age cohort: Manufacturing



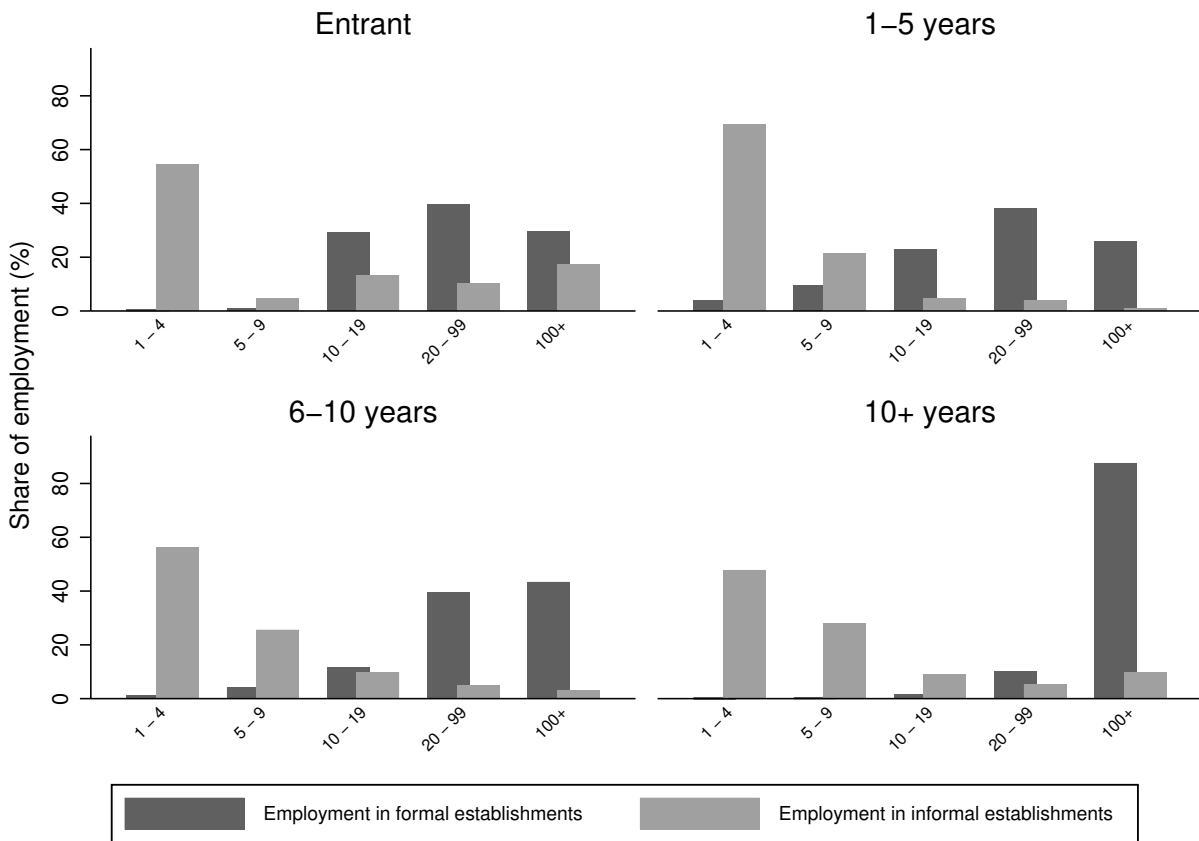
Source: Establishment censuses obtained from the statistical agencies of the select countries; see section 3. Note: The figure is constructed based on data for establishments in the manufacturing sector. Establishments with missing employment or age data and state-owned are excluded.

Figure 10: Formal and informal employment distribution by age cohort in manufacturing: Burkina Faso (2015)



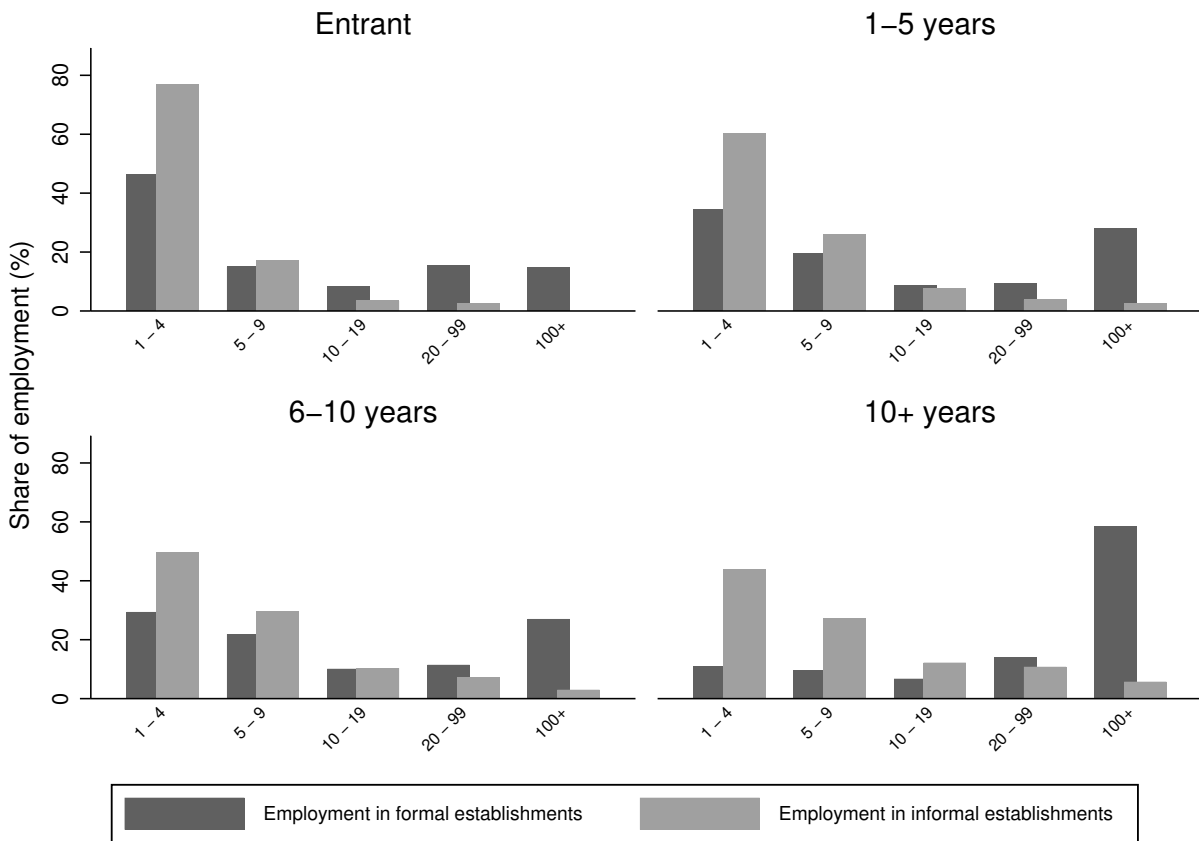
Source: Establishment censuses obtained from the statistical agencies of the selected countries; see section 3. Note: The reference is the total employment for each age cohort and formality status. The figure is constructed based on data for establishments in the manufacturing sector. Establishments with missing employment or age data and state-owned are excluded.

Figure 11: Formal and informal employment distribution by age cohort in manufacturing: Cameroon (2008)



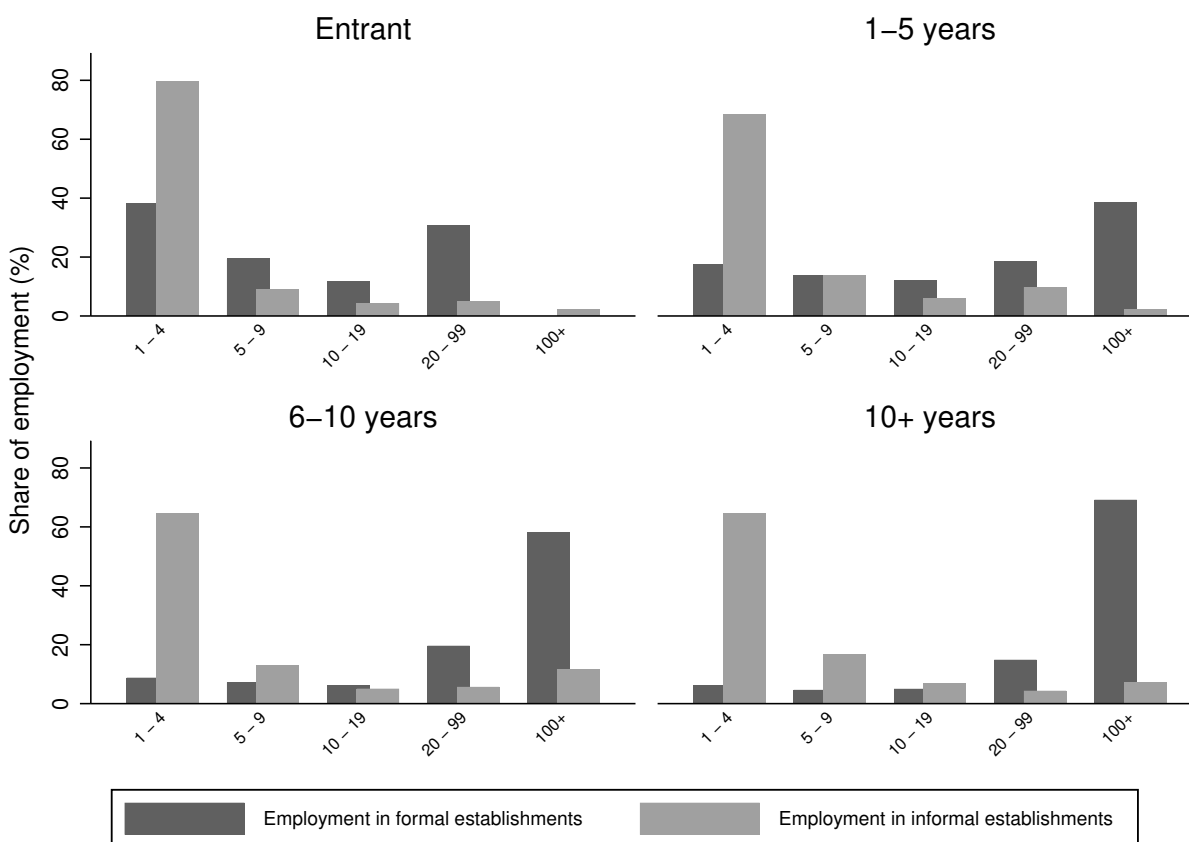
Source: Establishment censuses obtained from the statistical agencies of the selected countries; see section 3. Note: The reference is the total employment for each age cohort and formality status. The figure is constructed based on data for establishments in the manufacturing sector. Establishments with missing employment or age data and state-owned are excluded.

Figure 12: Formal and informal employment distribution by age cohort in manufacturing: Ghana (2013)



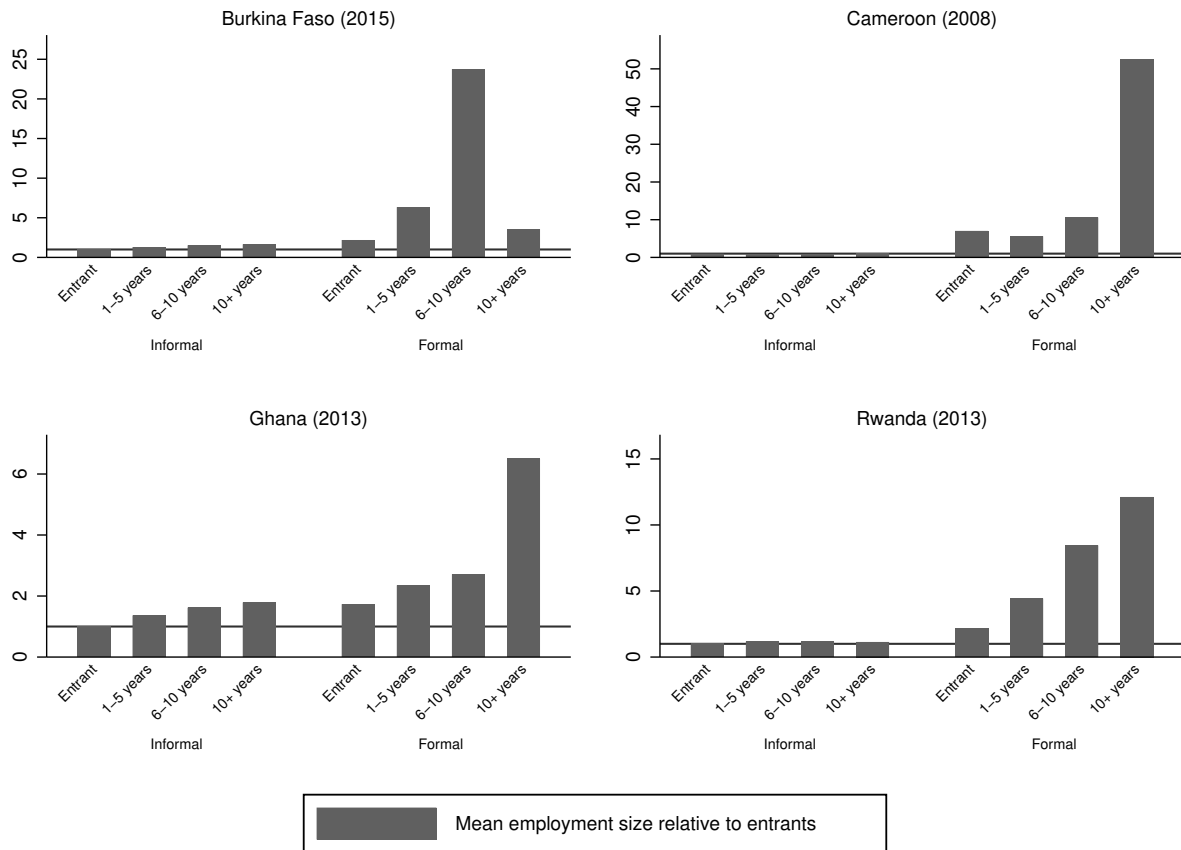
Source: Establishment censuses obtained from the statistical agencies of the selected countries; see section 3. Note: The reference is the total employment for each age cohort and formality status. The figure is constructed based on data for establishments in the manufacturing sector. Establishments with missing employment or age data and state-owned are excluded.

Figure 13: Formal and informal employment distribution by age cohort in manufacturing: Rwanda (2013)



Source: Establishment censuses obtained from the statistical agencies of the selected countries; see section 3. Note: The reference is the total employment for each age cohort and formality status. The figure is constructed based on data for establishments in the manufacturing sector. Establishments with missing employment or age data and state-owned are excluded.

Figure 14: Size of formal and informal establishments by age cohort: Manufacturing



Source: Establishment censuses obtained from the statistical agencies of the selected countries; see section 3. Note: The reference is the total employment for each age cohort and formality status. The figure is constructed based on data for establishments in the manufacturing sector. Establishments with missing employment or age data and state-owned are excluded.

Figures 10–13 illustrate growth differentials between formal and informal establishments by plotting the employment distribution at different age cohorts. Among formal establishments, the employment shares of larger class sizes continuously increase when looking at older age cohorts. These establishments tend to increase in size with additional years of market operation even if they face challenges in the business environment that stifle their growth. As for the informal establishments, the smallest size class has the highest employment shares and displays only a slight decline with older age cohorts. We also note the presence of a few informal establishments that are large or have been in operation for a number of years.

Similarly, figure 14 depicts the size of incumbents relative to new informal establishments. Here, we also see that incumbents experience size growth with age, but this tendency is most common among formal establishments rather than informal ones. Note that this pattern holds to a lesser extent in Burkina Faso.

5 Model

Our deconstruction of the missing middle led us to interpret the evidence as suggestive of distortions that encourage informality and discourage formal firm growth. Entry barriers and productivity-dependent idiosyncratic distortions have the potential of generating such outcomes. While the formal sector in the low-income countries under study exhibited a notable dynamism when compared with its informal counterpart, it still fell short of the dynamism of firms in advanced economies. This leaves a role to be played by productivity-dependent distortions, which discourage firm growth and misallocate resources across firms. Entry barriers, in turn, limit the incentives of firms to enter the formal sector and magnify the effect of idiosyncratic distortions.

We investigate this hypothesis formally in this section by feeding entry barriers and idiosyncratic distortions into a model of firm dynamics with endogenous informality. Our goal is to assess the extent to which these distortions can account for the prevalence of informality and the patterns of the size distribution we observed in the data and to quantify their aggregate implications. The model features a rich formal sector, with endogenous entry and exit of firms and endogenous innovation decisions, all responsive to the incentives created by the distortions. The informal sector also features endogenous entry but we assume that firm growth and the exit rate are exogenous so that the size distribution is governed by the rate of productivity growth over the firms' life cycle vis-a-vis the exit rate. We assume that formal and informal firms produce differentiated varieties that aggregate into differentiated baskets of formal and informal goods. Consumers face a constant elasticity of substitution across varieties within a type of good and across formal and informal goods. We calibrate the model to the U.S. and estimate entry barriers and idiosyncratic distortions in Ghana, the only country in our sample for which we also count with a companion formal sector census of industrial firms with balance sheet information.

5.1 Demand Side

Consumers derive utility from baskets of formal and informal goods, which they combine through a Constant Elasticity of Substitution demand system. Formally:

$$Q = \left[Q_1^{\frac{\lambda-1}{\lambda}} + Q_2^{\frac{\lambda-1}{\lambda}} \right]^{\frac{\lambda}{\lambda-1}}$$

where Q is the final good and Q_1 and Q_2 are formal and informal good baskets that go into the production of the final good¹⁶. In our calibration, parameters will be set so that the equilibrium expenditure and employment share in the informal sector are low, targeting values of the informal sector share in the U.S. economy. When distortions are identified and fed into the model, the elasticity of substitution λ will be a key determinant of the strength of the response of the informal sector to the distortions.

Formal and informal good baskets, in turn, are produced according to a Constant Elasticity of

¹⁶One could have directly defined Q to be aggregate consumption, where the aggregation is done at the level of the household. This is isomorphic to our specification of Q as a good produced by a final good aggregator.

Substitution aggregation of a continuum of differentiated varieties, given by:

$$Q_i = \left[\int \left[q_i^d(\omega) \right]^{\frac{\theta-1}{\theta}} M_i(e^\omega) d\omega \right]^{\frac{\theta}{\theta-1}} \quad (1)$$

where, hereafter, $i = 1$ and $i = 2$ shall denote the formal and the informal sectors respectively. In turn, $M_i(e^\omega)$ stands for the number of firms in sector i producing a variety with idiosyncratic productivity e^ω , to be determined endogenously in equilibrium.

Assuming a perfectly competitive representative producer of the aggregate final good and the aggregate basket of informal and formal varieties leads to the following demand functions for a given variety in a sector:

$$p_1(\omega) = P_1^{\frac{\theta-\lambda}{\theta}} Q^{\frac{1}{\theta}} q_1(\omega)^{\frac{-1}{\theta}} \quad (2)$$

Moreover, setting the final good as the numeraire, the aggregate and sectoral price indices are given by:

$$1 = \left[P_1^{1-\lambda} + P_2^{1-\lambda} \right]$$

$$P_i = \left[\int p_i(\omega)^{1-\theta} M_i f(e^\omega) d\omega \right]^{\frac{1}{1-\theta}} \quad (3)$$

5.2 Production Side

Consider now the producers of intermediate varieties of each good. Production is done under monopolistic competition according to the following technology:

$$y_i(\omega) = (A_i)^{\frac{1}{\theta-1}} (e^\omega)^{\frac{1}{\theta-1}} l(\omega) \quad (4)$$

where A_i stands for a sector-wide productivity term and e^ω denotes the idiosyncratic productivity of intermediate producers. Labor is the sole factor of production. In what follows, we shall assume that A_i is constant over time and common across all firms in sector i , while there is a stochastic process of idiosyncratic productivity to be specified below. It is in the nature of this stochastic process that the differences in the production possibilities across formal and informal producers rest.

Profit maximization under monopolistically competition involves solving the following optimization problem (we present it for a formal sector producer with the understanding that an equivalent problem is solved in the informal sector):

$$\max_{l_1} \left\{ (1 - \tau_\omega) P_1^{\frac{\theta-\lambda}{\theta}} Q^{\frac{1}{\theta}} (e^\omega)^{\frac{1}{\theta}} A_1^{\frac{1}{\theta}} l_1(\omega)^{\frac{\theta-1}{\theta}} - w l_1(\omega) \right\}$$

where we have plugged in the demand function solved in equation 2.

The idiosyncratic distortion, one of the types of distortion we consider, is reflected as a revenue tax in the producer's profit maximization problem. The key feature of this tax (interchangeably referred to as wedge) is that it is idiosyncratic and systematically related to the firm's idiosyncratic

productivity e^ω , consistently with the misallocation literature.¹⁷ Formally, we characterize this relationship with the following function:

$$(1 - \tau_\omega) = (e^\omega)^{\frac{-\gamma_i}{\theta-1}} \quad (5)$$

The key parameter governing the elasticity between $TFPR$, the measure of distortions, and $TFPQ$, the idiosyncratic productivity, is given by γ_i . Our calibration section discusses how we estimate this elasticity in the Ghanaian firm-level data.

Optimization leads to the following expressions for labor demand and variable profits:

$$l_1(\omega) = \left(\frac{\theta-1}{\theta}\right)^\theta \left(\frac{Q}{w^\theta}\right) P_1^{(\theta-\lambda)} e^\omega A_1 (1 - \tau_\omega)^\theta \quad (6)$$

$$\pi_1^v(\omega) = \frac{(\theta-1)^{\theta-1}}{\theta^\theta} P_1^{(\theta-\lambda)} \frac{Q}{w^{\theta-1}} e^\omega A_1 (1 - \tau_\omega)^\theta \quad (7)$$

5.3 Firm Dynamics, Entry, and Exit in the Formal Sector

Our characterization of the formal sector features the first-order margins of adjustments through which distortions affect micro and macro-level outcomes. These are an endogenous entry and exit of firms into the formal sector, and an innovation technology that allows for an endogenous component of firm dynamics.

Endogenous firm dynamics in the formal sector arise from a process of technological upgrading and downgrading similar to that in (Atkeson & Burstein, 2010). Specifically, a firm with current productivity e^ω can upgrade to $e^{\omega+\Delta}$ with probability $q_t(\omega)$ and can downgrade to $e^{\omega-\Delta}$ with probability $(1 - q_t(\omega))$. The expected growth rate, given by $q_t(\omega)$, is endogenous, as firms can allocate resources to innovation activities. The variance of the shock process (Δ), on the other hand, is exogenous. The labor-denominated cost for attaining a desired probability $q_t(\omega)$ is given by

$$\chi(q_t, \omega) = e^\omega \times \eta \left(e^{\phi q_t} - 1 \right).$$

Notice that the innovation cost is scaled by the entrepreneur's current productivity. This is an important assumption that allows the model to be consistent with innovation patterns of large firms in the US, our target economy for the calibration. The scale parameter η and the elasticity parameter ϕ will be calibrated to replicate the properties of the size distribution and the life cycle of firms in the US.

The value of an operating formal firm with current productivity e^ω is

$$v_{1,t}^o(\omega) = \max_{q_t(\omega)} \left\{ \begin{array}{l} \pi_{1,t}^v(\omega) - w_t \chi(q_t, \omega) - w_t f_c \\ + R_t (1 - \delta) [q_t(\omega) v_{1,t+1}(e^{\omega+\Delta}) + (1 - q_t(\omega)) v_{1,t+1}(e^{\omega-\Delta})] \end{array} \right\} \quad (8)$$

where $\pi_{1,t}^v(\omega)$ is the indirect variable profit function under optimal factor demands given prices and distortions, defined in equation 7, f_c denotes the labor-denominated fixed cost of production, and

¹⁷ See, for instance (Bento & Restuccia, 2014) and (Fattal-Jaef, 2022) for cross-country estimates of the relationship between idiosyncratic distortions and idiosyncratic productivity

δ stands for the exogenous exit probability. While there is an exogenous exit shock, the presence of fixed costs gives rise to an endogenous motivation for the least profitable businesses to exit the market.

The value of the formal firm is given by:

$$v_{1,t}(\omega) = \max_{\iota_{1,t}(\omega)} \{v_{1,t}^o(\omega), 0\},$$

where $\iota_{1,t}(\omega)$ encodes the firm's exit decision, equal to one if it operates and equal to zero if it exits.

The first-order condition with respect to the innovation choice yields

$$w_t \phi \mu e^{\phi p(\omega)} = R_t (1 - \delta) [v_{1,t+1}(e^{\omega+\Delta}) - v_{1,t+1}(e^{\omega-\Delta})].$$

Firms choose the probability of a technological upgrade so as to equate the marginal cost of innovation efforts with the gain in firm value. To the extent that idiosyncratic distortions are productivity dependent, as we are assuming in equation 5, these distortions will have a direct contribution on the rate of return to innovation expenses, in addition to general equilibrium effects on wages and final demand. Notice that entry barriers, which we have not introduced yet, have no direct effect on the firms' innovation decisions (though they will still have an indirect contribution through general equilibrium forces). Lastly, general equilibrium forces shaping the relative price between formal and informal goods and the wage rate will also impact the firm's incentives to innovate.

We turn now to discuss the entry technology into the formal sector. As standard in models of firm dynamics, we assume there is an infinite pool of potential entrants considering incurring a labor-denominated entry cost f_e to enter the market. At the time of confronting this cost, there is no knowledge of the idiosyncratic productivity but there is knowledge about the distribution $\Gamma(\omega)$ from which productivity is drawn. In equilibrium, a free-entry condition holds equating the cost of entry with the expected net present value of profits. In the distorted economy, the entry cost is inflated by an entry barrier τ^e , which is the second type of distortion we consider for rationalizing the informality in the data. The free-entry condition in this context is given by

$$w_t f_e (1 + \tau^e) = R_t (1 - \delta) \int v_{1,t+1}(\omega) d\Gamma(\omega). \quad (9)$$

Notice that we are assuming a one-period time to build between the payment of the entry cost and the actual entry.

5.4 Firm Dynamics, Entry, and Exit in the Informal Sector

Our characterization of the informal sector loads the replication of the informal firms' size distribution on the productivity distribution across firms. That is, the informal firms' size distribution will be matched by construction as an outcome of the calibration of the productivity distribution among informal producers. The whole economy's size distribution, on the contrary, will be shaped by equilibrium variables in the model, namely the fraction of informal employment and the distribution of innovation expenses across formal firms. Since these two are active margins of adjustment in response to the entry barrier and the idiosyncratic distortions, they deliver testable implications

to be validated by the data.

Given this calibration choice, the informal firms' productivity dynamics are given by a deterministic growth process subject to exogenous exit shocks. The exogenous growth and exit rates, μ and δ , jointly shape the age and size distributions across firms ((E. G. Luttmer, 2010). Formally the idiosyncratic productivity of a firm with age a is given by:

$$e^{\omega(a)} = e^{\mu_i a}$$

Given the exogenous death rate, the fraction of active informal firms of age a is given by:

$$f(a) = e^{-\delta_i a}$$

As said earlier, the production technology of informal firms adopts the same shape as that for formal producers, given in equation 4. Therefore, the outcome of the static optimization problem for the informal producer is given by the same expressions characterizing labor demand and variable profits in the formal sector. The sole difference between the two is that we abstract from idiosyncratic distortions across informal firms.¹⁸

Entry into the informal sector also entails a labor-denominated entry cost but is free from any entry barrier. The free-entry condition into the informal sector, then, is governed by:

$$w f_{e2} = \int \pi_2^v(a) e^{-(\rho+\delta)a} da$$

where f_{e2} is the entry cost into the informal sector, ρ is the subjective discount factor, and $e^{-(\rho+\delta)a}$ characterizes the expected productivity of the firm over its life-cycle. As was the case for the formal sector, f_{e2} captures non-distortionary elements of the cost of setting up an informal firm, such as the opportunity cost of the time spent implementing the business idea. Naturally, our calibration strategy will deliver a significantly lower cost of entry into the informal sector.

5.5 Households

In addition to the static problem of choosing formal and informal expenditure, the representative household confronts the dynamic decision of how to allocate its wealth into consumption and savings, and how to split the savings into a risk-free bond and the portfolio of formal and informal firms in the economy. Given an initial endowment of formal and informal firms $M_1(0)$ and $M_2(0)$, the household's dynamic problem is formally given by:

$$\int e^{-\rho t} \log [c(t)] dt,$$

¹⁸ Our rationale for abstracting from distortions across informal producers is that the scope for allocative efficiency gains is very limited given the dramatic concentration of production among very small producers. Our interpretation is that the size distribution of informal firms is a reflection of poor productivity or entrepreneurial ability rather than a manifestation of an idiosyncratic distortion.

subject to

$$\dot{a}(t) = \frac{r(t)a(t) + w(t)L + M_1 \int \pi_1^v(a, t) f(a) da + M_2 \int \pi_2^v(a, t) f(a) da - f_{e1} M_{e1}(t) w(t) - f_{e2} M_{e2}(t) w(t) + T}{M_2 \int \pi_2^v(a, t) f(a) da - f_{e1} M_{e1}(t) w(t) - f_{e2} M_{e2}(t) w(t) + T}$$

$$\dot{M}_1 = -\delta_1 M_1 + M_{e1} \quad (10)$$

$$\dot{M}_2 = -\delta_2 M_2 + M_{e2} \quad (11)$$

where T stands for the transfer to/from the household that balances the budget from the collection of revenue from entry barriers and idiosyncratic distortions.¹⁹ The Euler equation ruling the incentives to accumulate assets implies that, in the stationary equilibrium, $\rho = r$.

5.6 Equilibrium

A stationary competitive equilibrium in the model consists of (i) consumption, firm entry in each sector, and saving decisions of the household $[c, \dot{a}, M_{e1}, M_{e2}, \dot{M}_1, \dot{M}_2]$; (ii) labor demands, prices, and variable profits for each variety $[l_i(\omega), p_i(\omega), \pi_i(\omega)]$; (iii) demand functions for intermediate inputs and final output, and sectoral price indices $[y_i(\omega), Y_i, P_i]$; (iv) the age distributions of informal firms, the schedule of idiosyncratic distortions, and the entry barrier: $[f_2(a), [1 - \tau_w], \tau^E]$; (v) laws of motion for the productivity distribution of formal firms and the number of informal producers; (vi) innovation and exit decisions of formal firms $q(\omega), \iota(\omega)$; and (vii) wages and interest rates such that: a) given wages, interest rates, and the laws of motion for the distribution of formal and informal producers (i) solves the household's optimization problem; b) given wages and idiosyncratic distortions, (ii) solves each variety producer's profit maximization problem and (vi) solves the value function of a formal firm; c) given the prices of formal and informal varieties, (iii) solves the final good's representative producer profit optimization problem; d) the free entry conditions in each sector hold; e) the labor market clears; f) net asset demand is equal to zero.

The labor market clearing condition in the equilibrium is given by:

$$L = L_{p1} + L_{fc} + L_I + L_{p2} + f_{e1} M_{e1} + f_{e2} M_{e2}. \quad (12)$$

where L_{pi} stands for the labor demand devoted to the production of formal and informal varieties, L_{fc} and L_I denote the labor demand associated with fixed and innovation costs in formal firms, and the remaining elements capture the labor demand allocated to the entry of new formal and informal producers.

The law of motion for the distribution of formal firms, in turn, is given by the following equation:

$$M_{t+1}(\omega') = (1 - \delta) q_t(\omega' - \Delta) M_t(\omega' - \Delta) + (1 - \delta) [1 - q_t(\omega' + \Delta)] M_t(\omega' + \Delta) + (1 - \delta) M_{e,t} \Gamma(\omega'). \quad (13)$$

The expression establishes that a fraction $(1 - \delta) q_t(\omega' - \Delta)$ of firms with productivity less than or equal to $(\omega' - \Delta)$ survives the exogenous exit shock and transitions to a productivity level less than

¹⁹ The lump-sum transfer is how we implement the idea that we do not interpret τ^E and $[1 - \tau_w]$ literally as taxes that drag resources away from the economy beyond the distortion they generate. In our model, output and consumption are reduced only to the extent that entry barriers and allocative distortions affect aggregate productivity.

or equal to ω' . A fraction $(1 - \delta) [1 - q_t(\omega' + \Delta)]$ of the mass of firms with productivity between ω' and $(\omega' + \Delta)$ survives the exit shock and jumps downward to have productivity less than or equal to ω' . There is also an inflow of new firms into this group, which is given by the mass of entrants, a fraction $\Gamma(\omega')$ of which will feature productivity less than or equal to ω' . Endogenous exit will be driven by the mass of firms that transition downwards from the productivity cutoff, $(1 - \delta) q_t(\underline{\omega} + \Delta) M_t(\underline{\omega} + \Delta)$.

6 Quantitative Analysis

We turn now to the quantitative analysis of the role of entry barriers and idiosyncratic distortions in the formal sector in accounting for the extent of informal production and the shape of the size distribution. We first discuss the calibration strategy and then move on to quantifying the counterfactuals.

6.1 Calibration

Our calibration strategy proceeds in two steps. First, we calibrate structural parameters in the model, such as those governing the elasticity of substitution, the entry costs, and the innovation cost function, to match salient macro and firm-level properties of the United States. Then, we estimate idiosyncratic distortions from Ghanaian firm-level data and feed alternative estimates of entry barriers from the World Bank's Doing Business Indicators and from (Fattal-Jaef, 2022).

In the first step, we require the relative entry costs, f_{e1} and f_{e2} , and the relative sector-wide productivity levels A_1 and A_2 , to be consistent with an informal employment share of 8% in the U.S. and an average firm size in the U.S.' formal and Ghana's informal manufacturing sectors of 116 and 3.5 workers respectively.²⁰ The innovation cost function in the formal sector is parameterized to replicate the life-cycle growth and the employment share in the top 10 largest firms in the U.S. manufacturing sector. The growth and exit rate in the informal sector, in turn, are set so as to replicate the small share of employment accounted for by large firms in Ghana's informal firms' size distribution.

The elasticity of substitution between formal and informal goods is drawn from Bachas et al. (2020), while we adopt a standard value for the elasticity of substitution across varieties from the misallocation literature.

In the second step, we appeal to Ghana's National Industrial Census of 2003 to calibrate the idiosyncratic distortions and the entry barriers. The National Industrial Census is an alternative database covering industrial production only for registered establishments with ten workers or more, providing rich balance sheet information with which to measure idiosyncratic distortions following the approach in (Hsieh & Klenow, 2009).²¹

We follow (Fattal-Jaef, 2022)'s in measuring idiosyncratic distortions directly from the firm-level data and inferring the entry barrier so that the model's average firm size for the Ghanaian equilibrium matches the average firm size in the formal-sector data. Given the idiosyncratic distortion

²⁰ The average size in the U.S. manufacturing is conditional on firms with 10 or more workers, to make it comparable with Ghana's formal sector database

²¹ This is the database used in (Cirera et al., 2019) to characterize misallocation in Sub-Saharan Africa

profile in equation 5, we compute each firm’s $TFPR$ and $TFPQ$ and calibrate γ as the regression coefficient resulting from regressing $\log\left(\frac{TFPR_i}{TFPR_s}\right)$ on $\log\left(\frac{TFPQ_i}{TFPQ_s}\right)$, where industry averages are computed at the 4-digit level. Equipped with an estimate of the elasticity parameter of the idiosyncratic distortion profile, we infer the value of the entry tax that equates Ghana’s average firm size in equilibrium with Ghana’s observed average firm size in the manufacturing formal-sector data.

The parameter values resulting from the first and second calibration steps are reported in table 4.

Table 4: Parameter Values

Parameter	Values	Target
δ_1, δ_2	0.025, 0.0125	Exit rate large formal firms, exit rate informal firms
$\frac{\mu}{\delta_1}$	0.3	Employment share of top 10 largest firms in informal sector = 20%
λ	3	Bachas et.al. (2021)
θ	4	Hsieh and Klenow (2009)
ζ	4.5	Average size of entrant relative to incumbent U.S. anufacturing
η, ϕ	15, 0.00056	Life-Cycle growth and Employment Share top 10% largest firms U.S.
$\frac{f_{e1}}{f_{e2}}, \frac{A_1}{A_2}$	2.56, 7.15	Average firm size formal relative to informal and informal employment share -8%
γ_1	0.285	Regression Coefficient $\log(TFPR)$ vs $\log(TFPQ)$
τ^e	1.6	Average firm size in Ghana’s manufacturing sector relative to USA

Source: The average size, the employment concentration at the top 10 largest firms, and the average size ratio between 21-25 year old and 1 year old firms in the U.S.’ manufacturing sector are drawn from the Business Dynamics Statistics Database for the year 2007. To make the U.S. data comparable with Ghana’s formal Manufacturing Census, from which we infer idiosyncratic distortions and entry barriers, we compute statistics conditional on firms with ten workers or more, which is the smallest firm size covered in the Ghanaian data. The average firm size in Ghana’s informal sector is computed from the full establishment census. The $\log(TFPR)$ and $\log(TFPQ)$ are computed exactly as in (Hsieh & Klenow, 2009), estimated from Ghana’s National Industrial Census of 2003. The entry barrier τ^e is drawn directly from (Fattal-Jaef, 2022)’s estimate for Ghana.

6.2 Quantifying the Role of Entry Barriers and Idiosyncratic Distortions

Our main quantitative experiment involves starting from a stationary equilibrium calibrated to the U.S. economy and introducing the measures of distortions targeted to Ghana. Can the distorted stationary equilibrium resemble the micro-level features of the Ghanaian economy? In particular, can the extent of informal production and its implied consequences on the shape of the size distribution generate a missing middle as in the data? If so, what are the aggregate implications?

Consider first the effect of the various combinations of distortions on the shape of the size distribution, reported in figure 15. The top-left panel replicates the model’s representation of the U.S. economy’s size distribution, an outcome of the calibration. The bottom right panel illustrates the size distribution assuming only idiosyncratic distortions are in place, whereas the bottom left panel reproduces the outcome under entry barriers only. The top right plot corresponds to both distortions interacting.

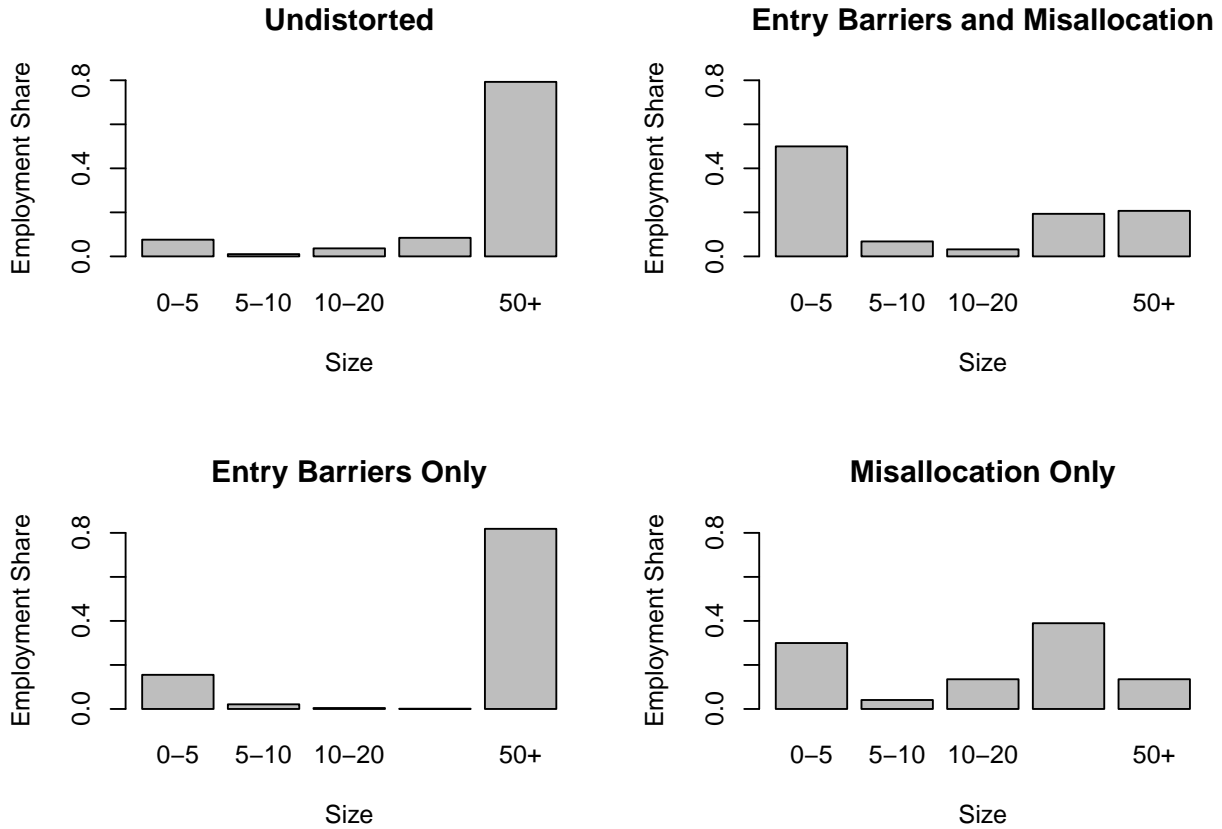
The size distribution under idiosyncratic distortions (bottom right) reallocates employment away from the top and towards the bottom but also expands some of the middle-size categories. Two forces are at play. Firstly, idiosyncratic distortions in the formal sector reduce the aggregate productivity among formal firms and increase the relative price of formal varieties, thereby reallocating

expenditure, firms, and employment into informal production. Since the size distribution of informal producers concentrates employment at the bottom, it explains the rise in the 0-5 employment share in the overall distribution. Secondly, the productivity dependence of idiosyncratic distortions misallocates employment from the largest to the smallest formal firms. Moreover, these distortions discourage formal firms from investing in innovation, which further compresses the dispersion in productivity distribution and concentrates production among middle-sized firms. On their own, then idiosyncratic distortions are unable to generate a missing middle as in the data.

The contribution of entry barriers to the size distribution is to polarize employment into smaller and larger firms, depleting the middle. On the one hand, the entry barrier constrains formal sector entry, which reduces aggregate productivity in the sector, increases its relative price, and motivates a rise in expenditure, entry, and employment in the informal sector. On the other hand, by constraining competition, the entry barrier reduces factor prices allowing the most productive firms to become even larger, rationalizing the expansion of the right tail of the distribution. While closer to reproducing the missing middle in the data, the entry barriers still exhibit excessive production at the top of the distribution.

The top right corner of figure 15 shows that the interaction between entry barriers and idiosyncratic distortions provides a closer approximation to the missing middle documented in the data. A limitation of the idiosyncratic distortions was that, due to taxing productive firms more heavily, they reallocated employment towards the middle of the distribution. Entry barriers revert this force. Moreover, the interaction between distortions worsens the decline in the formal sector's aggregate productivity, magnifies the increase in the relative price, and strengthens the rise of informal producers, all contributing to generating a missing middle in the distribution.

Figure 15: Distortions and the Size Distribution of Firms



Note: The top left panel corresponds to the equilibrium's size distribution for the undistorted economy calibrated to the U.S. The top right panel depicts the model's size distribution under our calibrated values of idiosyncratic distortions and entry barriers. The bottom left illustrates the entry barrier only economy, whereas the bottom right corresponds to the economy with idiosyncratic distortions.

Consider now the macroeconomic implications of the distortions. Table 5 reports the aggregate variables that we emphasized as the main drivers of the reallocation forces at the firm level, as well as the implications for Total Factor Productivity for the whole economy and for the formal sector only.

Table 5: Aggregate Results: Entry Barriers and Idiosyncratic Distortions

	Undistorted	Entry Barrier + Misalloc.	Entry Barrier	Misalloc.
Informal Empl. Share	0.09	0.58	0.18	0.35
Relative Price Formal to Informal	1.00	3.50	1.50	2.23
TFP	1.00	0.88	0.95	0.90
Formal GDP	1.00	0.41	0.84	0.65

Note: The informal employment shares are reported as levels in each stationary equilibrium. The relative price, *TFP*, and formal sector *GDP* are expressed relative to their corresponding values in the undistorted allocation. *TFP* corresponds to the whole economy's aggregate productivity, inclusive of both the formal and the informal sector.

The first two rows in table 5 report the effects of the distortions on the level of informality and the relative price. Informal employment rises significantly, from an 8% share in the U.S. to 18% under entry barriers, 35% under misallocation, and 58% for the combination of distortions. Relative to the informal employment share of 70% in our Ghanaian data, the model with both distortions can account for about three-quarters of the observed value. The table also shows the change in the relative price of formal to informal goods as the general equilibrium force driving the reallocation of expenditure, firm entry, and employment into informal production. The relative price increases more than threefold relative to its value in the undistorted economy.

In terms of the productivity losses, table 5 shows that these reach up to 12% in the case of entry barriers and idiosyncratic distortions. The last row of the table allows for a decomposition of the aggregate losses. Formal output plummets by up to 60% due to the multiple effects created by the distortions, namely, a misallocation of resources among surviving incumbents, a reduction in formal sector entry, and subdued innovation expenses. However, the informal sector, in so far as it is captured in official statistics, partially compensates for the losses in the formal sector, by increasing the number of varieties and thereby informal aggregate output. The aggregate productivity still falls due to the lower aggregate productivity in the formal sector and due to the reallocation of resources towards the informal sector, which is less productive than the formal one.

7 Conclusion

This paper investigates the size distribution of manufacturing plants in sub-Saharan Africa, where the distortions are all the more ubiquitous. It uses firsthand establishment censuses from Burkina Faso, Cameroon, Ghana, and Rwanda to identify four stylized facts: (i) there exists a drop in the employment share accounted for by the middle range of the size distribution, providing some support to the missing middle hypothesis; (ii) informal firms are necessary for the missing middle to appear; (iii) the size distribution of formal firms looks somehow more like the size distribution in developed economies; and (iv) the missing middle does not arise from entry but instead reflects firm growth patterns. To the best of our knowledge, our contribution is the first that embraces detailed and comprehensive censuses of establishments in the non-agricultural sectors of sub-Saharan countries in a comparative fashion.

The controversy about the existence of a missing middle appears to be definitional. A binary existence statement necessarily rests on identifying assumptions that arguably obscure the relevance of the matter. In that sense, we argue that the missing middle is an elusive concept. In this paper, we stepped back from the controversy and established a number of stylized facts with respect to patterns in the size distribution.

Our analysis underlines the important consequences of informality and questions the common perception that the missing middle arises from the size distribution of entrants. Beyond our exercise, we believe that methodically studying the size distribution in order to identify empirical regularities is useful and policy-relevant. The possibilities are currently rather limited, as one comes up against the issue of data availability. We can only hope that more establishment censuses will be conducted in the future – panels, in particular, would be greatly informative.

The policy-relevant question here is whether a missing middle is a cause for concern. In some

ways, it is: a firm size distribution exhibiting a missing middle means that there is a scope for shifting workers from small enterprises into medium-sized and larger ones. Given that larger firms tend to be more productive thanks to economies of scale, this shift would imply a productivity gain. A missing middle is therefore a symptom of unrealized potential for reallocation of labor across firms. It is therefore a useful metric to assess the scope for labor reallocation, which can be applied in relatively data-scarce settings.

Nevertheless, the missing middle as a term and concept can be a bit misleading. A distribution with a missing middle should be interpreted holistically: it is not just about a lack of medium-sized firms, but also about small firms not growing as well as larger firms failing to absorb the workforce employed in inefficiently small firms. A multitude of factors can contribute to this: lacking capabilities of entrepreneurs limiting the growth of micro-enterprises as well as distortions around formalization and labor market that medium-sized or large firms face to grow.

In terms of policy implications, a critical problem for African economies' productivity and formal employment is a lack of a large number of formal firms, and also a very large share of small informal firms that do not grow. Thus, policies that reduce the costs of registration by making it simple to become formal, and especially strengthening enforcement ([Ulyssea, 2020](#)) are important to reduce informality. However, most informal firms do not grow much, which can be the result of the need to stay small to avoid enforcement, but also lacking firm capabilities and subsistence entrepreneurship due to the lack of formal employment opportunities. Thus, policies that are size-dependent, if effective, do not necessarily increase the number of medium and large firms. Finally, despite the fact that formal firms in Africa resemble more the size distribution of formal firms in advanced economies, their growth is much less over the life-cycle due to existing distortions ([Cirera et al., 2019](#)), which implies that measures to reduce informality, need to go hand-in-hand with measures to reduce existing distortions to firm growth more generally.

References

- Alvarez, J., & Ruane, C. (2019). *Informality and aggregate productivity: The case of Mexico* (Working Paper No. 2019/257). IMF.
- Antunes, A. R., & Cavalcanti, T. V. d. V. (2007, January). Start up costs, limited enforcement, and the hidden economy. *European Economic Review*, 51(1), 203-224. Retrieved from <https://ideas.repec.org/a/eee/eecrev/v51y2007i1p203-224.html>
- Atkeson, A., & Burstein, A. T. (2010). Innovation, firm dynamics, and international trade. *Journal of Political Economy*, 118(3), 433-484. Retrieved from <https://doi.org/10.1086/653690> doi: 10.1086/653690
- Atkin, D., Faber, B., & Gonzalez-Navarro, M. (2018). Retail globalization and household welfare: Evidence from Mexico. *Journal of Political Economy*, 126(1), 1-73. Retrieved from <https://doi.org/10.1086/695476> doi: 10.1086/695476
- Axtell, R. L. (2001). Zipf Distribution of US Firm Sizes. *Science*, 293(5536), 1818-1820.
- Bachas, P., Gadenne, L., & Jensen, A. (2020, June). *Informality, consumption taxes and redistribution* (Working Paper No. 27429). National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w27429> doi: 10.3386/w27429
- Bento, P., & Restuccia, D. (2014). *Misallocation, establishment size, and productivity* (Tech. Rep.).
- Cirera, X., Fattal-Jaef, R., & Maemir, H. (2019, 08). Taxing the Good? Distortions, Misallocation, and Productivity in Sub-Saharan Africa. *World Bank Economic Review*, 34(1), 75-100.
- D'Erasmus, P. N., & Moscoso Boedo, H. J. (2012). Financial structure, informality and development. *Journal of Monetary Economics*, 59(3), 286-302. Retrieved from <https://ideas.repec.org/a/eee/moneco/v59y2012i3p286-302.html> doi: 10.1016/j.jmoneco.2012.03
- Di Giovanni, J., Levchenko, A. A., & Ranciere, R. (2011). Power Laws in Firm Size and Openness to Trade: Measurement and Implications. *Journal of International Economics*, 85(1), 42-52.
- Fattal-Jaef, R. N. (2022, April). Entry barriers, idiosyncratic distortions, and the firm size distribution. *American Economic Journal: Macroeconomics*, 14(2), 416-68. Retrieved from <https://www.aeaweb.org/articles?id=10.1257/mac.20200234> doi: 10.1257/mac.20200234
- Hopenhayn, H. A. (2014, August). *On the measure of distortions* (Working Paper No. 20404). National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w20404> doi: 10.3386/w20404
- Hsieh, C.-T., & Klenow, P. J. (2009). Misallocation and Manufacturing TFP in China and India. *Quarterly Journal of Economics*, 124(4), 1403-1448.
- Hsieh, C.-T., & Olken, B. A. (2014). The Missing "Missing Middle". *Journal of Economic Perspectives*, 28(3), 89-108.
- Lewis, W. A. (1954). Economic Development with Unlimited Supplies of Labour. *The Manchester School*, 22(2), 139-191.
- Lopez-Martin, B. (2019, December). Informal Sector Misallocation. *Macroeconomic Dynamics*, 23(8), 3065-3098. Retrieved from https://ideas.repec.org/a/cup/macdyn/v23y2019i8p3065-3098_2.html
- Luttmer, E. G. (2010). Models of growth and firm heterogeneity. *Annual Review of Economics*, 2(1), 547-576. Retrieved from <https://doi.org/10.1146/annurev.economics.102308.124410> doi: 10.1146/annurev.economics.102308.124410

10.1146/annurev.economics.102308.124410

- Luttmer, E. G. J. (2007, 08). Selection, growth, and the size distribution of firms. *The Quarterly Journal of Economics*, 122(3), 1103-1144.
- Ordoñez, J. C. L. (2014, November). *The informal sector in contemporary models of the aggregate economy* (Working Papers No. 2014-24). Banco de México. Retrieved from <https://ideas.repec.org/p/bdm/wpaper/2014-24.html>
- Rauch, J. (1991). Modelling the informal sector formally. *Journal of Development Economics*, 35(1), 33-47. Retrieved from <https://EconPapers.repec.org/RePEc:eee:deveco:v:35:y:1991:i:1:p:33-47>
- Tybout, J. R. (2000). Manufacturing Firms in Developing Countries: How Well Do They Do, and Why? *Journal of Economic Literature*, 38, 11-44.
- Tybout, J. R. (2014). The Missing Middle: Correspondence. *Journal of Economic Perspectives*, 28(4), 235-236.
- Ulyssea, G. (2018, August). Firms, Informality, and Development: Theory and Evidence from Brazil. *American Economic Review*, 108(8), 2015-2047. Retrieved from <https://ideas.repec.org/a/aea/aecrev/v108y2018i8p2015-47.html>
- Ulyssea, G. (2020). Informality: Causes and consequences for development. *Annual Review of Economics*, 12(1), 525-546. Retrieved from <https://doi.org/10.1146/annurev-economics-082119-121914> doi: 10.1146/annurev-economics-082119-121914

Appendix

Table A.1: Burkina Faso (2015)

Indicator	Sector coverage	Total	Establishment formality indicator							
			Keep regular account		Use SYSCOA accounting standard		Keep regular account or use SYSCOA accounting standard		Census (local) definition	
			#	%	#	%	#	%	#	%
Establishment	All	99,261	12.75	9.56	9,492	22.31	22,147	9,028	9.10	
	Manufacturing	16,705	9.03	3.89	650	12.92	2,159	620	3.71	
Employment	All	255,908	12.09	31.81	81,412	43.90	112,339	74,067	28.94	
	Manufacturing	53,116	10.05	27.48	14,594	37.53	19,932	13,420	25.27	

Source: Establishment censuses obtained from the statistical agencies of the select countries; see section 3. According to the census data from Burkina Faso, an establishment is considered formal if it has registration number (with mobile trade and mobile credit, tax identification number) and uses SYSCOHADA as a type of accounting.

Table A.2: Cameroon (2008)

Indicator	Sector coverage	Total	Establishment formality indicator							
			Registration at tax authority		Keep formal account		Registration at tax authority and keeps formal account		Census (local) definition	
			#	%	#	%	#	%	#	%
Establishment	All	88,144	7.78	13.24	11,673	7.47	6,584	11,948	13.56	
	Manufacturing	10,922	5.26	8.02	876	5.16	564	887	8.12	
Employment	All	429,758	46.11	53.87	231,512	45.72	196,477	233,186	54.26	
	Manufacturing	82,502	58.65	61.39	50,647	58.54	48,298	50,733	61.49	

Source: Establishment censuses obtained from the statistical agencies of the select countries; see section 3. According to the census data from Cameroon, formal establishments are the ones that prepare a Statistical and Fiscal Declaration (DSF) or those that do not prepare DSF but which keep an operating account and a partial account of balance sheets.

Table A.3: Ghana (2013)

Indicator	Sector coverage	Total	Establishment formality indicator											
			Registration at business register		Registration at tax authority		Registration at business register or tax authority		Keep formal account		Registration at business register/tax authority or keep formal account		Census (local) definition	
			#	%	#	%	#	%	#	%	#	%	#	%
Establishment	All	638,234	114,285	17.91	208,191	32.62	247,771	38.82	88,141	13.81	269,083	42.16	60,314	9.45
	Manufacturing	99,437	8,220	8.27	27,964	28.12	31,082	31.26	4,267	4.29	31,947	32.13	2,936	2.95
Employment	All	3,383,206	1,683,640	49.76	1,709,699	50.53	2,083,013	61.57	1,700,999	50.28	2,360,573	69.77	1,355,326	40.06
	Manufacturing	437,316	166,783	38.14	218,970	50.07	241,252	55.17	145,925	33.37	246,580	56.38	137,246	31.38

Source: Establishment censuses obtained from the statistical agencies of the select countries; see section 3. According to the census data from Ghana, formal establishments are defined as those that are registered with the Registrar General's Department (RGD) and keep formal accounts. In Rwanda, an enterprise is considered to operate in the formal sector if it is registered at Rwanda Revenue Authority (RRA) and maintains regular operational accounts.

Table A.4: Rwanda (2013)

Indicator	Sector coverage	Total	Establishment formality indicator											
			Registration at business register		Registration at tax authority		Registration at business register or tax authority		Keep formal account		Registration at business register/tax authority or keep formal account		Census (local) definition	
			#	%	#	%	#	%	#	%	#	%	#	%
Establishment	All	154,236	124,423	80.67	36,538	23.69	125,443	81.33	11,889	126,041	81.72	14,487	10,051	6.52
	Manufacturing	10,742	8,758	81.53	2,306	21.47	8,844	82.33	971	8,881	82.68	707	405	3.77
Employment	All	493,302	452,631	91.76	316,605	64.18	455,233	92.28	270,976	456,078	92.45	264,649	204,679	41.49
	Manufacturing	39,708	36,441	91.77	23,473	59.11	36,724	92.49	19,142	36,785	92.64	18,691	15,064	37.94

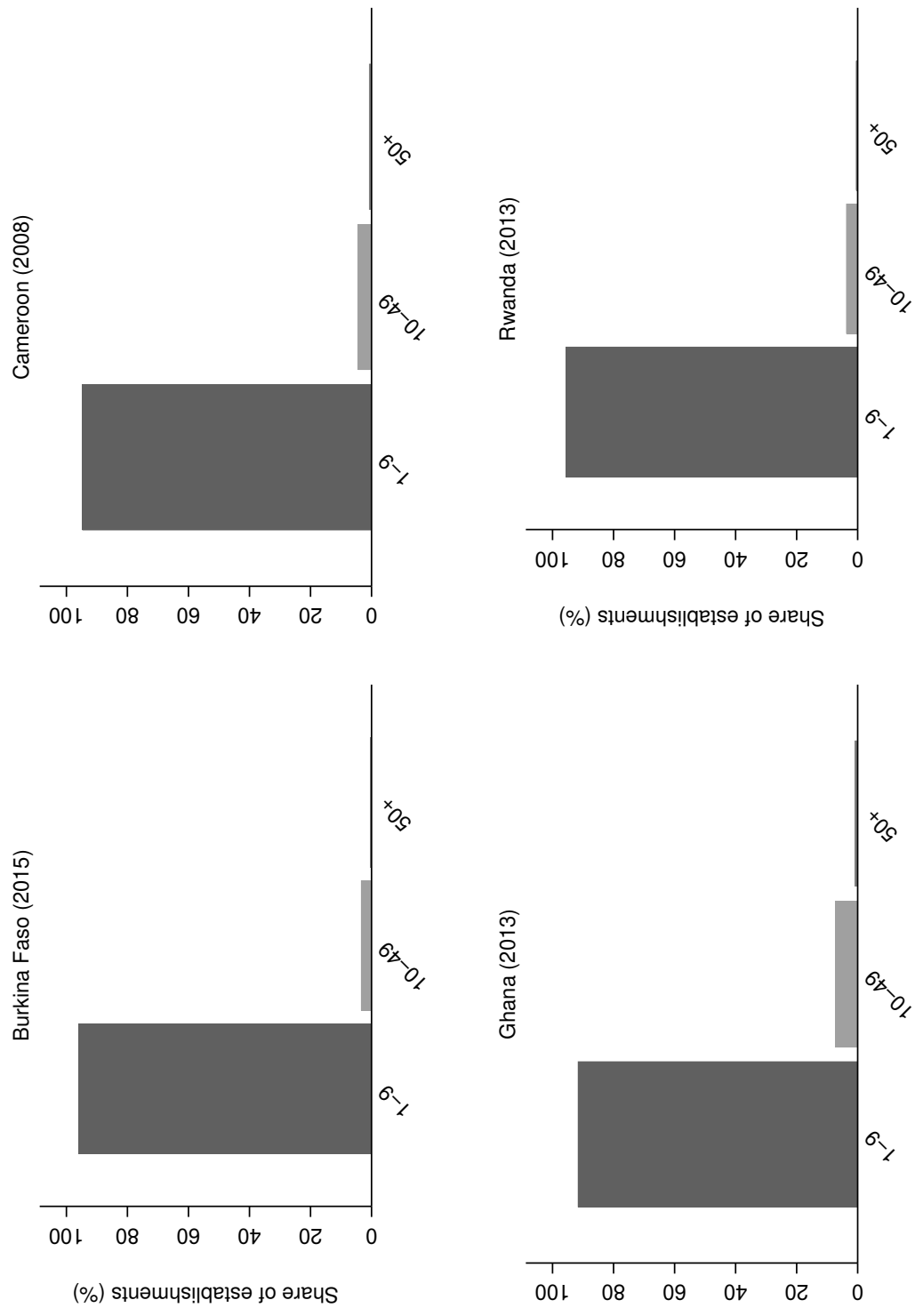
Source: Establishment censuses obtained from the statistical agencies of the select countries; see section 3. According to the census data from Rwanda, an enterprise is considered to operate in the formal sector if it is registered at Rwanda Revenue Authority (RRA) and maintains regular operational accounts.

Table A.5: Difference Between Actual vs Predicted Employment Employment Shares in the Formal Manufacturing (based on Pareto distribution)

	κ	Size class – Number of workers				
		1 – 4	5 – 9	10 – 19	20 – 99	100+
Panel a - Narrow size classes						
Burkina Faso (2015)	1.079	-0.092	-0.002	0.009	0.117	-0.005
Cameroon (2008)	1.051	-0.064	-0.014	0.015	0.080	0.001
Ghana (2013)	1.178	-0.010	0.068	0.006	-0.018	0.002
Rwanda (2013)	1.155	-0.050	0.035	0.020	0.049	-0.010
Panel b - Broad size classes						
	κ	1 – 9	10 – 49	50+		
Burkina Faso (2015)	1.061	-0.076	0.097	-0.014		
Cameroon (2008)	1.030	-0.045	0.056	-0.006		
Ghana (2013)	1.186	0.017	-0.012	0.010		
Rwanda (2013)	1.150	-0.035	0.058	-0.011		

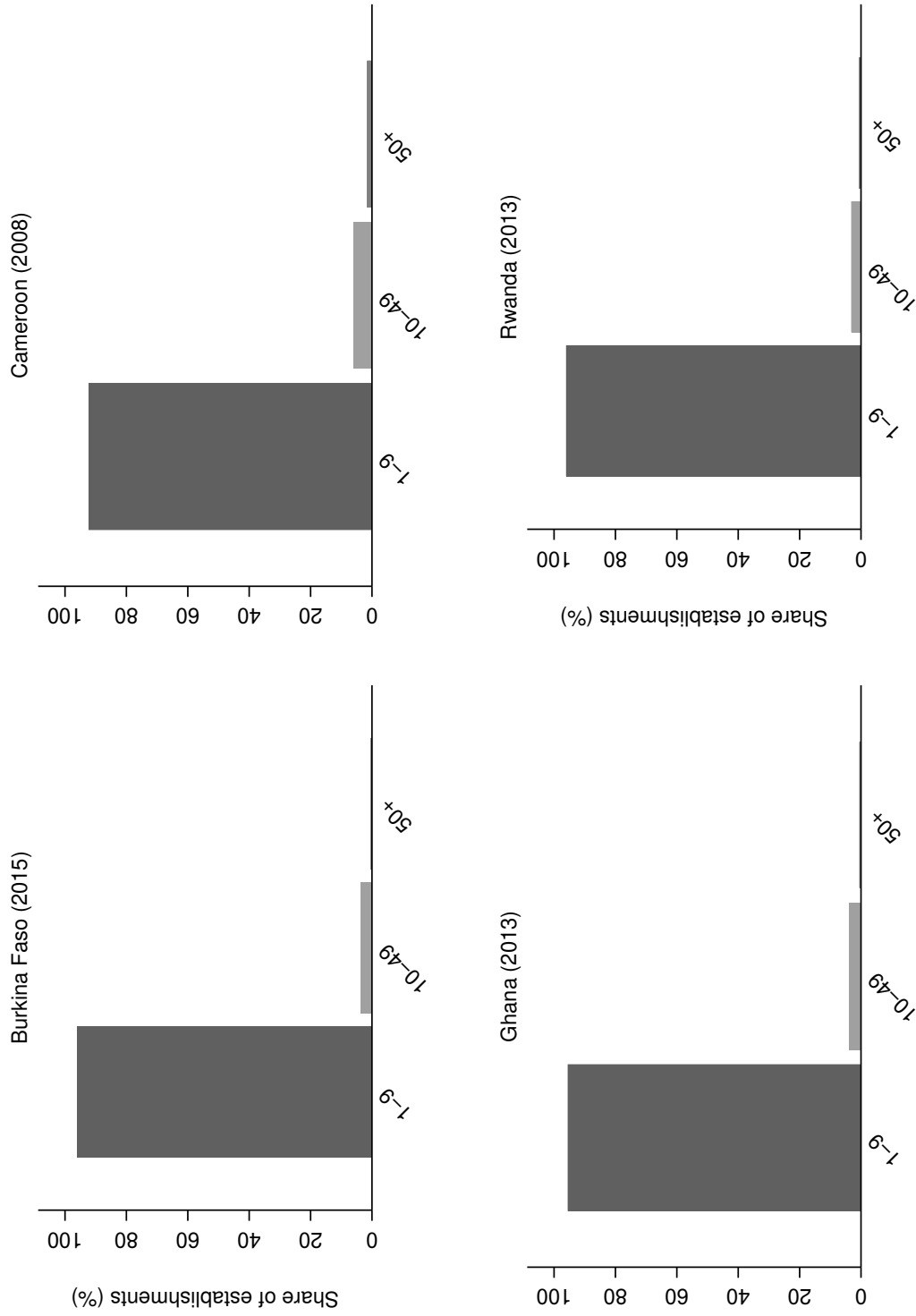
Source: Establishment censuses obtained from the statistical agencies of the select countries; see section 3. *Note:* The Pareto shape parameter κ is estimated based on minimizing Euclidean distance between observed and theoretical shares.

Figure A.1: Establishment distribution (broad size classification): Economy-wide



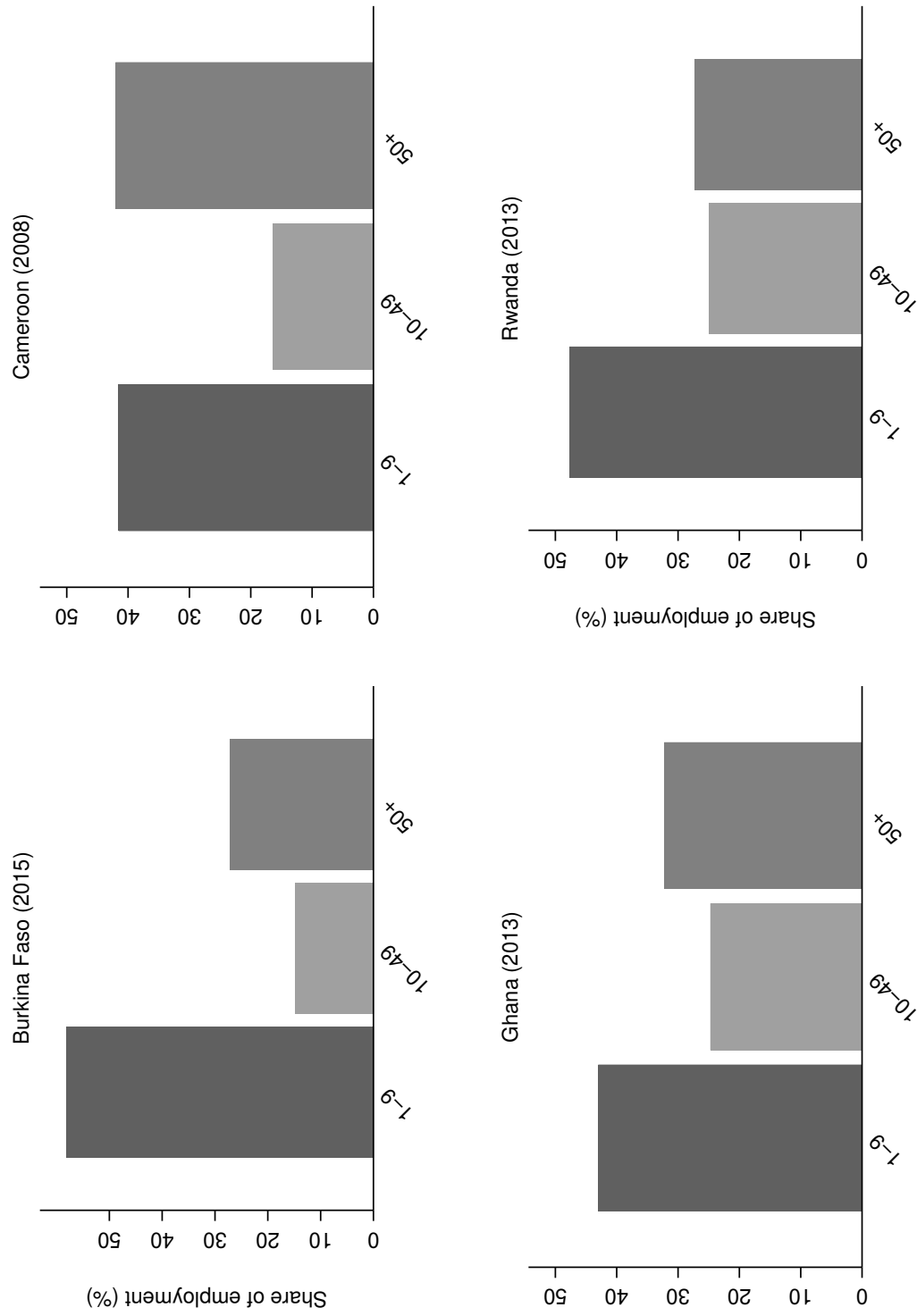
Source: Establishment censuses obtained from the statistical agencies of the select countries; see section 3. For the U.S., the data comes from the 2019 Business Dynamics Statistics (BDS) dataset. Note: The figure is constructed based on data for establishments in all sectors irrespective of their state or foreign ownership status. Establishments with missing employment data are excluded.

Figure A.2: Establishment distribution (broad size classification): Manufacturing



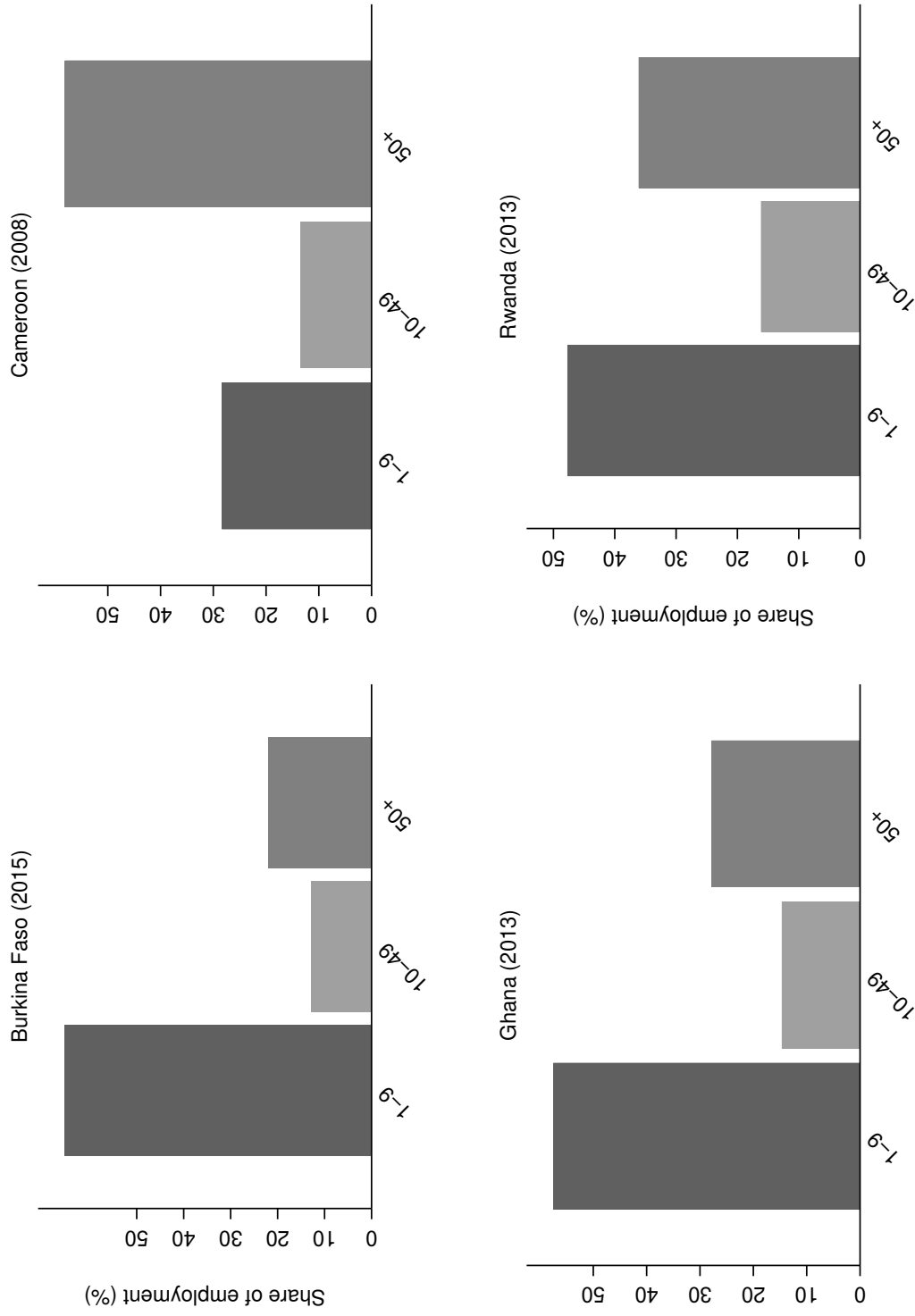
Source: Establishment censuses obtained from the statistical agencies of the select countries; see section 3. For the U.S., the data comes from the 2019 Business Dynamics Statistics (BDS) dataset. Note: The figure is constructed based on data for establishments in the manufacturing sector irrespective of their state or foreign ownership status. Establishments with missing employment data are excluded.

Figure A.3: Employment distribution (broad size classification): Economy-wide



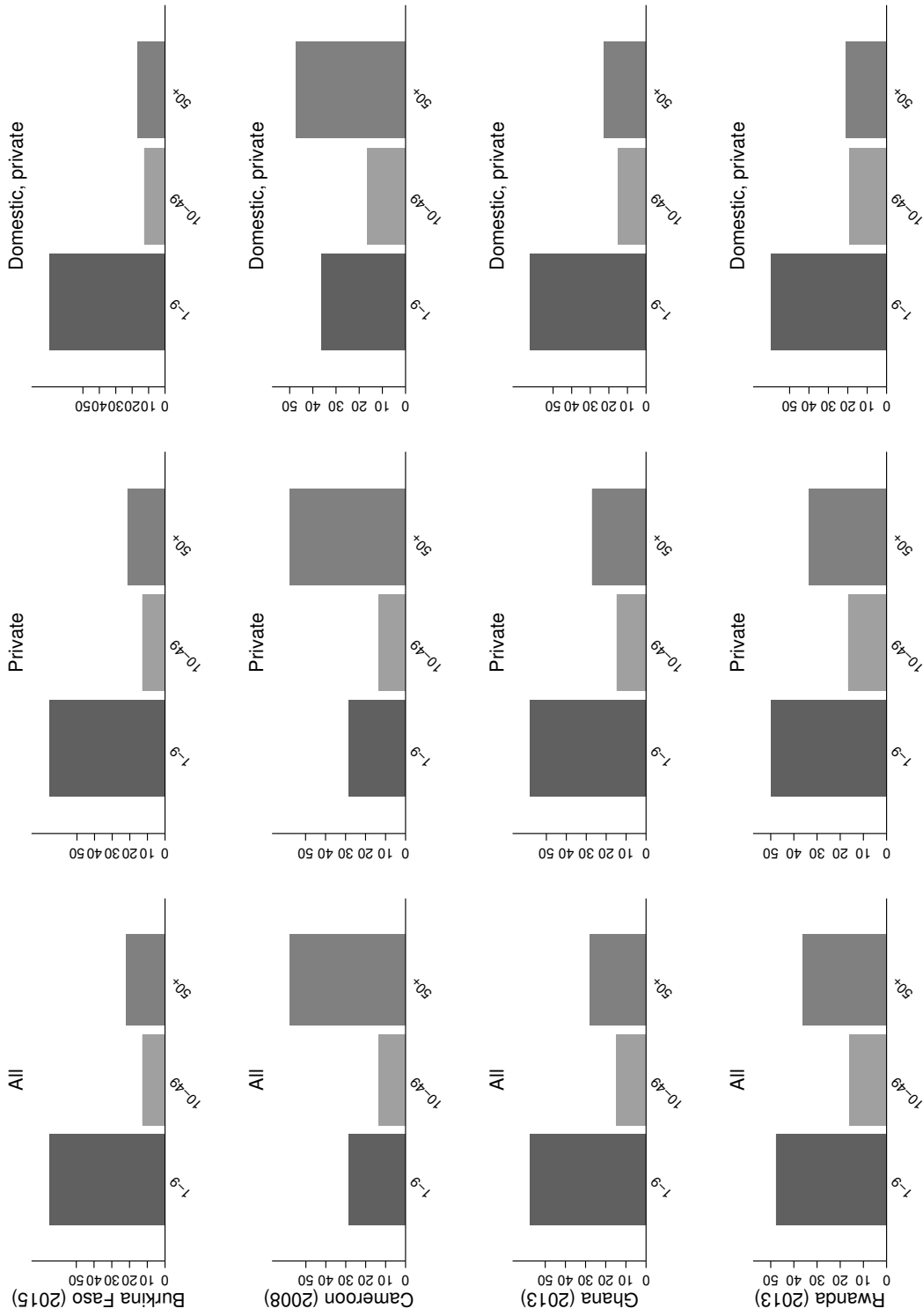
Source: Establishment censuses obtained from the statistical agencies of the select countries; see section 3. For the U.S., the data comes from the 2019 Business Dynamics Statistics (BDS) dataset. Note: The figure is constructed based on data for establishments in all sectors irrespective of their state or foreign ownership status. Establishments with missing employment data are excluded.

Figure A.4: Employment distribution (broad size classification): Manufacturing



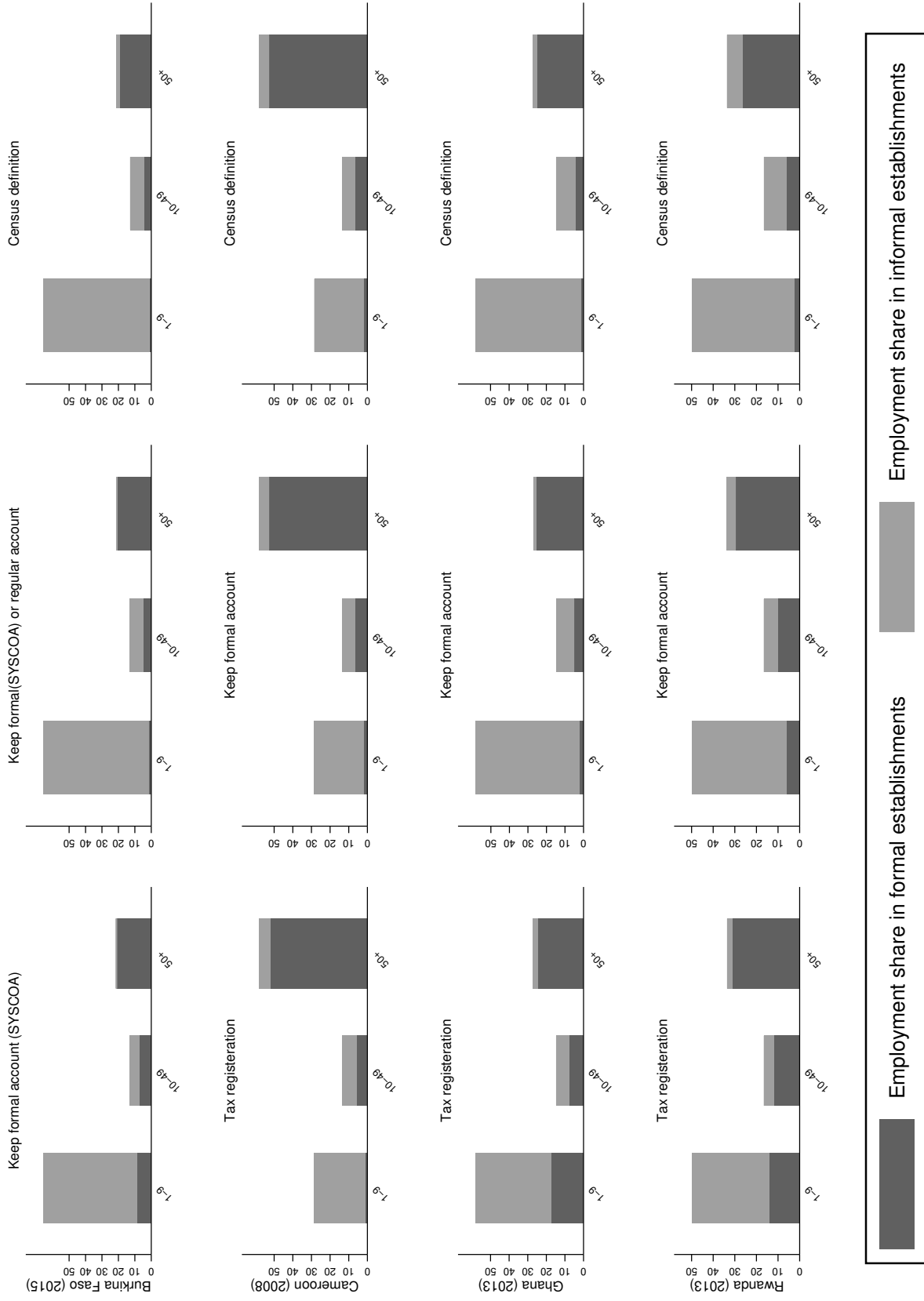
Source: Establishment censuses obtained from the statistical agencies of the select countries; see section 3. For the U.S., the data comes from the 2019 Business Dynamics Statistics (BDS) dataset. Note: The figure is constructed based on data for establishments in the manufacturing sector irrespective of their state or foreign ownership status. Establishments with missing employment data are excluded.

Figure A.5: Employment distribution by state and foreign ownership (broad size classification): Manufacturing



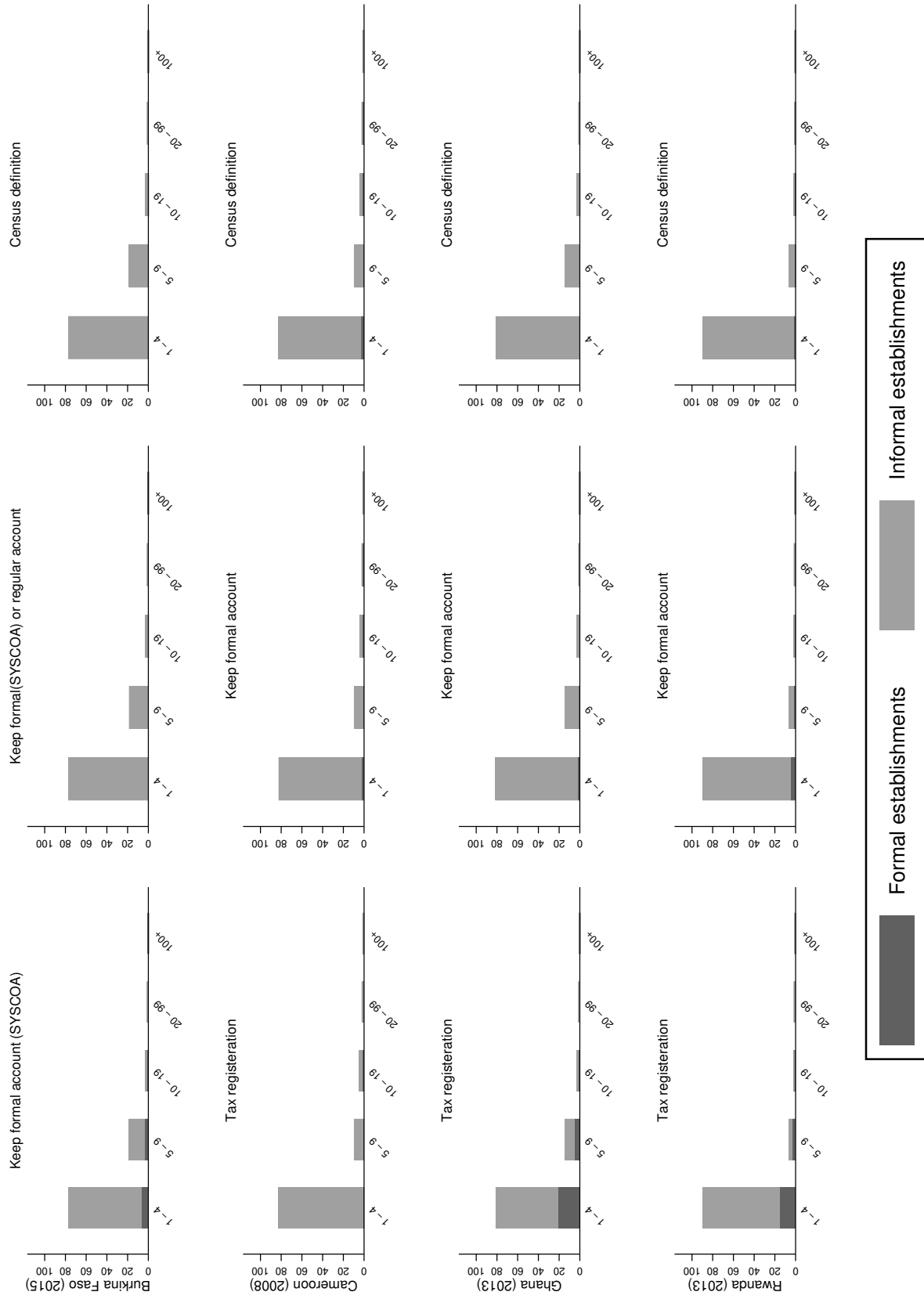
Source: Establishment censuses obtained from the statistical agencies of the select countries; see section 3. Note: The figure is constructed based on data for establishments in the manufacturing sector. Establishments with missing employment data are excluded.

Figure A.6: Formal and informal employment distribution (broad size classification): Manufacturing



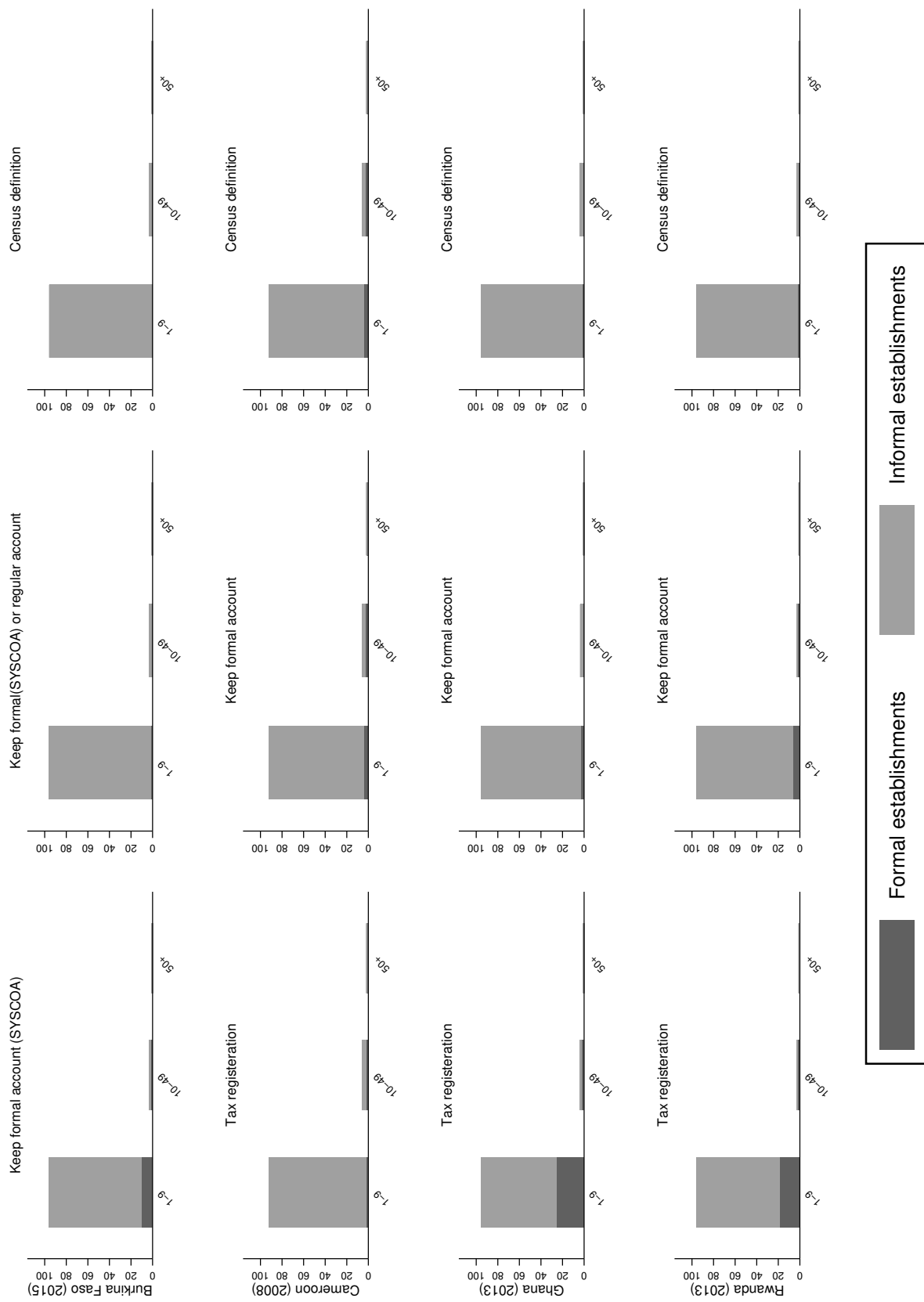
Source: Establishment censuses obtained from the statistical agencies of the select countries; see section 3. Note: The figure is constructed based on data for establishments in the manufacturing sector. Establishments with missing employment data and state-owned are excluded. The statistical offices of the sample countries have slightly different yet comparable definitions of formality. In Burkina Faso, an establishment is considered formal if it has registration number (with mobile trade and mobile credit, tax identification number) and uses SYSCOA as an accounting standard. In Cameroon, formal establishments are the ones that prepare a Statistical and Fiscal Declaration (DSF) or those that do not prepare DSF but which keep an operating account and a partial account of balance sheets. In Ghana, formal establishments are defined as those that are registered with the Registrar General's Department (RGD) and keep formal accounts. In Rwanda, an establishment is considered to operate in the formal sector if it is registered at Rwanda Revenue Authority (RRA) and maintains regular operational accounts.

Figure A.7: Formal and informal establishment distribution: Manufacturing



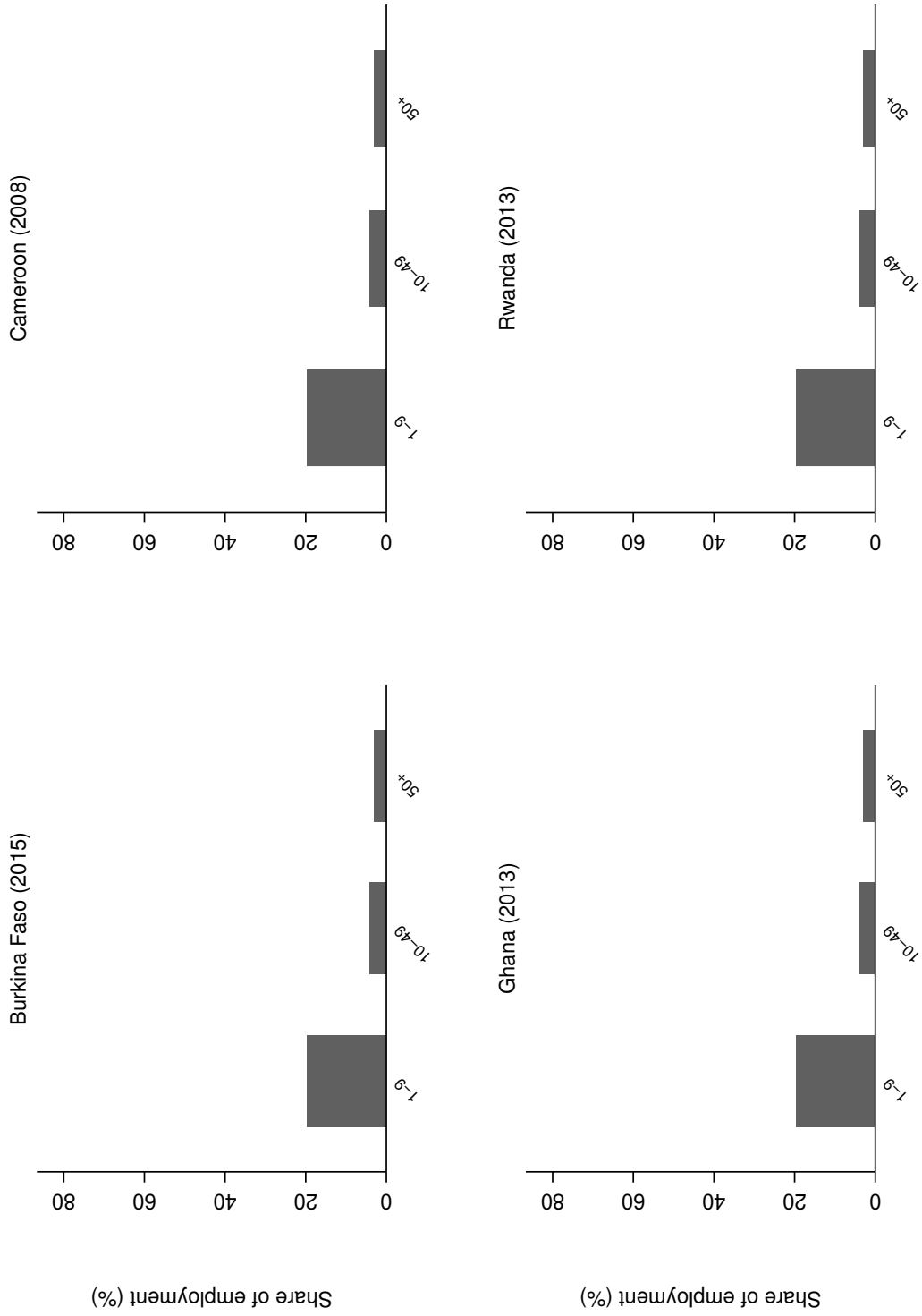
Source: Establishment censuses obtained from the statistical agencies of the select countries; see section 3. *Note:* The figure is constructed based on data for establishments in the manufacturing sector. Establishments with missing employment data and state-owned are excluded. The statistical offices of the sample countries have slightly different yet comparable definitions of formality. In Burkina Faso, an establishment is considered formal if it has registration number (with mobile trade and mobile credit, tax identification number) and uses SYSCOA as an accounting standard. In Cameroon, formal establishments are the ones that prepare a Statistical and Fiscal Declaration (DSF) or those that do not prepare DSF but which keep an operating account and a partial account of balance sheets. In Ghana, formal establishments are defined as those that are registered with the Registrar General's Department (RGD) and keep formal accounts. In Rwanda, an establishment is considered to operate in the formal sector if it is registered at Rwanda Revenue Authority (RRA) and maintains regular operational accounts.

Figure A.8: Formal and informal establishment distribution (broad size classification): Manufacturing



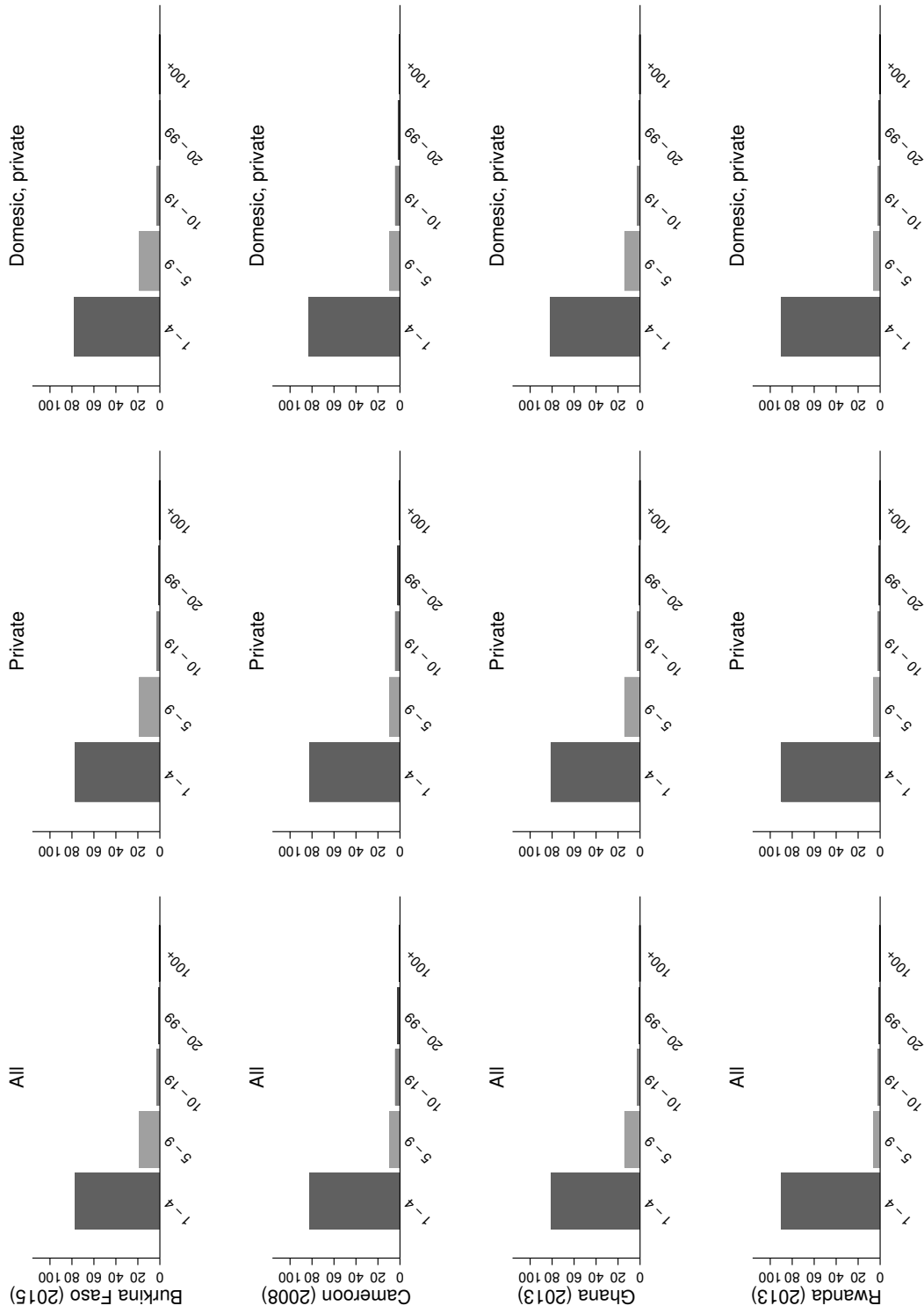
Source: Establishment censuses obtained from the statistical agencies of the select countries; see section 3. *Note:* The figure is constructed based on data for establishments in the manufacturing sector. Establishments with missing employment data and state-owned are excluded. The statistical offices of the sample countries have slightly different yet comparable definitions of formality. In Burkina Faso, an establishment is considered formal if it has registration number (with mobile trade and mobile credit, tax identification number) and uses SYSCOA as an accounting standard. In Cameroon, formal establishments are the ones that prepare a Statistical and Fiscal Declaration (DSF) or those that do not prepare DSF but which keep an operating account and a partial account of balance sheets. In Ghana, formal establishments are defined as those that are registered with the Registrar General's Department (RGD) and keep formal accounts. In Rwanda, an establishment is considered to operate in the formal sector if it is registered at Rwanda Revenue Authority (RRA) and maintains regular operational accounts.

Figure A.9: Employment distribution among entrants (broad size classification): Manufacturing



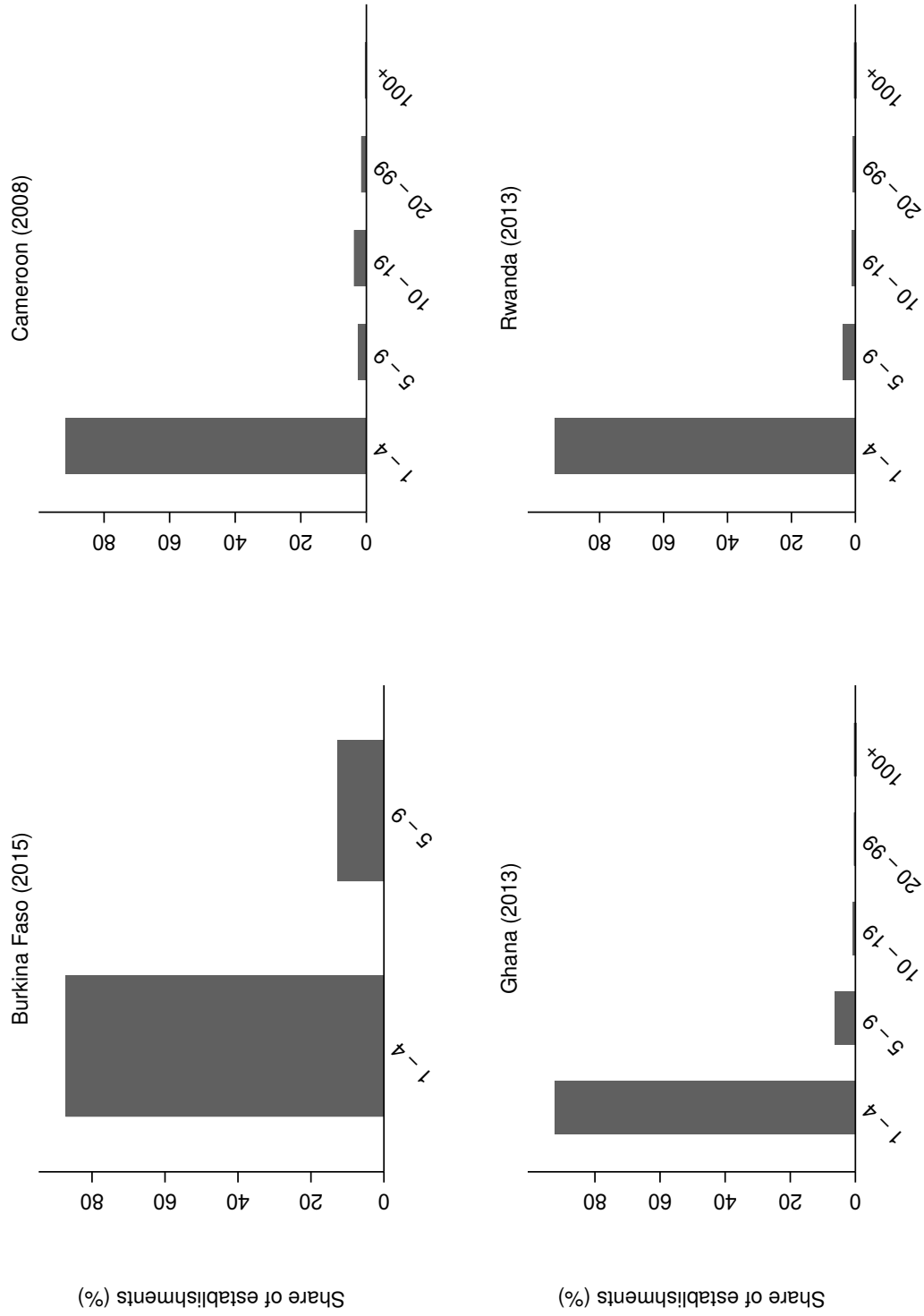
Source: Establishment censuses obtained from the statistical agencies of the select countries; see section 3. Note: The figure is constructed based on data for establishments in the manufacturing sector. Establishments with missing employment or age data and state-owned are excluded.

Figure A.10: Establishment distribution by state and foreign ownership: Manufacturing



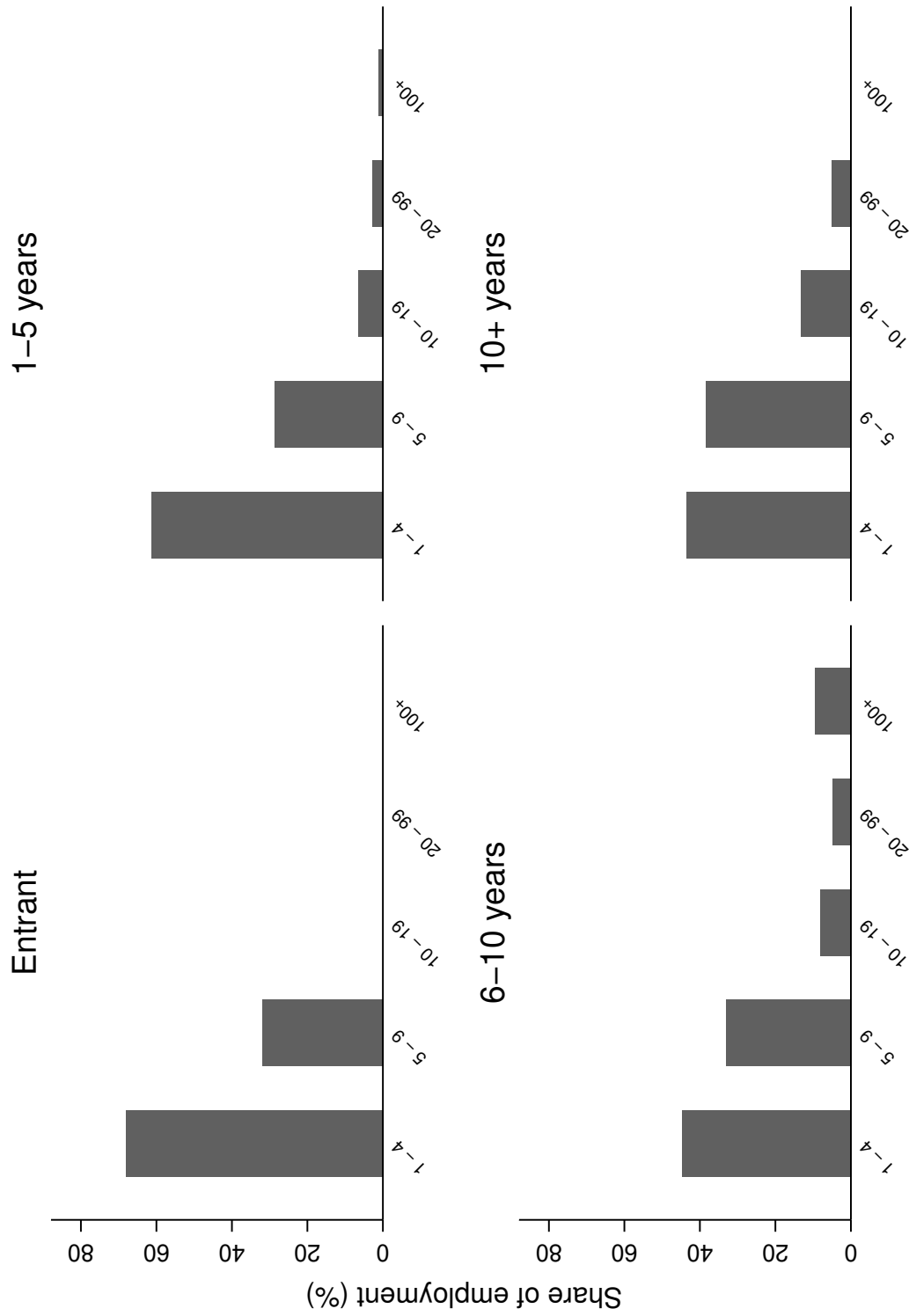
Source: Establishment censuses obtained from the statistical agencies of the select countries; see section 3. Note: The figure is constructed based on data for establishments in the manufacturing sector. Establishments with missing employment data are excluded.

Figure A.11: size difference by age group manufacturing



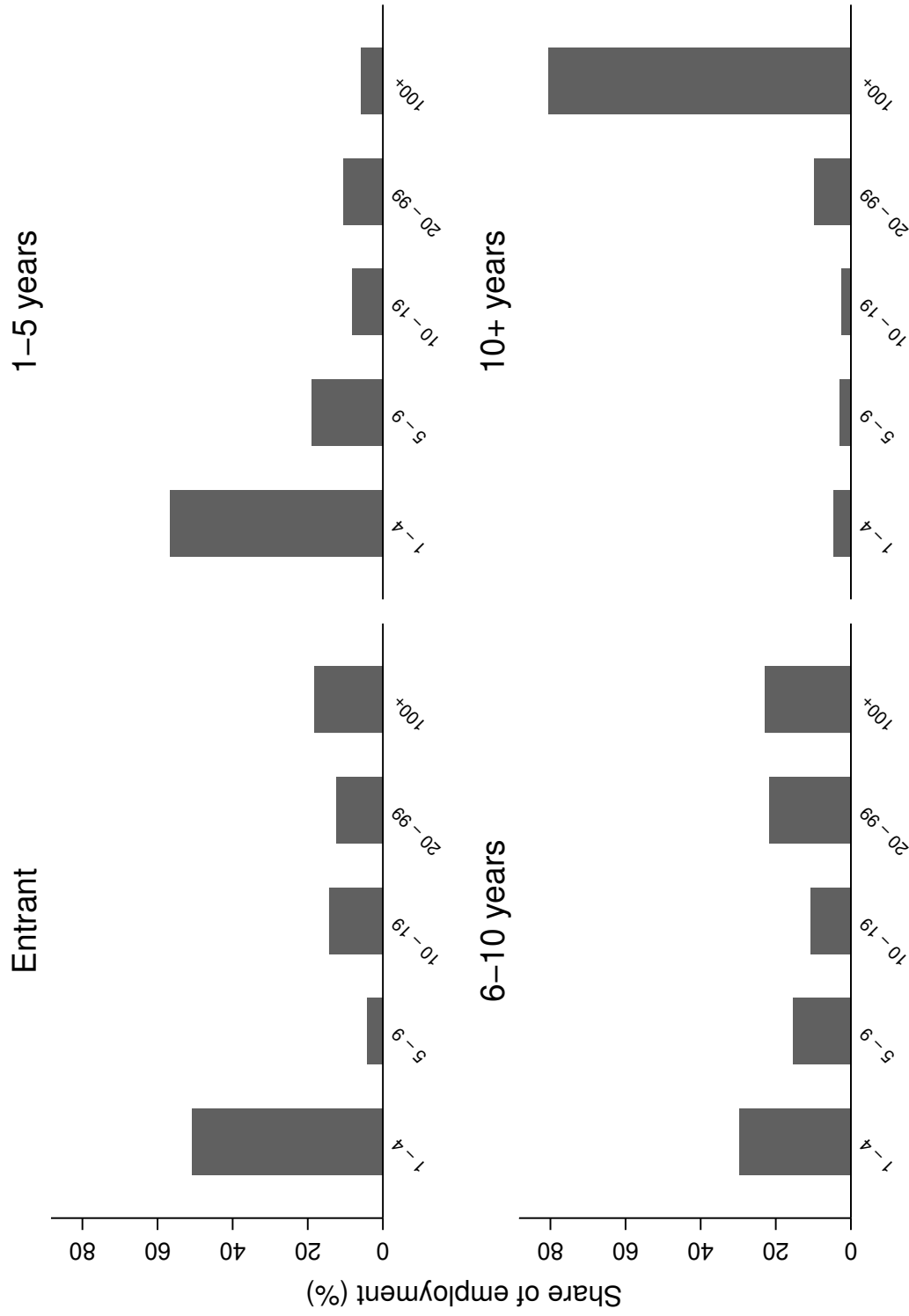
Source: Establishment censuses obtained from the statistical agencies of the select countries; see section 3. Note: The figure is constructed based on data for establishments in the manufacturing sector. Establishments with missing employment or age data and state-owned are excluded.

Figure A.12: Formal and informal establishment distribution by age cohort in manufacturing: Burkina Faso (2015)



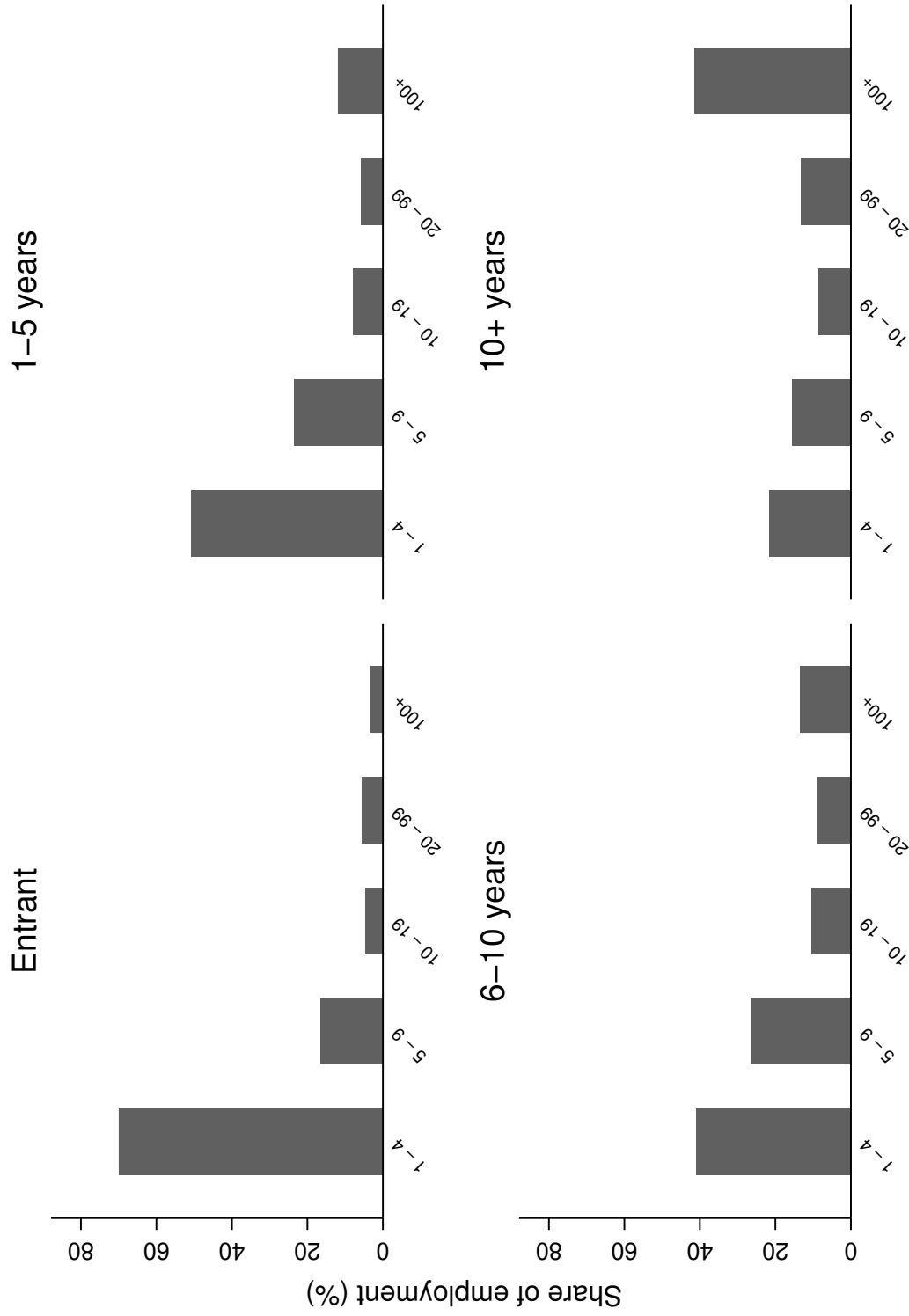
Source: Establishment censuses obtained from the statistical agencies of the selected countries; see section 3. Note: The reference is the total employment for each age cohort and formality status. The figure is constructed based on data for establishments in the manufacturing sector. Establishments with missing employment or age data and state-owned are excluded.

Figure A.13: Formal and informal establishment distribution by age cohort in manufacturing: Cameroon (2008)



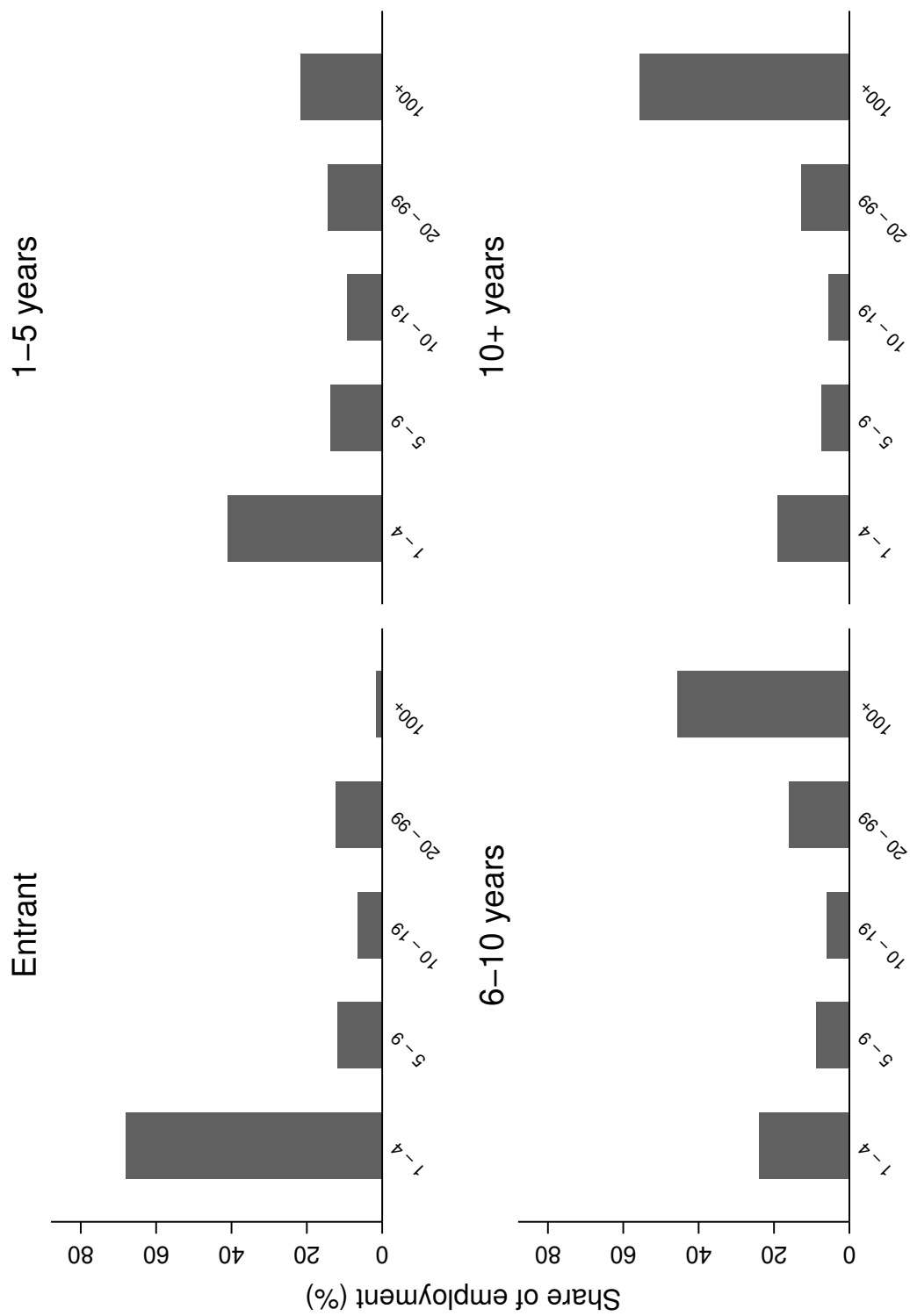
Source: Establishment censuses obtained from the statistical agencies of the selected countries; see section 3. Note: The reference is the total employment for each age cohort and formality status. The figure is constructed based on data for establishments in the manufacturing sector. Establishments with missing employment or age data and state-owned are excluded.

Figure A.14: Formal and informal establishment distribution by age cohort in manufacturing: Ghana (2013)



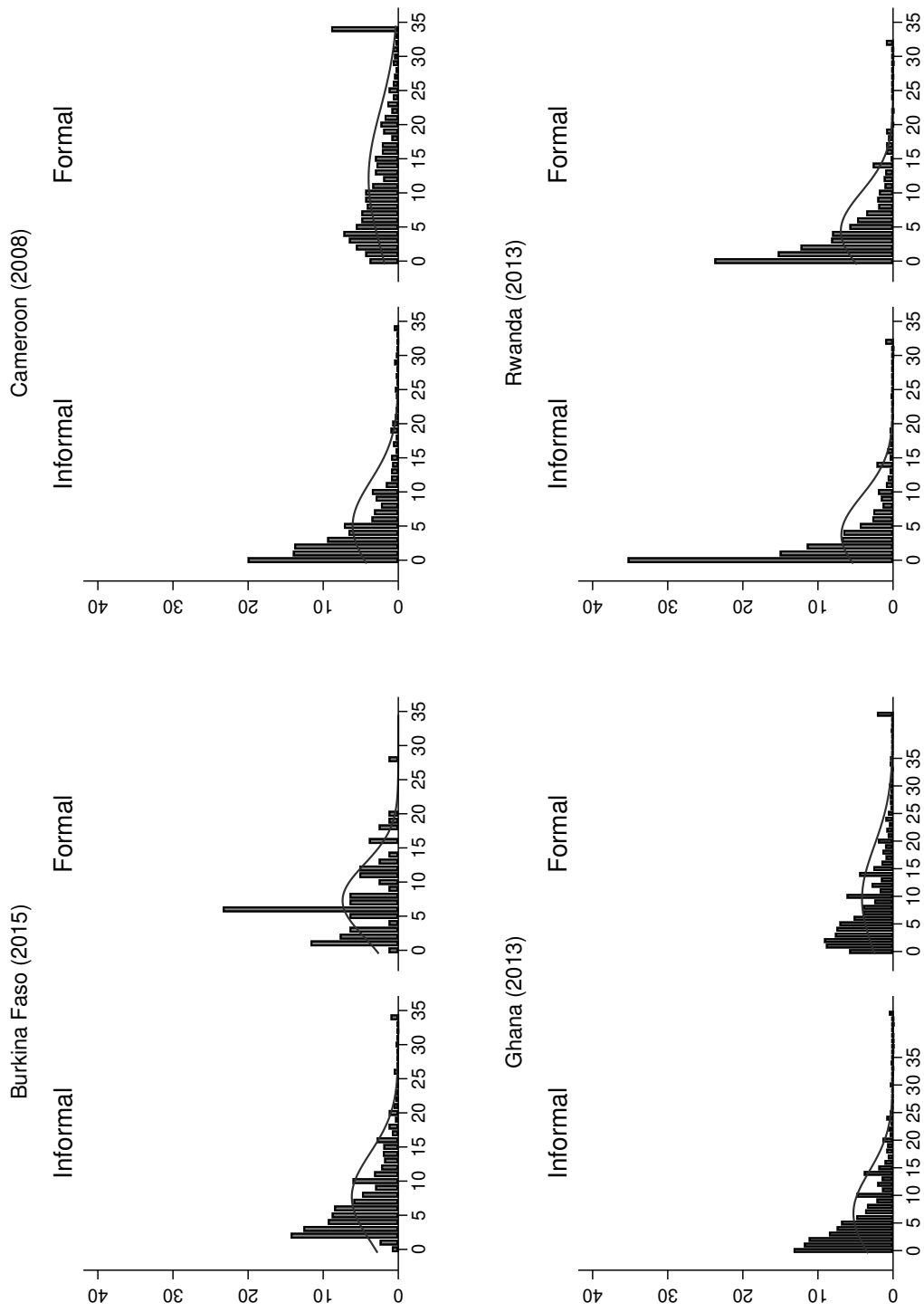
Source: Establishment censuses obtained from the statistical agencies of the selected countries; see section 3. Note: The reference is the total employment for each age cohort and formality status. The figure is constructed based on data for establishments in the manufacturing sector. Establishments with missing employment or age data and state-owned are excluded.

Figure A.15: Formal and informal establishment distribution by age cohort in manufacturing: Rwanda (2013)



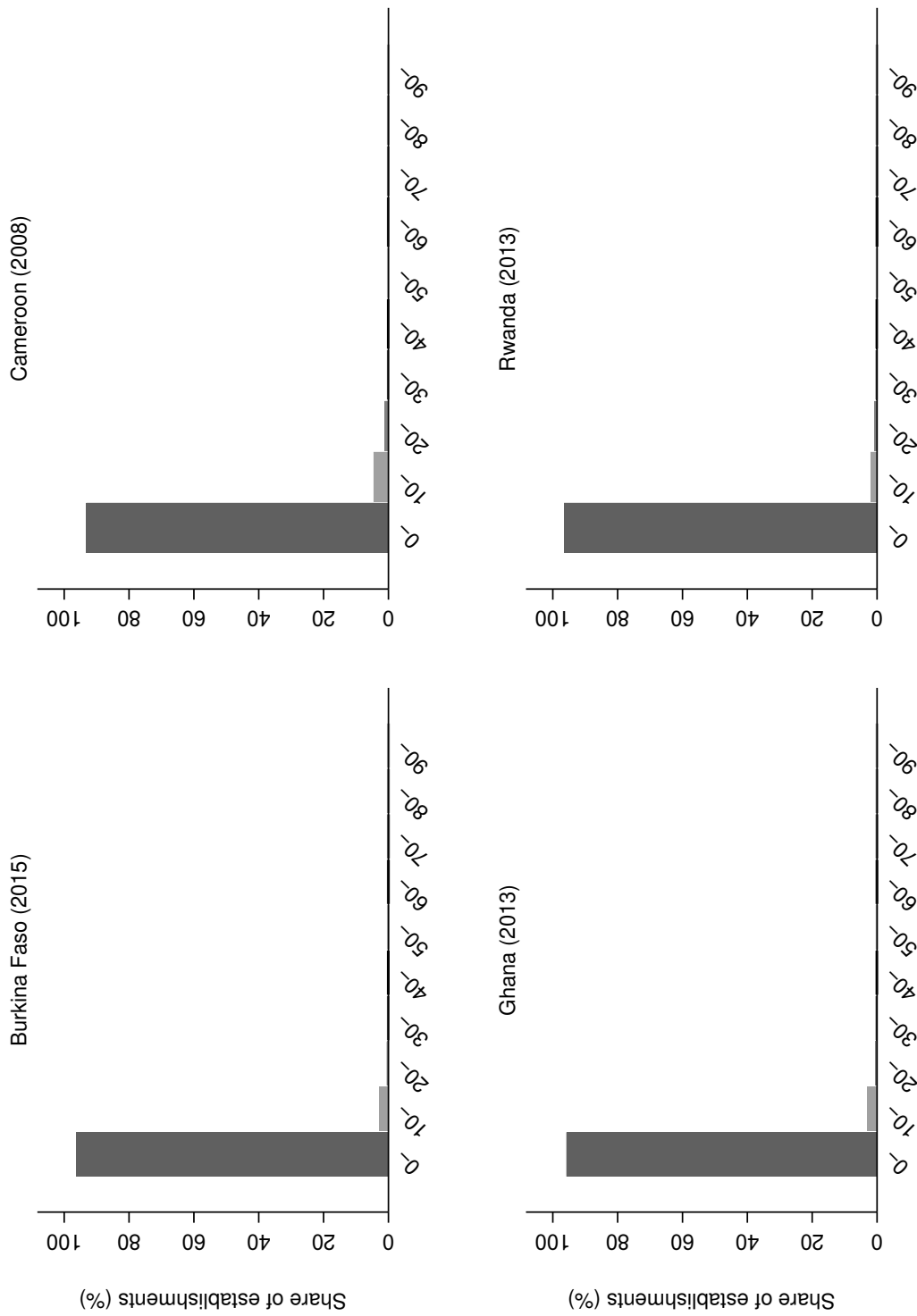
Source: Establishment censuses obtained from the statistical agencies of the selected countries; see section 3. Note: The reference is the total employment for each age cohort and formality status. The figure is constructed based on data for establishments in the manufacturing sector. Establishments with missing employment or age data and state-owned are excluded.

Figure A.16: Formality, age and establishment distribution in manufacturing



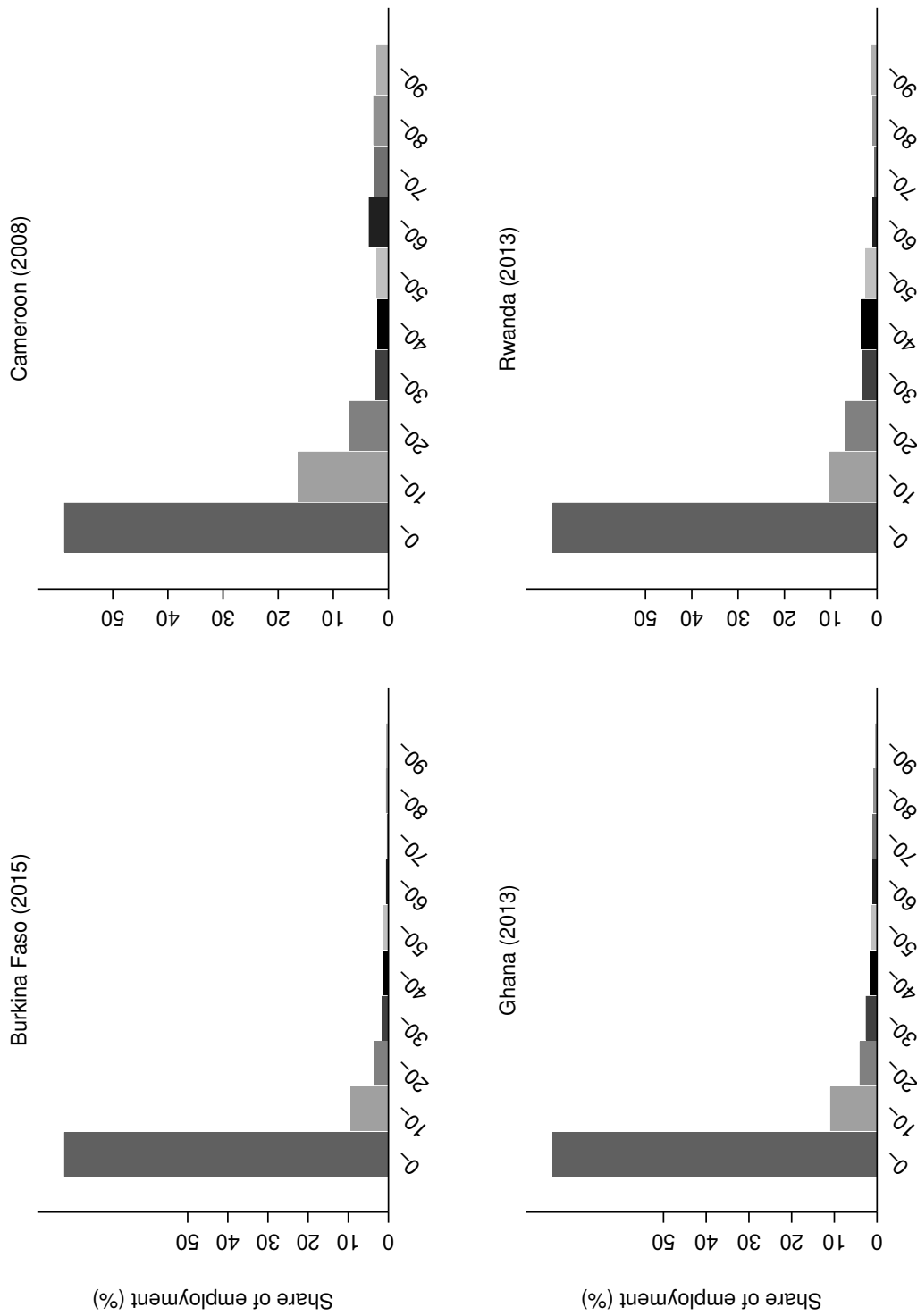
Sources: Establishment censuses obtained from the statistical agencies of the selected countries; see section 3.

Figure A.17: Establishment distribution in manufacturing (bins of equal width)



Sources: Establishment censuses obtained from the statistical agencies of the selected countries; see section 3.

Figure A.18: Employment distribution in manufacturing (bins of equal width)



Sources: Establishment censuses obtained from the statistical agencies of the selected countries; see section 3.