



PEDL Research Papers

This research was partly or entirely supported by funding from the research initiative Private Enterprise Development in Low-Income Countries (PEDL), a Foreign, Commonwealth & Development Office (FCDO) funded programme run by the Centre for Economic Policy Research (CEPR).

This is a PEDL Research Paper which emanates from a PEDL funded project. Any views expressed here are those of the author(s) and not those of the programme nor of the affiliated organizations. Although research disseminated by PEDL may include views on policy, the programme itself takes no institutional policy positions.

What is a firm census in a developing country? An answer from Ghana

Andrew Kerr and Bruce McDougall

DataFirst, University of Cape Town¹

CSAE Working Paper WPS/2020-11

Abstract

A burgeoning literature in economics uses firm census data to provide explanations for the very large differences in income per capita across countries. Much of this literature takes for granted that the coverage of firm censuses across and within countries is similar. In this paper we use data from four Ghanaian firm censuses conducted between 1962 and 2014 to show that the coverage of each census was very different. Treated as is, the four censuses show dramatic and unbelievable changes in the scale of manufacturing production in Ghana over this period. As a result, we examine and document important changes in what undertaking a “firm census” has meant over 50 years in Ghana, as well as documenting variation in the coverage of firm censuses from several other African countries. We show that it is possible to obtain a believable evolution of the firm size distribution in Ghana over the period for which we have firm microdata, but that this requires substantial work to understand how the coverage of each firm census has varied over time. Our paper shows that the coverage of firm censuses both within and across countries can differ quite dramatically, and that this can impact research that uses firm census data.

¹ andrew.kerr@uct.ac.za

This paper has been funded by PEDL (Private Enterprise Development in Low Income Countries). We thank the Ghanaian Statistical Service (GSS) for making the data available and to Anthony Krakah, Isaac Dadson and Jacqueline Anum from GSS for the extremely helpful assistance in understanding the data used in this paper. We thank Francis Teal for providing several helpful suggestions and comments, extra data and census documentation. Participants at SALDRU, IZA/GLM (Addis Ababa), PEDL (LSE), CAED (Michigan) and CSAE (Oxford) conferences and seminars provided very useful comments.

Part of the writing of the paper took place while Andrew Kerr was a CSAE fellow in Oxford from January to March 2020. I thank the CSAE for their generous hospitality.

As part of the PEDL project we will be making a sample of the 2014 IBES publicly available.

The population census data from Ghana that we use in this paper was downloaded from IPUMS International at the University of Minnesota. We acknowledge both IPUMS and the Ghana Statistical Service in using this data.

1. Introduction

A burgeoning literature in economics uses firm census data to provide explanations for the very large differences in income per capita across countries, including Hsieh and Olken (2014), Hsieh and Klenow (2014) and Cirera et al (2017). These authors all stress the importance of firm census data that covers firms in the informal sector, the size of which is large in the countries they study. Bento and Restuccia (2017, 2020) go further, arguing that their research on misallocation and the implications for the firm size distribution across countries requires the inclusion of the smallest informal firms, even own account workers without employees. Hsieh and Olken (2014) and Hsieh and Klenow (2014) simply assert that the data they use cover “all firms”, including those in the informal sector. And in working with census data from several countries the authors implicitly assume that the coverage across countries is similar, or that differences in coverage are not important.

The question we seek to answer in this paper is how firm census coverage can vary over time within one country, or across countries, and how this impacts research that uses firm census data. In reviewing the literature on measuring misallocation and distortions using data from firms and farms Gollin and Udry (2019:10) note that “The data used for analyses [of firm-level productivity] come from firm surveys that vary in quality and coverage.” These authors find that the impact of measurement error and heterogeneity on the measurement of misallocation across farms is substantial. Rotemberg and White (2019) show that the results from Hsieh and Klenow (2009), who use firm census data from China, India and the US to argue that extent of misallocation across firms can explain a substantial part of the cross-country variation in income per capita, are extremely sensitive to measurement error and the techniques used to correct for this and for item missing data. In this paper we focus on the variation in the *coverage* of firm census data in developing countries, which we argue has not been sufficiently interrogated.

Our contribution in this paper is to use micro data and firm census reports for four firm and population censuses in Ghana over 55 years to carefully investigate how coverage has changed, and in doing so make the data more comparable. When doing so the conclusions we reach are very different to those from a comparison which takes the data at face value, such as Sandefur (2010). Our analysis highlights that the coverage of firm census data can differ very dramatically across countries, and that coverage can also vary over time within the same country. We also show that population census or representative household survey data can provide a useful and simple check on the coverage of firm censuses of the smallest firms in developing countries, since these data are available for many countries.

We use micro data from the 1987, 2003 and 2014 Ghanaian firm censuses, as well as the 1962 firm census report and microdata from the 1984, 2000 and 2010 population censuses as well as three household surveys, to describe the evolution of the firm size distribution and show that coverage of the Ghanaian firm censuses has varied dramatically over time. We use these data to provide

several explanations for the varying coverage, and then to make the censuses as comparable as possible. We also use documentation from firm and population censuses in six African countries to show that there is huge variation in the coverage of firm census data.

The 1962, 1987, 2003 and 2014 Ghanaian firm census data shows unbelievable trends in the firm size distribution. The average firm size was 2.7 persons engaged in 1962 and then increased to 17 persons engaged in the 1987 census, despite GDP per capita declining by 40% over this period. By 2003 the average firm size had decreased to 9 persons engaged and then further declined to 4.5 persons engaged in 2014, despite GDP per capita tripling over this period. These results are even stranger considering that the 1987, 2003 and 2014 censuses all had the same stated target population.

We find that part of the explanation for these unbelievable trends is that the coverage of the smallest firms, of rural areas, and of associations (groups of self-employed workers) all varied quite substantially over the censuses, and that when taking this into account the adjusted firm size distributions are much more stable and believable. We also show that across several African countries for which there is firm census data both the target population and actual coverage differs quite dramatically. We argue that there are many sensible definitions of a firm when deciding which firms to include when conducting a firm census, but that researchers using cross country firm data should pay more attention to how this may differ within and across countries.

In section 2 we describe the changes in the Ghanaian firm size distribution and explore how coverage has changed over time by comparing firm and population census data. In section 3 we describe how the evolution of firm size distribution changes as we make the censuses more comparable. Section 4 shows that there are substantial differences in the coverage of firm census data from several African countries, and that this has not been considered in two recent papers using firm census data. Section 5 concludes.

2. Coverage Differences in the Ghanaian firm censuses

In this section we use firm and population census data from Ghana to show that the four firm censuses are not comparable.

2.1 Census Descriptions and Firm Size Distributions

The 1987, 2003 and 2014 Ghanaian firm censuses had, in theory, the same target population. This was all non-household establishments with a fixed location and household-based establishments if there was a sign advertising the establishment (GSS, 1989: 3, GSS, 2006, GSS 2015). The 1962 census had broader coverage that included “all small, medium and large establishments... including handicrafts and household type of units producing industrial commodities or rendering industrial services” (Ghana Central Bureau of Statistics, 1965). All four censuses included manufacturing, which is what we limit our analysis to below. There are almost no multi-establishment manufacturing firms in the Ghana census data, and we use the words firm and establishment interchangeably.

The censuses all incorporated a phase 1 where basic data on employment, industry and firm characteristics were collected on all firms in the target population. Phase 2 was a subsample of the firms enumerated in phase 1. In 1962 phase 2 was planned but not undertaken. In 1987 phase 2 was a census of firms with 10 or more persons engaged but no smaller firms were included. In 2003 phase 2 was again a census of firms with 10 or more persons engaged, but a probability sample of smaller firms was taken. In 2014 phase 2 was a census of firms with 50 or more persons engaged, and a probability sample was taken for smaller firms. In our analysis below we use data from phase 1 only.

The census documentation also indicates important changes in coverage of rural areas. The 2014 census was called the Integrated Business Establishment Survey (IBES). The IBES documentation states that the coverage was all urban and rural establishments. The 2003 and 1987 censuses were called the National Industrial Censuses (NICs). The 2003 NIC report states that “Enumeration was also generally confined to urban areas except for establishments listed in the interim register that were located in rural areas. Enumerators were also instructed to enquire from each settlement of the existence of any establishment in surrounding areas” (GSS, 2006: 6). The 1987 report states that enumeration was confined to settlements with a population of more than 500, although again enumerators were instructed to enquire about the existence of an establishment in surrounding areas (GSS, 1989: 3).

Table 1 below provides firm size distributions from the 1962, 1987, 2003 and 2014 censuses for the manufacturing sector. Firm size is measured throughout the paper by the number of persons engaged, which includes employees involved in production, employees not involved in production (e.g. administrators, supervisors), unpaid working proprietors and business partners, and paid and unpaid apprentices or family members. The last three rows of the table also shows the total number of persons engaged in the firms in the table, the share of small firm employment (firms with less than 10 persons engaged) and the mean firm size.

The changes in the firm size distribution shown in Table 1 are unbelievable. The number of firms enumerated in the 1962 census was around 95000, and the average size was 2.67. The number of firms enumerated in the 1987 census fell to 7 847, while the average persons engaged increased to 17 persons per establishment. Between 1987 and 2003 the number of firms enumerated tripled, whilst the average firm size nearly halved. Between 2003 and 2014 the number of firms enumerated quadrupled, whilst the average firm size again more than halved. These changes are also reflected in the share of small firm employment, shown in the second to last row, which declined from approximately 75% to 22% between 1962 and 1987, and then rose to 35% in 2003 and 58% in 2014. The overall trend seems to be one of dramatic growth in the small firm sector, and the stagnation and decline of the importance of large firms.

Table 1: The firm size distribution in Ghana.

Persons Engaged	1962		1987		2003		2014	
	Num	%	Num	%	Num	%	Num	%
1	49902	52.40	611	7.90	3055	11.79	31935	32.12
2-4	36521	38.40	2307	29.82	11155	43.07	48878	49.15
5-9	6787	7.10	3391	43.83	7846	30.29	14202	14.28
10-19	1364	1.43	727	9.40	2441	9.42	2935	2.95
20-29	246	0.26	226	2.92	528	2.04	588	0.59
30-49	132	0.14	155	2.00	372	1.44	394	0.40
50-99	105	0.11	142	1.84	260	1.00	237	0.24
100-199	58	0.06	78	1.01	118	0.46	128	0.13
200-499	38	0.04	52	0.67	88	0.34	87	0.09
500+	14	0.01	48	0.62	39	0.15	53	0.05
Number of firms	95167	100	7737	100	25902	100	99437	100
Total PE in all firms	254096		128777		240531		437316	
Share of PE in firms with < 10 PE	75.00		22.34		35.27		57.46	
Mean firm size	2.67		16.64		9.23		4.40	

Note: Own calculations from 1987, 2003 and 2014 firm censuses. 1962 data is from Ghana Central Bureau of Statistics (1965).

2.2 Using Population Censuses to Describe Firm Census Coverage

The previous section showed unbelievable trends in the Ghanaian firm size distribution, as well as the total firm count, which suggests that coverage of the censuses differed. In this section we use population census data to compare the coverage of rural areas and of the smallest firms in the Ghanaian firm censuses. In doing so we follow Sandefur (2010), and Teal (2016)², but substantially expand the scope of their analysis, and correct some important errors, resulting in substantially different conclusions. Sandefur (2010) neglected all the self-employed, which is where coverage varied the most across the censuses, whilst Teal (2016) incorrectly asserted that own account workers were not enumerated at all, when the variation in their coverage is a key part of the story we describe below. Our methods can be used to compare the coverage of firm censuses across countries, which we undertake for Ghana and five other African countries in Section 3.

Persons engaged by firms (employees, apprentices and unpaid workers) would be enumerated in the population census as such, and we can thus compare the number of persons engaged in the population and firm censuses, to check comparability across years. Almost all firms in Ghana are either sole proprietorships or partnerships owned by one or a few owners, who would all be enumerated as self-employed in the population census (which distinguishes between own account

² Wittenberg (2004) undertook similar comparisons for the 1996 South African manufacturing and population censuses.

self-employed and employers). This enables us to compare estimates of employers, own account workers and other non-owners (employees, unpaid workers and apprentices) from the firm and population censuses.

Table 2 shows data from the Ghanaian firm and population censuses and provides comparisons for manufacturing only. The three sets of rows in Table 2 show the number of own account self-employed, self-employed employers and then employees, apprentices and unpaid workers together. The columns show these for each year, broken down by rural and urban where possible. The population census estimates are shown in blue and firm census estimates are shown in red. The ratio of firm to population census is shown in the last row of each of the 3 sets of rows, which allows us to compare the coverage of the firm and population censuses over time. The population censuses were undertaken in 1960, 1984, 2000 and 2010, and the estimates are adjusted using linear interpolation to obtain estimates for the firm census years³. We believe that the comparisons show changes in the firm census coverage, since, whilst there are population census undercounts, the coverage is more stable than the firm censuses- which can be seen in the sensible changes in the population census numbers in Table 2.

Table 2: Firm and Population census coverage

		NIC 1987		NIC 2003		IBES 2014		NIC 1962			
		1	2	3	4	5	6	7	8		
Firm or Worker Cat.		Total	Total	Urban	Rural	Total	Urban	Rural	Total		
Own account workers (1 persons)	Pop censu	463421	544346	315038	229308	782501	444647	337854	Employers and own account workers	Pop census	177600
	NIC/IBES	611	3039	2025	764	31667	21191	9821			
	Ratio (%)	0.13	0.56	0.64	0.33	4.05	4.77	2.91			
Employers (>1 person engaged)	Pop censu	27538	76076	58974	17102	105141	83658	21483	Firm census	Ratio (%)	53.59
	NIC/IBES	7126	20776	15798	3875	66641	47835	17566			
	Ratio (%)	25.88	27.31	26.79	22.66	63.38	57.18	81.77			
Employees, apprentices, unpaid workers	Pop censu	155974	252841	188830	64011	362500	270144	92356	Employees, apprentices, unpaid workers	Pop census	57620
	NIC/IBES	128777.2	216317	177655	31506	334020	266452	64020			
	Ratio (%)	82.56	85.55	94.08	49.22	92.14	98.63	69.32			
Notes: own calculations from firm + population censuses											
RED= Firm Census		Blue= Population Census									

The first set of rows in Table 2 show the population and firm census estimates of the number of own account workers. There were 611 own account workers enumerated in the 1987 firm census, whilst the population census around the same time enumerated 463 000. This means that the firm census enumerated only 0.13% of all own account workers enumerated in the population census. For 2003 this ratio was 0.56, meaning that coverage was 4 times larger, but still very small. In 2014 firm census coverage increased substantially again, so that around the ratio of own account workers in the firm census to those in the population census was around 4%. Urban coverage was higher than rural coverage in both 2003 and 2014. For 1962 we cannot separate own account workers

³ The estimates for 1960 are taken from the census population report, which were derived from a 10% systematic stratified sample of individual-level records (Ghana Central Bureau of Statistics,1962) The estimates from 1984, 2000 and 2010 are estimated using the 10% samples available from IPUMS at the University of Minnesota.

and employers, so we discuss this only after the main results from the other 3 censuses. Details of how we classified firms as either rural or urban are in the Appendix.

The two main takeaways for own account workers are that the firm censuses enumerate only a very small fraction of all own account workers in Ghana, but that coverage increased very dramatically between 1987 and 2014. That such a small fraction of own account workers are enumerated in the firm censuses is not a problem per se- such workers would be enumerated in household surveys and policy makers would probably not want detailed data collected again in a firm census. But what is problematic is that coverage has varied so dramatically. If one is to use firm census data to undertake research on development and changes over time then one needs to use comparable data.

The second set of rows show the estimates of employers, who can be identified in the population census (through a direct question) and the firm census (through questions on ownership of the firm). These rows show that a much larger fraction of the estimated number of employers in the population census are also enumerated in the firm census. This was 26% and 28% in the 1987 and 2003 firm censuses respectively but more than doubled to 63% in the 2014 firm census. Thus, again the firm censuses coverage is lower than the population census, and again the coverage varied, but this time only between 2014 and the previous 2 censuses. Coverage of rural and urban areas was similar.

Given the large variation in coverage of different types of self-employed it would be useful to know more about which types of own account workers and employers are not included in the firm censuses but are in the population censuses. Unfortunately, there is very little overlap in the questions asked in the firm and population censuses and so this is not possible.

The last set of rows shows the coverage of paid employees, apprentices and unpaid family workers. Here the coverage is much higher and more similar, around 83% in 1987, 86% in 2003 and 92% in 2014. Sandefur (2010) showed that coverage in 1987 and 2003 was similar for this category of workers but did not check this for own account workers or employers, which we have just showed varied very dramatically.

For the 1960 population and 1962 firm censuses we do not have micro data and must rely on the published reports. They do not allow us to separate own account workers and employers but do give the total number of own account workers and employers. The last column of Table 2 shows that the coverage of own account workers and employers was much higher than in the three later censuses, as expected given the much wider target population. The firm census found 54% of the self-employed enumerated in the population census, whereas this was 11% in the 2014 firm census, and substantially less than this in 1987 and 2003.

The number of employees, apprentices and unpaid family workers enumerated in the 1962 firm census was two and a half times higher than the population census, which implies the firm census coverage of employees was much higher than in the population census. This issue is discussed in the 1962 firm census report, which notes that those employed in agro-processing were incorrectly

counted as being in agriculture in the population census, rather than in manufacturing (Ghana Central Bureau of Statistics,1965)⁴. Unfortunately, without further detail it is impossible to explore whether this is the full explanation.

3. Making the Ghanaian Firm Censuses Comparable

Sandefur (2010) argues that between 1987 and 2003 the mean firm size in Ghana declined and the number of small firms increased, that this is a sign of increased informality and that this flies in the face of a literature that suggests that the scale of production and formality should increase when market friendly reforms such as privatisation, capital account opening and trade liberalisation are implemented, as they were in Ghana in the 1980s and early 1990s. But we have shown that there was huge variation in the coverage of the firm censuses relative to the population censuses, and this likely explains some of the unbelievable trends in the firm size distribution. Thus, in this section we attempt to make the censuses as comparable as possible. In doing so we can estimate changes in the firm size distribution in Ghana that we believe more accurately reflect changes in Ghana over the last 30 years.

3.1 Limiting Coverage to Urban Firms

We showed in the previous section that the firm census coverage varied across urban and rural areas over time. We thus use the urban/rural classification to limit our analysis to urban areas, where coverage seems to have been more comparable. How we classified firms as rural or urban is described in the Appendix. Table 3 shows the firm size distributions limiting to urban firms only. We exclude 1962 since the report does not distinguish between rural and urban firms. Compared to the naïve results in Table 1, the mean firm size in 1987 has dropped by 2.8 persons engaged, because 1987 rural firms were larger than urban ones. In 2003 and 2014 the firm size increased by 0.8 and 0.5 persons engaged, because rural firms were smaller than urban firms⁵. Thus, limiting to urban firms, where coverage was more comparable, has substantially reduced the differences between the 1987 census, and the 2003 and 2014 censuses. The share of persons engaged in small firms has increased in 1987, and decreased in 2003 and 2014, although there are still large differences, suggesting that Ghana has seen a proliferation of small firms. In the next section we show that this trend is primarily driven by the coverage differences shown above in the comparisons with the population censuses.

⁴ This makes sense of the firm census persons engaged in manufacturing being much higher than in the population census. However, the report also states that those employed in food processing and distribution were classified as in manufacturing in the population census but that those involved in selling (ie distribution) were counted as being in retail trade in the industrial census. This is the opposite to what one would expect given the numbers in Table 2.

⁵ The reason that rural firms were larger on average than urban firms in 1987 is likely to be because coverage was more limited and so only the largest rural firms were enumerated. In 2003 and 2014, as one would expect, rural firms were smaller than urban firms.

Table 3: The Urban firm size distribution

Persons Engaged	1987		2003		2014	
	Num	%	Num	%	Num	%
1	567	8.11	2040	10.57	21354	30.59
2-4	2162	30.94	8263	42.81	34492	49.42
5-9	3160	45.22	6017	31.17	10626	15.22
10-19	569	8.14	1909	9.89	2176	3.12
20-29	181	2.59	409	2.12	421	0.60
30-49	121	1.73	252	1.31	311	0.45
50-99	111	1.59	192	0.99	175	0.25
100-199	53	0.76	102	0.53	112	0.16
200-499	35	0.50	82	0.42	80	0.11
500+	29	0.41	35	0.18	49	0.07
Number of firms	6988	100	19301	100	69796	100
Total PE in all firms	96179		195770		338571	
Share of PE in firms with < 10	27.97		32.81		53.43	
PE	13.76		10.08		4.85	
Mean firm size	13.76		10.08		4.85	

Note: Own calculations from 1987, 2003 and 2014 firm census microdata.

3.2 Making Coverage of the Bottom of the Firm Size distribution Comparable

In this section we use the firm and population census data to improve the comparability of the bottom end of the firm size distribution across the 3 most recent firm censuses. To do so we assume that the population census own account worker estimates are correct, and then add or remove small firms so that both the 2014 and 1987 firm censuses have the same ratio of firm census own account workers to population census own account workers in 2003 in urban areas, which was 0.6%. This ratio (shown in Table 2) was 0.13 in 1987⁶ and 5% in 2014. This means that to make the 2014 census comparable to 2003 we must exclude 84% of the own account workers enumerated in urban areas in 2014 and triple the number of own account workers in 1987 (from 611 to 1833). Table 4 shows the outcome of this adjustment. As expected, it reduces mean firm size in 1987 (by nearly 2 persons engaged) and increases it in 2014 (by around 1.5 persons engaged). As a result of the large number of own account workers enumerated in 2014 the adjustment decreases the total number of firms in 2014 by around 50%.

⁶ We cannot differentiate between urban and rural coverage in 1987.

Table 4: The firm size distribution for Urban firms with comparable coverage of own account workers

Persons Engaged	1987		2003		2014	
	Num	%	Num	%	Num	%
1	1701	20.94	2040	10.57	3048	5.92
2-4	2162	26.62	8263	42.81	34492	66.99
5-9	3160	38.91	6017	31.17	10626	20.64
10-19	569	7.01	1909	9.89	2176	4.23
20-29	181	2.23	409	2.12	421	0.82
30-49	121	1.49	252	1.31	311	0.60
50-99	111	1.37	192	0.99	175	0.34
100-199	53	0.65	102	0.53	112	0.22
200-499	35	0.43	82	0.42	80	0.16
500+	29	0.36	35	0.18	49	0.10
Number of firms	8122	100	19301	100	51490	100
Total PE in all firms	97313		195770		320265	
Share of PE in firms with < 10	PE		32.81		50.77	
Mean firm size	11.98		10.08		6.22	

Note: Own calculations from 1987, 2003 and 2014 firm census microdata.

In addition to large differences in the proportion of own account workers enumerated in the firm censuses, we also showed in Section 2 that coverage of employers was higher in 2014 than in 1987 and 2003, where it was roughly the same. In 2003 29% of employing sole proprietors in the population census were enumerated in the firm census. This was 26% in 1987⁷ so we do not adjust 1987 but it was 58% in 2014. Thus, to make 2014 comparable with 2003 we must exclude around 60% of IBES sole proprietors with employees, apprentices or unpaid workers.

Making coverage of employers comparable is more complicated than for own account workers because we do not know the sizes of the firms the employing sole proprietors own in the population census. We only know that we need to exclude 60% of employers in 2014 so that the coverage relative to the population census is the same as it was in 2003- 28%. There is also an additional constraint, because if we exclude employers, then we are excluding both owners and other persons engaged in these firms are dropped, which affects the count of non-owner persons engaged in the firms, which we also want to be comparable across censuses.

To undertake this adjustment, we assumed that the difference in the coverage of firms of size 2-9 between 2014 and 2003 was greatest in the smallest firms and then decreased linearly so that

⁷ Again, we cannot differentiate between urban and rural coverage in 1987, so we adjust using the overall ratio of firm to population census employers.

coverage of firms of size 10 and higher was the same and no adjustment was necessary. This means we dropped 77% of 2 person firms and 0.5% of 9 person firms, so that the overall effect was to drop 60% of all employing sole proprietors. This method resulted in the coverage of employing sole proprietors and employees, apprentices and unpaid family workers being the same in 2003 and 2014.

The key result of this exercise, shown in Table 5, is that mean firm size increases by 3 persons engaged in 2014, not far from the 2003 mean firm size of 10. The share of employment in small firms is now only slightly higher in 2014 compared to 2003. It should be noted that making coverage of small firms the same does not *automatically* imply mean firm sizes should be similar- this depends on the entire firm size distribution in both 2003 and 2014.

Whilst we can adjust the census data to improve comparability, the dramatic change in the coverage of small firms in Ghana between 2003 and 2014 is still puzzling and requires an explanation. One is the massive scale and budget of the 2014 census (which expanded to cover all sectors of the economy and ended up enumerating 635000 firms), so that the coverage of the smallest firms improved. Such a large operation could also have resulted in lower fieldwork quality and higher estimates if fieldworkers were not as well supervised. However, we have not been able to pin down an explanation, even after discussions with GSS. A further important difference between the NIC 2003 and IBES 2014 is in the way in which associations were enumerated, a subject we turn to next.

Table 5: Urban Firm Size Distribution adjusted for higher 2014 Coverage of small firms

Persons Engaged	1987		2003		2014	
	Num	%	Num	%	Num	%
1	1701	20.94	2040	10.57	3048	11.23
2-4	2162	26.62	8263	42.81	13034	48.04
5-9	3160	38.91	6017	31.17	7728	28.48
10-19	569	7.01	1909	9.89	2176	8.02
20-29	181	2.23	409	2.12	421	1.55
30-49	121	1.49	252	1.31	311	1.15
50-99	111	1.37	192	0.99	175	0.64
100-199	53	0.65	102	0.53	112	0.41
200-499	35	0.43	82	0.42	80	0.29
500+	29	0.36	35	0.18	49	0.18
Number of firms	8122	100	19301	100	27134	100
Total PE in all firms	97313		195770		247523	
Share of PE in firms with < 10 PE	28.81		32.81		36.30	
Mean firm size	11.98		10.08		9.12	

Note: Own calculations from 1987, 2003 and 2014 firm census microdata.

3.3 The Coverage of Associations

In Ghana associations are groupings of self-employed individuals organising to improve their working conditions or to arrange bulk discounts or joint purchases for their members. The 2003 NIC enumerated a substantial number of manufacturing associations, and they were included in the microdata as establishments, where the number of “persons engaged” was actually the number of association members. As a result 5% of “firms” enumerated in 2003 (1145 firms), were associations, whilst 14% of persons engaged in the firms enumerated in 2003 were in associations. The mean size of associations was 26, the largest had 1600 “persons engaged” and there were 60 associations of size 100 or more, out of only 250 firms in total with more than 100 persons engaged. By contrast, in the 2014 IBES and 1987 NIC only around 0.2% of manufacturing firms enumerated were associations, and their members were less than 0.5% of total persons engaged.

Including associations as firms does not seem correct. Whilst associations may undertake bulk buying, negotiating discounts on behalf of members or lobbying local government, their members are independent owners of their own businesses. Given the varying coverage of associations it thus seems appropriate to exclude associations from any analysis of the changes in the Ghanaian firm size distribution.

Table 6 below shows the size distribution when excluding associations, making coverage of the bottom of the firm size distribution comparable and excluding rural firms. The impact of excluding associations is negligible in 1987 and 2014 but in 2003 results in a decline of the mean firm size of around 0.7 persons engaged. The result is that mean firm size is now only slightly higher in 2014 than 2003 and the share of all persons engaged in small firms is the same in 2003 and 2014.

Undertaking 3 important adjustments for varying coverage has thus already radically altered the story of the changes in the Ghanaian firm size distribution between 1987 and 2014. Instead of an explosion of small firms, we have found an increase in the share of persons engaged in small firms and a decrease in mean firm size between 1987 and 2003, but almost no change in either of these between 2003 and 2014. Changes in coverage of the Ghanaian firm censuses are thus a very important part in describing changes in Ghana over the last 30 years.

Table 6: Firm size distribution excluding rural firms and associations and adjusted coverage of small firms.

Persons Engaged	1987		2003		2014	
	Num	%	Num	%	Num	%
1	1701	20.94	2035	10.99	3021	11.17
2-4	2162	26.62	8125	43.88	13001	48.08
5-9	3160	38.91	5803	31.34	7709	28.51
10-19	569	7.01	1733	9.36	2170	8.03
20-29	181	2.23	348	1.88	419	1.55
30-49	121	1.49	170	0.92	308	1.14
50-99	111	1.37	128	0.69	171	0.63
100-199	53	0.65	75	0.41	111	0.41
200-499	35	0.43	66	0.36	80	0.30
500+	29	0.36	32	0.17	49	0.18
Number of firms	8122	100	18515	100	27039	100
Total PE in all firms	97313		172683		246684	
Share of PE in firms with < 10	28.81		36.13		36.33	
Mean firm size	11.98		9.26		9.12	

Note: Own calculations from 1987, 2003 and 2014 firm census microdata.

3.4 State-owned Firms and Privatisation

Ghana undertook several reforms in the 1980s and 1990s as a result of an IMF structural adjustment programme. One of these was the privatisation of some state-owned firms, which could have led to these large privatised firms shrinking or exiting and reduce the number of large firms and mean firm size, which we investigate in this section. On the other hand, Frazer (2005) discusses the possibility that smaller and younger state-owned firms were not privatised by the state but were simply allowed to exit. This is an explanation for his result that state-owned firms alive in 1987 were more likely to exit than privately owned firms between 1991 and 1997, which could result in fewer small firms. The firm census micro data allow us to investigate which of these two effects can be seen in the changes in the firm size distribution.

State-owned firms were a substantial part of manufacturing in Ghana in 1987. They employed at least 16% of persons engaged in 1987, whilst another 10% of persons engaged were in firms with joint state and private ownership⁸. In contrast, the total persons engaged in either fully public or

⁸ We say “at least” because the 1987 data lacks information on the ownership status of all firms. We only know whether a firm is state-owned for the largest 583 firms in the 1987 census. Of these it is the 81 state-owned firms and 38 part-state, part privately-owned firms that employed 25% of all persons engaged, so if some smaller firms were also state-owned, the actual fraction of persons engaged in state-owned firms in 1987 could have been larger than this⁸. The mean firm size of the 81 state-owned firms in the 583 firms for which we do have ownership

jointly public and private firms was only 4% of all persons engaged in 1962, 5% in 2003 and 1% in 2014.

It is thus of interest to examine whether changes in state-owned firms had any impact on the firm size distribution. Table 7 shows the firm size distribution excluding state or partly state-owned firms, and again including only urban, non-association firms with the bottom of the firm size distribution made comparable. The mean firm size declines by 2 persons engaged in 1987 relative to Table 6, and only slightly in 2003 and 2014, whilst the share of persons engaged in small firms increases in 1987. This means the average firm size and the share of workers in small firms is now very similar across all 3 censuses.

Our efforts to improve the comparability of the firm censuses have thus led to a very different conclusion from the one drawn from Table 1 at the start of the paper. Rather than a huge influx of small firms between 1987 and 2014, the urban private sector firm size distribution in Ghana has been very stable. Our conclusion is that firm censuses within one country with the same or similar target populations can have very different coverage. In the next section we examine the changes in the top tail of the firm size distribution in Ghana.

Table 7: Private, urban, non-association firms with Comparable small firm coverage

Persons Engaged	1987		2003		2014	
	Num	%	Num	%	Num	%
1	1701	21.13	2032	11.04	3013	11.17
2-4	2162	26.86	8102	44.01	12978	48.11
5-9	3160	39.26	5784	31.42	7702	28.55
10-19	569	7.07	1715	9.32	2161	8.01
20-29	175	2.17	343	1.86	415	1.54
30-49	109	1.35	164	0.89	306	1.13
50-99	96	1.19	116	0.63	167	0.62
100-199	37	0.46	70	0.38	107	0.40
200-499	28	0.35	61	0.33	78	0.29
500+	12	0.15	24	0.13	46	0.17
Number of firms	8049	100	18411	100	26973	100
Total PE in all firms	76062		162437		241978	
Share of PE in firms with < 10	PE		38.29		36.98	
Mean firm size	9.45		8.76		8.97	

Note: Own calculations from 1987, 2003 and 2014 firm census microdata.

information in 1987 was 254, it was 320 for jointly owned firms and 90 for the other 464 largest privately-owned firms

3.5 The Shape of the Firm Size Distribution in Ghana

In this section we describe the shape of the firm size distributions in Ghana. A stylised fact is that the firm size distribution in many countries follows a Pareto distribution (Axtell 2001, Tybout 2014, Amirapu and Gechter 2020). The Pareto cumulative distribution function is $F(x) = 1 - \left(\frac{x_0}{x}\right)^\alpha$, where x_0 is the minimum firm size and α is the parameter governing the shape of the distribution (Axtell 2001). Smaller α values mean there is more mass in the right-hand tail of the distribution. Axtell (2001) finds that this parameter is between 1 and 1.1 for the US firm size distribution. In the individual income inequality literature a key decision is the choice of threshold value x_0 , which can make a substantial difference to inequality estimates (Jenkins, 2017). In the case of the firm size distribution we follow Axtell (2001) and use the minimum possible firm size, 1 person engaged, ie an own account worker.

The Pareto cdf can be rearranged as

$$\log(1-F(x)) = \alpha \log x_0 - \alpha \log x, \quad (1)$$

Equation 1 implies a linear relationship between $\log(1-F(x))$ and $\log x$. A simple non-parametric check is thus to graph this relationship (Wittenberg, (2017)). Figure 1 shows this for the 4 censuses, using a minimum firm size of 1. We do not have microdata for 1962 but can use the numbers of firms reported in 10 size categories and the proportion of all firms above the top value of each size category to plot 9 points where we know $\log(1-F(x))$ and $\log x$. All four censuses seem to show an approximately linear relationship between $\log(1-F(x))$ and $\log x$, until the very top of the distribution, where there are very few firms, suggesting that the shape of the distribution is fairly stable. The location of the line for each year reflects the variation in the coverage of the smallest firms. 1962 and 2014 had the greatest coverage of the smallest firms and so they lie below 1987 and 2003.

Figure 2 graphs the same relationship but using the set of firms for which the 1987, 2003 and 2014 census years are more comparable- ie urban, non-association, non-state-owned firms with the bottom tail made more comparable, as discussed in sections 2.1-2.4. This shows that the distribution looks remarkably stable across the latest 3 censuses. There are small humps where there are too few firms of size approximately 3-7 persons engaged. 1962 looks quite different, since we cannot undertake any adjustments to make it more comparable by reducing the number of small firms to equalise coverage of the bottom tail. The much larger coverage of small firms disguises that the rest of the distribution has remained remarkably similar since 1962, despite the very different economic strategies implemented by successive governments over time.

Figure 1: Non-parametric Pareto check

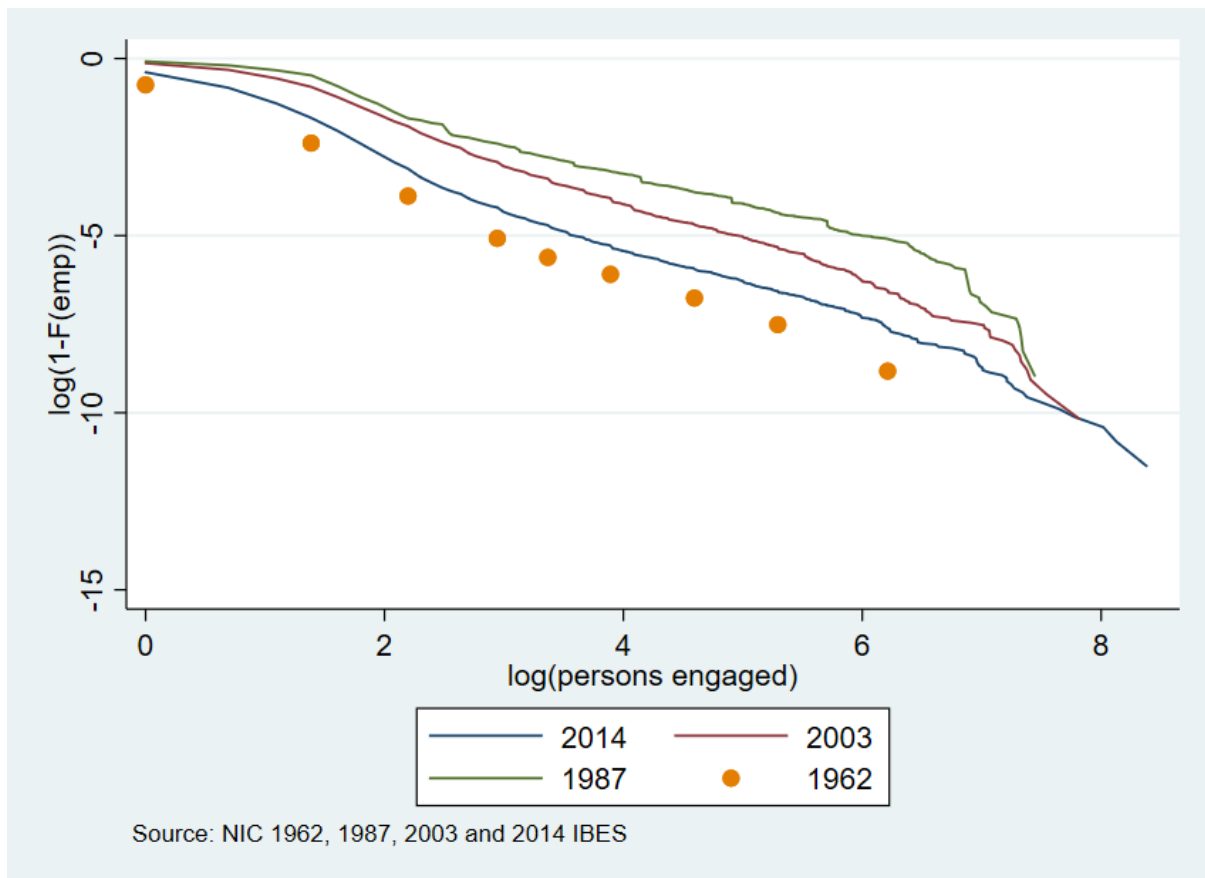
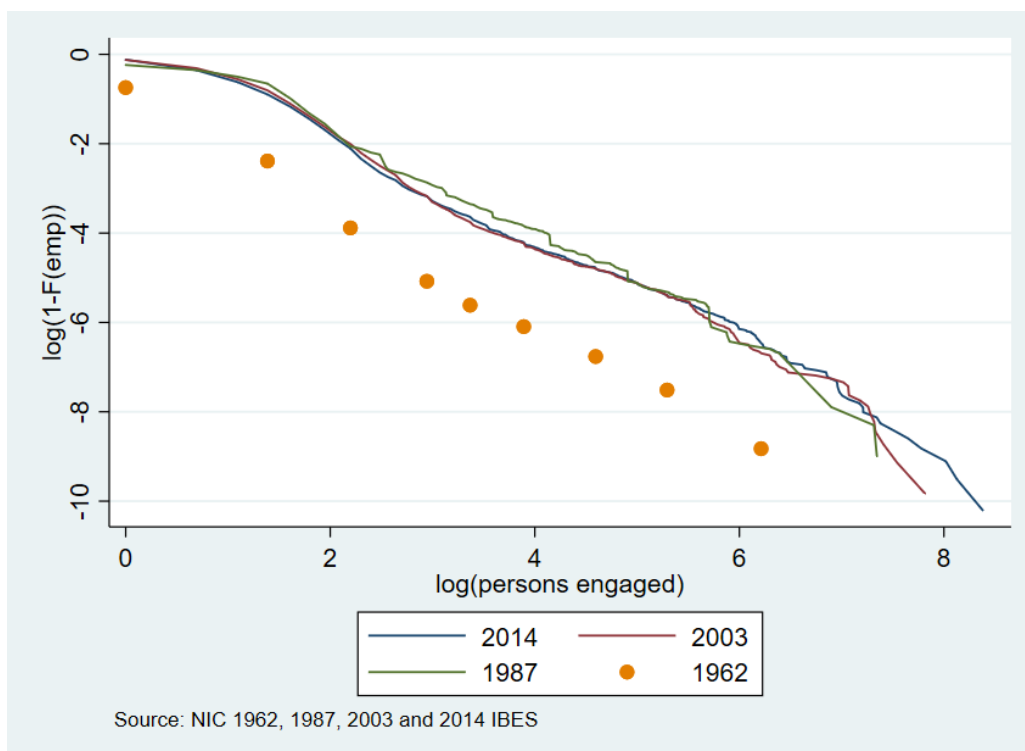


Figure 2: Non-parametric Pareto check with more comparable coverage



Axtell (2001) uses equation 1 and OLS to estimate α ⁹. Table 8 shows the OLS estimates of α for the Ghanaian firm size distribution. The left-hand side treats the data as is, whilst the right-hand side uses the more comparable data for 1987, 2003 and 2014. For 1962 we have the 9 size categories only, so the regression has only 9 observations. Smaller alpha values mean larger mass in the right tail of the distribution. The estimated α values in 1987 and 2003 are fairly close to 1, whilst they are larger for 1962 and 2014, where the coverage of the bottom of the distribution is better. The right-hand side shows alpha values after making the 3 latest censuses as comparable as possible, with all being close to 1, and similar to the estimates of Axtell (2001) for the US. One conclusion is that, as in the previous sections, if different censuses have differing levels of coverage this will impact the Pareto coefficient estimates.

Table 8: OLS Estimates of Pareto coefficient:

	Raw				Adjusted		
	1962	1987	2003	2014	1987	2003	2014
Alpha	1.29	0.92	1.03	1.21	0.92	1.07	1.08
Standard error	0.05	0.00	0.00	0.00	0.00	0.00	0.00

Note: Own calculations from 1987, 2003 and 2014 firm census microdata and 1962 census report

3.6 A robustness check using household survey data

We have argued that the coverage of associations, rural areas and small firms have all contributed to varying coverage of the Ghanaian firm censuses, which give very similar firm size distributions when made comparable. One check on this result is to examine changes in average firm size using the Ghana Living Standards Surveys (GLSS). These are household surveys that ask paid workers, apprentices and unpaid family workers what size firm they work in. We use GLSS 3 (1991), 5 (2005/6) and 6 (2012/3) to explore changes in the firm size distribution and follow an approach used in the World Development Report 2013 (World Bank, 2013) to measuring the average firm size using household survey data. We divide the individual weight by the firm size reported by each paid and unpaid person engaged. This can then give us an unbiased estimate of the correct firm size distribution, albeit one that excludes owners and is estimated in a very different way to directly using the firm censuses. One weakness of this method is the relatively small sample sizes (134 individuals in manufacturing in GLSS 3, 283 in GLSS 5 and 348 in GLSS 6).

⁹ Tybout (2014) estimates Pareto parameters without firm level data. He uses the fact that the Pareto distribution implies regularities in the share of employees in different firm size categories to compare estimated and actual shares of employees in 3 firm size categories. He finds that countries with lower average incomes have higher Pareto parameters and thus less fat tails than the US. This method is similar to methods used in the “top incomes” literature, see Atkinson (2016). Jenkins (2017) uses the maximum likelihood estimator of alpha in his work on income inequality, implemented in the Stata command `paretofit`. The MLE of alpha is constrained to be <1 when the cutoff value is 1, as it is in work on the firm size distribution, so we do not use the MLE in the work in this paper.

Table 9 shows that in the 1991 GLSS the mean firm size was 7.5, 4 in 2005/6 and 3.7 in 2012/3. The differences in magnitudes of the mean imply that the coverage of the smallest firms is broader in household surveys, similar to the population census. But what is interesting is that there is no change in the mean between 2005 and 2012, which is the result we obtained for the firm censuses once we had made coverage comparable. And there is also a decrease in mean firm size between 1991 and 2005, which we also found once we made coverage comparable, though the decline is larger than in our results. Thus, whilst the GLSS household surveys are not the ideal method for estimating changes in average firm size, they do broadly fit with our results from the firm census data.

Table 9: Mean firm size using GLSS household survey data

	GLSS		
	1991	2005/6	2012/13
Mean Firm size	7.5	4	3.7
95% Confidence interval	5.6-9.5	3.6-4.5	3.4-4.0

Source: Own calculations from GLSS 1991, 2005/6 and 2012/13.

4. Firm census coverage in sub-Saharan Africa

We have shown that coverage of the Ghanaian firm censuses varied substantially, despite similar target populations in the 3 most recent censuses. In this section we show that the target populations and actual coverage varies substantially across firm censuses in 5 other African countries where firm census data has been collected. As described above, the 1987, 2003 and 2014 Ghanaian firm census target population was all firms with their own premises, plus household-based firms if there was a sign advertising the business. Our view is that this is a pragmatic and sensible target population, even if actual coverage varied substantially, but that other sensible target populations have been used. We wish to highlight that these differences are large and thus very important for cross country comparisons.

The Ethiopian Large and Medium Manufacturing Industries Survey has been used in several recent papers (eg. Shiferaw et al 2015). It includes all manufacturing firms of larger than 10 workers that use electricity in production. There is also a separate Small-Scale Manufacturing Industries Survey of firms of size less than 10 that use power driven machinery (Soderbom 2012).

Soderbom (2012) provides some helpful descriptive statistics using these two Ethiopian surveys in 2007. There were around 45300 firms in total, with 43 338 in the small firm survey and around 2000 in the large and medium firm census. This could mean roughly 50 000 sole proprietors and partners (a firm can have more than one owner) are enumerated in the census, along with at most 2000 firms whose owners are not people and so would not be enumerated as employers in the population census. The population census of 2007 did not ask questions about industry of the self-employed or employers, but the 2005 Ethiopian Labour Force Survey found there were 705 000 self-employed with no employees in manufacturing (Central Statistical Agency of Ethiopia, 2006).

Thus, a rough approximation is that the Ethiopian censuses included about 6% (45 000/705 000) of all the self-employed enumerated in the population census.

That “only” 6% of all the self-employed were included is not to say that the Ethiopian National Statistics Office made an error, but simply that the criteria for inclusion (power driven machinery in Ethiopia’s case) excludes a large fraction of the smallest firms/own account operators. The 2003 Ghana firm census included only around 0.5% of own account workers, whilst the 2014 census included about 4%. The key point is that coverage can differ quite substantially across different countries, or even within countries over time (as we have shown in Ghana), and this is likely to impact any cross-country analysis or analysis over time within countries.

The Rwandan firm census of 2011 is an example of a firm census that had a much broader target population, and thus captured a much larger share of the self-employed in manufacturing enumerated in the population census. The 2011 firm census was supposed to include all formal and informal establishments with a fixed site (National Institute of Statistics of Rwanda, 2011). The firm census found 123000 manufacturing establishments, around 33% of the number of self-employed identified as working in manufacturing in the 2012 population census.

The Ivory Coast and Kenyan firm censuses used by Cirera et al (2017) use a much stricter definition of a firm as only firms registered for tax are included. Since almost all of these are not owned by individuals, approximately 0% of the self-employed enumerated in the Kenyan and Ivory Coast population census would be included in these firm censuses.

South Africa conducts a survey of tax (VAT) registered firms and a nationally representative household survey of non-tax registered household businesses. When combined, as in Bento and Restuccia (2017), there is 100% coverage of own account workers, compared to 33% in the Rwandan firm census, approximately 7% in the Ethiopian censuses, 4% and 0.5% in the 2014 and 2003 Ghanaian firm censuses and 0% in the Ivory Coast and Kenyan firm censuses. There is thus massive variation in coverage of the smallest firms across countries, in addition to differences in actual coverage even with the same target population that we have shown for Ghana. Bento and Restuccia (2017) claim that the data they used is comparable and includes all firms, including own account workers. But these authors used the Rwandan, Ghanaian and South African firm data that we have described above as is, despite the three data sources including 100%, 33% and 4% of own account workers respectively. This is incredibly large variation and suggests cross country comparisons using such data is very problematic. We suggest that paying attention to both stated and actual coverage of firm censuses is important and that population census data or appropriately weighted household survey data provide a simple check on firm census data coverage.

Table 10: Firm census coverage in Sub-Saharan Africa

Country	Definition
Kenya and Ivory Coast	Tax-registered firms only Implies 0% coverage of own account self-employed and ~0% self-employed employers.
Ghana	“All formal and informal firms, only including hh firms if there is a sign indicating the presence of the firm” 0.5% own account self-employed enumerated in 2003, 4% in 2014.
Ethiopia	All firms with more than 10 workers, and a sample of firms of size less than 10 that use power driven machinery. 2007 firm census had 45000 firms <10 enumerated, 2000 10+, LFS shows 700k own account workers, and 9000 employers. (Soderbom, 2012). Approx. 6% of self-employed included
Rwanda	“All formal and informal establishments with a fixed site ” (incl hh) 1 in 3 self-employed own account workers enumerated in 2012.
South Africa	Representative survey of VAT registered firms, including a census of larger firms, unregistered firms identified through household sample surveys Implies 100% coverage of all self-employed including own account workers (when weighted).

The inclusion of the smallest firms and own account workers is not just an important issue for developing countries. Axtell (2001) highlights that in 1997 there were 5.5 million firms with at least 1 employee in the US but another 15 million with no employees. Using the numbers reported in Axtell (2001) and including these firms as size 1 would decrease the US mean firm size in 1997 from 19 to around 5.8. This shows that coverage of the smallest firms and own account workers makes a huge difference to the firm size distribution in a developed country also.

5. Conclusion

In this paper we have described changes in the scale of production of Ghanaian manufacturing between 1962 and 2014 using data from four firm censuses. Taken as is the censuses suggest unbelievable trends in the size distribution of manufacturing firms in Ghana, despite the same stated target populations in the 1987, 2003 and 2014 censuses. Using population census data allowed us to show that there were large differences in the coverage of the firm censuses of rural and urban areas, as well as the smallest firms, and allowed us to improve comparability and obtain a more accurate picture of the changes in the firm size distribution in Ghana between 1987 and 2014. When doing so, we find very different results than when using the data as is. Instead of a burgeoning small firm sector since 1987 we find a roughly constant share of employment in small firms, and almost no decline in mean firm size in privately owned firms. Privatisation of large firms between 1987 and 2003 likely explains some of the decline in mean firm size over this period.

Sandefur (2010) argued the proliferation of small firms in Ghana following business friendly policy changes in the 1980s was a puzzle. Our analysis of the data suggests less of a puzzle- since there was no proliferation of small firms. Nevertheless, there has not been an *increase* in the importance of large firms or in the scale of production in Ghana following the business-friendly policies put in place since the 1980s.

In addition to large variation in firm census coverage within Ghana, we have also shown that the coverage of small firms and own account workers in firm censuses across six African countries varies dramatically. This will matter for attempts to compare mean firm size or firm size distributions across countries, such as Bento and Restuccia (2017), who used supposedly comparable firm census data that actually cover 0.5% of own account workers in Ghana, 33% in Rwanda and a combination of household and firm survey data for South Africa covering (when weighted) 100% of own account workers.

One suggestion that would improve attempts to investigate firm census coverage would be to have questions that allow firm and population census (or household survey, appropriately weighted) to speak to each other. A question in the Ghana Living Standards Survey on whether the household business has a sign would have allowed direct comparison of coverage of the stated target population of the firm censuses over time. Including urban or rural location in the firm census would also help to compare coverage. Unfortunately, the number of African countries where both firm and population census data are publicly available and that allow detailed coverage checks is very low.

Firm census data are usually unavailable because they are more sensitive, and because secure methods of data access are not available in most developing countries. An alternative way of obtaining more information on firms is for household surveys to ask more questions about firms, including a more detailed firm module for larger firms or firms operating from a non-household premises. The Ghana Living Standards Surveys are a good example of this. But given how few large (say 10 or more employees) firms exist, relatively small household sample surveys (perhaps 5000-30000 households) will enumerate very few of these. The upside of this situation is that a firm census covering larger firms could only require enumerating a few thousand firms in most low or lower-middle income countries, where resources are limited anyway, which could then be complemented with household survey data on small firms. But this would require coordination between industrial statistics and household survey sections of national statistics offices, which can be siloed.

The results in this paper are of relevance for several reasons. The first is that the firm size distribution has been increasingly recognised as an important marker of development (Hsieh and Klenow (2014), Bento and Restuccia (2017, 2020)). Our analysis shows that measuring changes in the firm size distribution in developing countries is difficult, even in Ghana, where supposedly comparable firm censuses have been undertaken and the data made publicly available. The conclusions reached using the naïve comparison and the most comparable data are dramatically different. The naïve comparison suggests there has been a proliferation of small firms, and a

decline in the importance of large firms. When looking at the more comparable data there has been little change in the share of persons engaged in small firms and the firm size distribution. This difference is important in any attempt to describe and explain economic outcomes and evaluate policy in Ghana. Our discussion of how firm censuses in six African economies were conducted and what the target population and actual coverage were also means that what a firm census is can differ markedly across countries. We do not think this point has received enough recognition in the recent literature using firm census data from developing countries.

6. References

- Amirapu, A. & Gechter, M. (2020). Labor Regulations and the Cost of Corruption: Evidence from the Indian Firm Size Distribution. *The Review of Economics and Statistics*, 102(1), 34-48.
- Atkinson, A. B., (2016). Pareto and the upper tail of the income distribution in the UK: 1799 to the present. *Economica*. 84, 129–156.
- Axtell, R. (2001). Zipf Distribution of U.S. Firm Sizes. *Science*. 293 (5536). 1818-1820.
- Bento and Restuccia (2017). Misallocation, Establishment Size, and Productivity. *American Economics Journal: Macroeconomics*. 9(3) 267-303.
- Bento, P. and Restuccia, D. (2020). On average establishment size across sectors and countries. *Journal of Monetary Economics*. Available at <https://doi.org/10.1016/j.jmoneco.2020.01.001>
- Central Statistical Agency of Ethiopia (2006). Report on the 2005 National Labour Force Survey. Statistical Bulletin 365. Available at <http://www.social-protection.org/gimi/RessourceDownload.action?ressource.ressourceId=14555>
- Cirera, X. Fattal Jaef, R and Maemir, H. (2020). Taxing the Good? Distortions, Misallocation, and Productivity in Sub-Saharan Africa. *The World Bank Economic Review*, 34(1). 75–100.
- Ghana Central Bureau of Statistics (1965). 1962 Industrial Census Report. Accra. Available at the British Library.
- Ghana Central Bureau of Statistics. (1962), Population Census of Ghana. Advance Report of Volumes III and IV. Accra. Available at the British Library.
- Ghana Statistical Service (1989). Ghana National Industrial Census 1987 Phase 1 report.
- Ghana Statistical Service (2006). 2003 National Industrial Census Report. Available at <http://www2.statsghana.gov.gh/nada/index.php/catalog/33> Accessed 11 February 2020.
- Ghana Statistical Service (2015). Integrated Business Establishment Survey Summary Report. Available at https://statsghana.gov.gh/gssmain/fileUpload/pressrelease/SUMMARY%20REPORT_FINAL_FINAL_24-5-16.pdf
- Frazer, G (2005). Which Firms Die? A Look at Manufacturing Firm Exit in Ghana, *Economic Development and Cultural Change*, 53(3), 585-617.
- Hsieh, T and Klenow, P. (2009). Misallocation and manufacturing TFP in China and India. *Quarterly Journal of Economics*, 124(4):1403-1448.
- Hsieh and Klenow (2014). The Life Cycle Of Plants In India And Mexico. *The Quarterly Journal Of Economics*. 129(3). 1035-1084.

Jenkins, S (2017). Pareto Models, Top Incomes and Recent Trends in UK Income Inequality, *Economica*, 84(334), 261-289.

National Institute of Statistics Of Rwanda, 2011. Establishment Census 2011 Analytical Report. Available at

http://www.minicom.gov.rw/fileadmin/minicom_publications/Planning_documents/Establishment_Census_2011_FINAL_.pdf

Sandefur, J. (2010). On the Evolution of the Firm Size Distribution in an African Economy. *CSAE Working Paper*, 2010-5.

Admasu Shiferaw and Måns Söderbom and Eyerusalem Siba and Getnet Alemu (2015). Road Infrastructure and Enterprise Dynamics in Ethiopia. *The Journal of Development Studies*. 51(11). 1541-1558.

Soderbom, M. (2012). Firm Size and Structural Change: A Case Study of Ethiopia. *Journal of African Economies*, 21(2), ii126–ii151.

Teal, F. (2016). Firm Size, Employment and Value-added in African Manufacturing Firms: Why Ghana needs its 1 per cent. *CSAE Working Paper* 2016-07.

Teal, F. (2017). What constrains the demand for labour in firms in Sub-Saharan Africa? Some evidence from Ghana. Working Paper.

World Bank. (2013). *World Development Report 2013: Jobs*. Washington, DC: World Bank. DOI: 10.1596/978-0-8213-9575-2

Wittenberg, M. (2004). "The Mystery of South Africa's Ghost Workers In 1996: Measurement And Mismeasurement In The Manufacturing Census, Population Census And October Household Surveys." *South African Journal of Economics*, 72(5), 1003-1022, December.

Wittenberg, M. (2017). The top tail of South Africa's earnings distribution 1993-2014: Evidence from the Pareto distribution. *REDI3x3 Working paper* 46.

Data sets

Minnesota Population Center. Integrated Public Use Microdata Series, International: Version 7.2 [dataset]. Minneapolis, MN: IPUMS, 2019. <https://doi.org/10.18128/D020.V7.2>

Ghana Statistical Service (1987) National Industrial Census 1987 [dataset]. Available at <https://www.empiricalde.com/#!ghana-firm-census-1987-and-2003/c6h5>

Ghana Statistical Service (2003) National Industrial Census 2003 [dataset]. Available at <https://www.empiricalde.com/#!ghana-firm-census-1987-and-2003/c6h5>

Ghana Statistical Service (2014) Integrated Business Establishment Survey [dataset]. Not currently publicly available.

7. Appendix: Classification of firms as Urban or Rural

To identify whether firms are urban or rural, we use firm address data directly from the firm census microdata, as well as 2010 population census enumerator area-level data. We obtained a list of 2010 population census EAs provided by GSS, classified as either rural or urban. Unfortunately, we did not have access to such a file for the 2000 or 1984 population censuses. The 2014 IBES and 2003 NIC firm censuses we use have location data on each firm, including the region, district, sub-metro, locality name and suburb, which we then merged to 2010 population census EA names.

We were able to classify 98.5% of 2014 firms and 94% of 2003 firms as rural or urban using the 2010 population census data. We began by merging the IBES/NIC firm location to 2010 population census EAs using region and locality name. This matched two thirds of the firms in 2014. We next merged on region and locality name, which matched another 25%. Using region and locality name classified another 2%, as did classifying firms whose locality name was Accra, Tema or Kumasi as urban. We then merged on suburb and district, and then classified a further 1% in 2014 as rural or urban based on whether an entire district's EAs were either rural or urban. Another 3 matching options resulted in another 2.5% being matched.

For 2003 we tried to match 2003 location data to the 2010 population census EA list. Matching on region and town name matched 46% of firms. Matching on region and suburb matched another 24%. We then used the 2000 population census to find districts that were all rural or all urban and then classified another 13% of firms if they were in those types of districts. Various other matching techniques resulted in another 11% of firms being classified as rural or urban.

Classifying 1987 firms as either rural or urban using the 2010 population census data is more difficult, because of the greater time that has elapsed and because we only have firm location data about the district the firm was in (there were 106 districts in 1987). To impute firm location as urban or rural we matched 1987 districts to 2003 districts. We then assumed a district and all firms in that district were rural if the majority of the enumeration areas in the district were rural in 2003. Undertaking this imputation results in rural firms comprising 9.6% of all firms in 1987, 21% in 2003 and 28% in 2014. This implies increasing coverage of rural areas, especially considering that Ghana was urbanising rapidly over this period. We do not have 1984 population census rural or urban data, so we cannot compare coverage of the 1987 firm census in urban versus rural firms. We use the location data in our analysis in section 2.3.