DataFirst

Firm Size and Economic Development: Evidence from Population Censuses and Household Surveys

Andrew Kerr and Nicholas Forster

Technical Paper Series
Number 43

About the Author(s)

Andrew Kerr: School of Economics and DataFirst, University of Cape Town

Nicholas Forster: School of Economics, University of Cape Town

This research has been funded by the Structural Transformation and Economic Growth (STEG) research programme, through grant number 1182.
The authors acknowledge all the statistical offices that provided the underlying data making this research possible.

Recommended citation

Kerr, A. Forster, N. (2023). Firm Size and Economic Development: Evidence from Population Censuses and Household Surveys. DataFirst Technical Paper 43. Cape Town: DataFirst, University of Cape Town.

© DataFirst, University of Cape Town, 2023

DataFirst, University of Cape Town, Private Bag, Rondebosch, 7701, Tel: (021) 650 5708,
Email: datafirst@uct.ac.za/support@data1st.org

Firm Size and Economic Development: Evidence from Population Censuses and Household Surveys

Andrew Kerr[1] and Nicholas Forster[2]

DataFirst Technical Paper 43
University of Cape Town
December 2023

Abstract

In this paper we investigate the relationship between mean firm size and economic development. This relationship has received attention in the recent literature on resource misallocation and the extent to which this drives differences in economic development. To investigate this relationship, we use the insight that mean firm size can be estimated using population census or household survey data that does not contain information on the size of the establishment a worker works in or a firm owner owns. We find a positive relationship between firm size and GDP per capita using OLS, as in previous work, but show that this relationship is substantially stronger than those that have been estimated previously. We show that this difference is not due to differences between the composition of our sample and those in prior work. But because we have multiple observations for some countries we can also show that changes in GDP per capita are not correlated with changes in mean firm size. This is an inconvenient complication for the literature on firm size and economic development.

[1] School of Economics and DataFirst, University of Cape Town
[2] School of Economics, University of Cape Town
This research has been funded by the Structural Transformation and Economic Growth (STEG) research programme, through grant number 1182.
The authors acknowledge all the statistical offices that provided the underlying data making this research possible.

1. Introduction

Do cross-country differences in resource misallocation amongst firms explain differences in economic growth and structural transformation, through their impact on Total Factor Productivity (TFP)? Several recent papers have argued that examining cross-country differences in mean firm size helps answer this question (Hsieh and Klenow, 2009; Bento and Restuccia, 2017; 2021; Poschke, 2018). This paper investigates whether the stylised fact shown by Bento and Restuccia (2017; 2021) and Poschke (2018)– a positive relationship between mean firm size and the level of economic development- holds up when using more comparable data for a large set of countries over many time periods.

The contribution made in this paper is to use comparable sources of data to estimate firm size- population census microdata from IPUMS, aggregated population census data from the United Nations Statistics Division (UNSD) and comparable household survey data aggregates from the International Labour Organisation- and a simple method to measure mean firm size across many countries and time periods. We extend the work of Bento and Restuccia (2017; 2021) by using a more comparable source of cross-country data on mean firm size over many time periods. We extend the work of Poschke (2018) by using more time periods and countries, particularly in sub-Saharan Africa, which were largely excluded.

Our research also relates to the literature on the relationship between self-employment shares and economic development (La Porta and Schleifer, 2014), which has been investigated in much previous work. In our analysis we use the insight of Salas-Fumás and Sanchez-Asin (2019), that the self-employment share is the inverse of mean firm size. This insight allows us to use household survey and population census data to contribute to the recent literature on the relationship between mean firm size and economic development but has been overlooked.

Population census microdata for many countries over a long period of time has been harmonised and made publicly available by IPUMS at the University of Minnesota. Population census data has also been harmonised across country by the UNSD, although this is only available at aggregate levels. In addition, the ILO has aggregate statistics on different types of workers from household surveys for many countries and time periods. These data and the insight of Salas-Fumás and Sanchez-Asin (2019) that mean firm size is the inverse of the self-employment share means that we can create measures of mean firm size using population census and household survey data, that are more comparable than those previously estimated, even though these data sources do not have direct questions about firm size.

In the next section we review the literature relevant to our key question. Section 3 describes the data that we use. Section 4 provides some descriptive statistics from these data. In Section 5 we investigate the relationship between mean firm size and GDP per capita. Section 6 concludes.

2. Literature Review

Recent research in macro-development has focused on the misallocation of resources as a key driver of cross-country differences in total factor productivity, which in turn drives differences in economic development across countries (Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009; Bento and Restuccia, 2017; 2021). Some of these papers have emphasised that examining cross-country differences in mean firm size is important for understanding the extent to which misallocation explains differences in economic development across countries is correct (Bento and Restuccia, 2017; 2021, Poschke, 2018), because the misallocation of resources may occur across firms, resulting in unproductive firms having more capital and labour than is optimal (Bento and Restuccia, 2017).

In this section we review the existing literature on the relationship between firm size and economic development. Several recent papers, using different types of data, have found a positive relationship between firm size and GDP per capita. A helpful way of distinguishing between these approaches is by the different types of data used. Thus, we begin with surveys or censuses that have the firm or owner as the subject. We then move on to surveys or censuses that have the employed as the subject.

2.1. Firm and Firm Owner Survey Approach

The key papers with the firm or firm owner being the subject of the survey or census used are Bento and Restuccia (2017; 2021) and Poschke (2018). Bento and Restuccia (2017; 2021) modelled the effect of correlated distortions on productivity investments by establishments and the resulting firm size distribution in a country. Correlated distortions are added costs and constraints faced by larger firms. They act as implicit and explicit taxes imposed on larger firms and thus disincentivise firms from growing larger through productivity-increasing investments. Bento and Restuccia (2017; 2021) assume that firms in developed countries face lower correlated distortions than firms in developing countries, resulting in firms in developed countries being larger on average than those in developing countries. Bento and Restuccia (2017) study manufacturing, while in 2021, they add in services.

To validate their theoretical predictions, Bento and Restuccia (2017) made an extensive effort to create a comparable cross-country measure of mean firm size in manufacturing using hundreds of firm censuses and nationally representative surveys undertaken between 2000 and 2012, in which developing countries are well represented. They then investigated the relationship between mean firm size and GDP per capita in the 132 countries which they obtained data for. They find a significant and positive elasticity of 0.29 of mean firm size with respect to GDP per capita in manufacturing. Bento and Restuccia (2021) added new data on mean firm size in services and found an elasticity of firm size with respect to GDP of 0.33.

Poschke (2018) acknowledges the importance of correlated distortions but argues that different levels of technological development also drive mean firm size differences. In Poschke's (2018) occupational choice model, exogenous technological improvements benefit more highly skilled entrepreneurs the most and result in lesser-skilled entrepreneurs moving to wage employment. Fewer entrepreneurs and better technology for the remaining entrepreneurs increases mean firm size.

To represent the global firm size distribution across countries, Poschke (2018) uses two datasets. First, he uses data from the Amadeus database compiled by Bureau van Dijk. This dataset covers 33 different European countries and is compiled from information filed by both public and private companies in local registers that are publicly accessible. Second, to include developing countries, Poschke (2018) uses the Global Entrepreneurship Monitor (GEM). This includes individual surveys of entrepreneurs and business owners from over 50 countries, although there are only two African countries.

Like Bento and Restuccia (2017; 2021), Poschke (2018) finds a positive relationship between mean firm size GDP per worker (not per capita), obtaining a positive elasticity of 0.72. He also finds the self-employment rate and GDP per worker are negatively related, and notes that this result is broadly related the positive mean firm size and economic development result. Poschke (2018) theorises that differing cross-country levels of technology primarily drive the relationship between mean firm size and economic development. Poschke (2018) investigates this by calibrating his model to US data and inferring cross-country technological levels from the differences between US data and data for other countries. He estimates that differences in technology explain half the positive elasticity between mean firm size and GDP per capita. Poschke (2018) also finds a negative relationship between correlated distortions and mean firm size, like Bento and Restuccia (2017). The author infers the level of these distortions across countries using the cross-country variation in distortions found by Hsieh and Klenow (2009). Poschke (2018) thus concludes that both distortions and technology impact the firm size distribution.

Several other papers imply a positive relationship between mean firm size and economic development but do not investigate it directly. Garcia-Santana and Ramos (2015) study the relationship between distortions and plant size in a cross-section of developing countries. They calculate the share of labour allocated to small plants across country, where small is defined as less than twenty employees. In a cross-country regression, they then regress GDP per worker on the share of labour in small firms and find a negative and significant association, meaning that countries with more labour devoted to smaller firms generally have a lower GDP per capita. Garcia-Santana and Ramos (2015) then regress the share of labour allocated to small firms on firm distortions, measured by the World Bank's Ease of Doing Business Index. They find a negative coefficient, meaning that countries with more distortions allocate more labour to smaller firms. Therefore, the implication is that the negative relationship between distortions and share of labour allocated to small firms can explain some of the relationship between economic development and that same share. Developing countries are thus argued to have more labour devoted to smaller firms because they have higher levels of distortions. This is a similar argument to that made by Bento and Restuccia (2017; 2021) and Poschke (2018).

Chen (2022) investigates the relationship between the prevalence of large firms and economic growth across countries. He uses two data sources- the World Bank Enterprise Surveys (WBES) for developing countries and the OECD's Structural Business Statistics for developed countries. The OECD data is a panel with each of the 33 countries containing 9 yearly observations between 2008 and 2017. In the WBES, Chen (2022) analyses 113 countries between 2006 and 2019. Most countries only have 1 or 2 years of data, and the total number of observations is 274. Chen (2022) measures the prevalence of large firms in a country by the thickness of the right tail of the firm size distribution. Right tail thickness is defined as the log of the ratio between the share of firms categorised as large and the share of firms categorised as not small, divided by the log of the ratio between the bin size definitions for large and small. In the OECD data a small firm is defined as one with less than 10 employees, while a large firm has more than 250. Using the WBES data, these thresholds are 5 and 100 respectively.

Chen (2022) finds a positive relationship between right-tail thickness and GDP per capita in both the OECD and WBES datasets, i.e. large firms are more prevalent in higher income countries. Chen (2022) finds a slightly smaller coefficient when not controlling for country fixed effects in the OECD, but the author focuses on the results when country fixed effects are controlled for in the OECD, as he argues that this shows that country-specific exogenous factors are not driving firm size differences.

Both Chen (2022) and Garcia-Santana and Ramos (2015) use the WBES. But this is not the ideal dataset in which to compare the mean firm size across country, since it excludes all informal firms, and many surveys exclude firms with fewer than than 5 employees. Since developing countries have more informal and small firms, the WBES does not provide an accurate estimate of the mean firm size in these countries, nor the share of small firms. Chen (2022) focuses on large firms, and Garcia-Santana and Ramos (2015) acknowledge these issues. In our analysis below we estimate measures of firm size that incorporate the smallest firms, as Bento and Restuccia (2017) recommend.

The research discussed thus far finds a positive relationship between various measures of firm size and GDP per capita or between the importance of large firms and GDP per capita. But two older papers find a negative relationship between mean firm size and economic development. Both have significant data limitations, which Bento and Restuccia (2017) argue is likely explain the negative relationship. Alfaro, Charlton and Kanczuk (2009) use the WorldBase dataset from Dun and Bradstreet to calculate mean firm size across countries and regress it on GDP per worker. The authors notes some significant limitations-as the WorldBase dataset is primarily used as an administrative database for contact details and basic information, it only contains formal firms, and has very low sample sizes for developing countries.

Bollard, Klenow and Li (2014) also report results that imply a negative relationship between mean firm size and economic development. They use UNIDO's industrial statistics database and find a positive relationship between firms per worker and value added per worker in manufacturing. Firms per worker is simply the inverse of the mean firm size, and thus their positive result implies a negative relationship if mean firm size was used. However, once again, UNIDO likely only surveys formal firms, and thus the exclusion of informal firms in developing countries likely biases this result.

2.2. Individual Approach

It is also possible to measure mean firm size using household surveys or population censuses that interview individuals and ask them about their employment, even when there is no question directly about firm size (which there almost never is in population censuses, the main source of data used in our analysis below). This insight comes from Salas-Fumás and Sanchez-Asin (2019), who aim to investigate the cross-country differences in the self-employment rate and the importance of social capital, using data from the International Labour Organisation (ILO, 2023a), which collects and harmonises data from labour force surveys in a wide range of countries over time and releases aggregated data on employment outcomes. Salas-Fumás and Sanchez-Asin (2019) use the number of employers, employees, and own-account workers. In undertaking their analysis, Salas-Fumás and Sanchez-Asin (2019) explicitly link mean firm size and the self-employed share. As far as we are aware, it is the only paper other than ours to do so. To better understand the similarities, consider the three measures that Salas-Fumás and Sanchez-Asin (2019) use in their analysis:

The Share of Self-Employed with Employees (Employers):

$$\mathrm{SE}_{\mathrm{it}} = \frac{\mathrm{employers}_{\mathrm{it}}}{\mathrm{employees}_{\mathrm{it}} + own{-}account_{it} + \mathrm{employers}_{\mathrm{it}}} \qquad (1)$$

The Share of Self-Employed without Employees (Own-Account Workers):[3]

$$\mathrm{SOA}_{\mathrm{it}} = \frac{own{-}account_{it}}{\mathrm{employees}_{\mathrm{it}} + own{-}account_{it} + \mathrm{employers}_{\mathrm{it}}} \qquad (2)$$

Average Span of Control (Entrepreneur Efficiency):

$$\mathrm{SoC}_{\mathrm{it}} = \frac{\mathrm{employees}_{\mathrm{it}}}{\mathrm{employers}_{\mathrm{it}}} \qquad (3)$$

The fourth measure is the all-self-employed share. It includes both employers and own-account workers.

The Share of all Self-Employed (Own-Account and Employers):

$$\mathrm{ASE}_{\mathrm{it}} = \frac{\mathrm{employer}_{\mathrm{it}} + own{-}account_{it}}{\mathrm{employees}_{\mathrm{it}} + own{-}account_{it} + \mathrm{employers}_{\mathrm{it}}} \qquad (4)$$

The relationships between the measures of firm size we construct and theirs are as follows. Our mean firm size including own-account workers is the inverse of the all-self-employed share. The sum of the inverse of the employer share and the own-account share is the mean firm size measure we use. The span of control variable is our mean firm size excluding own-account workers less 1. An important difference, however, is that Salas-Fumás and Sanchez-Asin (2019) use the number of employees in their measures, whereas we use persons engaged. This is because in developing countries many workers are not working for pay- they could be apprentices or unpaid family workers.

[3] For ease of reference, where Salas-Fumás and Sanchez-Asin (2019) refer to solo self-employed, this paper refers to own-account workers.

Salas-Fumás and Sanchez-Asin (2019) show that an increase in GDP per capita by 1% is associated with a 0.063 percentage point decrease in the self-employed worker share. They show that the own-account share drives this result because the own-account share decreases by 0.065 percentage points with every 1% increase in GDP per capita. The employer share is uncorrelated with GDP per capita. Therefore, the negative relationship between real GDP per capita and the all-self-employed share found by Salas-Fumás and Sanchez-Asin (2019) is driven by differences in the share of own-account workers across countries.

What implications would this have for our analysis below? It suggests that any positive relationship found between mean firm size and GDP per capita is driven by the inclusion of own-account workers.

Salas-Fumás and Sanchez-Asin (2019) were not the first to link self-employment shares and GDP per capita. La Porta and Schleifer (2014) found a negative relationship between the all-self-employed share and GDP per capita, using data from the World Development Indicators, which use ILO-modelled estimates.

Schoellman, Lu and Donovan (2017) investigate the relationship between the firm size distribution and labour market outcomes. They do so by harmonising labour force survey data from 17 different countries. Although they do not focus on the self-employment share, they use household survey data to explore mean firm size. This is similar to our approach. However, they calculate mean firm size from individuals' responses to questions about the size of the firm they work in, which is not available in many surveys and almost all population censuses, part of the reason they only include 17 countries in their analysis. The authors regress the employment share of each firm size bin on GDP per capita, using country-years. They find a negative association with the smallest category and a positive association with the largest, meaning that developed countries devote more labour towards large firms than less developed countries do, as in Garcia-Santana and Ramos (2015) and Chen (2022), who use firm data. A large share of the labour force working in large firms implies a larger mean firm size and thus implies a positive relationship between economic development and the mean firm size.

2.3. Summary

The relationship between cross-country mean firm size and economic development has been investigated using firm and individual level data by Bento and Restuccia (2017; 2021) and Poschke (2018), who find positive relationships with GDP per capita and per worker respectively. Other research implying this positive relationship has used WBES and OECD data. Firms in more developed countries are larger on average, which may seem obvious but was not found in early papers addressing this question (Alfaro, Charlton and Kanczuk,2009; Bollard, Klenow and Li, 2014). The literature on self-employment shares and development is also relevant because mean firm size including all own account workers as firms is the inverse of the all-self-employed share used by Salas-Fumás and Sanchez-Asin (2019). These authors, La Porta and Schleifer (2014) and others find a negative relationship between the all-self-employed share and GDP per capita. This implies a positive relationship between mean firm size and GDP per capita.

3. Data Description

In this section we document the sources of data used in our analysis. The first is the IPUMS harmonised population census data for 86 countries, and 263 census-year observations. The second is census aggregates from the UN Demographic and Social Statistics Department, which we use to add observations to IPUMS. The third is the ILO aggregated household survey data described above used by Salas-Fumas and Sanchez-Asin (2019).

3.1. IPUMS Description

The international version of the IPUMS dataset consists of harmonised micro-data from population censuses in various countries over multiple years. The collecting and harmonising are undertaken by the Minnesota Population Centre (MPC). We use 263 of the 553 population censuses and household surveys available from IPUMS. Not all the censuses can be used due to missing industry or status in employment variables.

We use the IPUMS micro-data to create a dataset including the country, year, and estimated mean firm size. The estimated mean firm size is calculated using information on employment activity from the various censuses and surveys. The formula used to calculate the primary estimate is found below:

$$mean\ firm\ size_1 = \frac{persons\ engaged + employers + own\ account}{employers + own\ account} \quad (5)$$

Persons engaged is a broad employment definition. The ILO defines it as workers who "could be in paid employment or in self-employment, including in less obvious forms of work, some of which are dealt with in detail in the resolution adopted by the 19th ICLS, such as unpaid family work, apprenticeship or non-market production" (ILO, 2023b). This category is constructed in IPUMS by including all paid employees, unpaid workers, apprentices, and family workers.

Separately, among self-employed workers we distinguish between employers and own-account workers. Own-account workers are those who work on their own and do not employ anyone else. This distinction is not important in the above equation, as both appear in the numerator and denominator. However, it does become important if one uses a measure which excludes own-account self-employed, as in equation 6 below. This definition does not consider own-account workers to be firms, whereas the first approach does.

$$mean\ firm\ size_2 = \frac{persons\ engaged + employers}{employers} \quad (6)$$

As noted above, not all censuses made available through IPUMS include the necessary questions to calculate mean firm size do not include information on industry or status in employment. Others do not include separate categories for employers and own-account workers. We cannot create the second measure of mean firm size without this last distinction. For the first measure, the IPUMS data contains censuses for 263 country-year combinations from 86 countries. This includes 48 observations in 21 countries in sub-Saharan Africa. For the second measure, the data contains 238 country-year observations covering 80 countries. This includes 46 observations from 20 countries in sub-Saharan Africa.

A similar data source is used to complement the IPUMS data. The United Nations' Demographic and Social Statistics Department (UNSD) harmonises census data from 1995 onwards from many countries into a single dataset. Although the micro-data is not available, total employment by status in employment and industry is provided. This dataset adds 58 extra country-year observations to the IPUMS dataset. The comparability across the datasets is high. The mean firm size estimates are extremely similar for most countries which appear in both datasets- the median ratio of firm sizes between country-year observations in each dataset is 1.004, and the mean ratio is 1.12.

3.2. ILO Data

We also use household survey data compiled by the ILO to calculate mean firm size using the same methods as described above. The ILO compiles data from national labour force surveys and processes them to produce harmonised and comparable aggregated employment estimates for many countries (ILO, 2023). As discussed above, Salas-Fumás and Sanchez-Asin (2019) used this data. The ILO data stretches from 1980 through to 2023, with 130 countries and 1159 observations. However, data before 2000 is quite limited.

3.3. Differences in the Mean Firm Size Measures

There are some important differences in the mean firm size measures created from the ILO and IPUMS/UN datasets. We discuss these in this section. The IPUMS/UN individual data will exclude publicly listed firms and private firms with dispersed ownership. This is because individuals would not report themselves as an employer when they own a small share in a listed or private company. Second, the number of country-year observations we could use would decline by 82% if we excluded public sector workers. We thus include them but are unable to include public sector firms in the denominator. These two issues both imply that the IPUMS/UN measure overestimates the true mean firm size. . When we compare mean firm size estimates in countries where we can exclude public sector workers, the estimates do not differ much,[4] meaning that the upward bias caused by the public sector issue is not large.

Two other issues should be mentioned that imply that the number of employers may not equal the number of firms. First, we might double count a firm due to including two employers who have a share in the same firm. We would, therefore, overestimate the number of firms, thereby underestimating the mean firm size. Secondly, an employer may own multiple firms causing us to underestimate the number of firms. The

[4] There are 63 country-year observations in which private and public sector could be distinguished. We compared the mean firm size measures including and excluding public sector workers by calculating the ratio of mean firm size with public sector to mean firm size without public sector. The mean ratio was 0.94, and the median ratio was 0.98, indicating that the measures were very similar. The correlation coefficient between the two was 0.99.

mean firm size would then be overestimated. This would, therefore, result in a smaller mean firm size estimate in the IPUMS/UN data as compared to the other two datasets.

The issues with the ILO data are similar to the IPUMS/UN data, as the data is created from labour force surveys of individuals and households, although the IPUMS/UN population census data estimates should have less sampling error, due to 5% or 10% stratified random samples drawn from the entire census that IPUMS provides. Measurement error in employment type in each source of data may be important and we discuss this below.

3.4. Comparison with Bento and Restuccia (2017) and Poschke (2018) Datasets

The IPUMS/UN and ILO datasets are the primary datasets used in this paper. However, an important part of the analysis below is a comparison with existing work on mean firm size across countries and time. Bento and Restuccia (2017) and Poschke (2018) undertook analysis similar to ours. Therefore, a more detailed comparison of the datasets used in those papers to the IPUMS/UN and ILO data used in this paper is of interest.

Bento and Restuccia (2017) constructed a dataset using firm censuses and nationally representative firm and household surveys from 132 countries between 2000 and 2012. They produce a single mean firm size estimate for the period between 2000 and 2012. The benefit of their approach, as argued by the authors, is twofold. First, their dataset is broadly representative of the world income distribution as it includes developing and developed countries. Second, they attempt to ensure that their measure includes the coverage of small firms and own-account workers. As they note, and as discussed above, data sources such as the World Bank Enterprise Surveys do not cover these firms. Bento and Restuccia (2017) define firm size as the number of persons engaged.

Kerr and McDougall (2020) raised the concern that the data compiled by Bento and Restuccia (2017) had large differences in the coverage of own-account workers. By comparing Bento and Restuccia (2017) firm survey sources to population censuses or household surveys, Kerr and McDougall (2020) show that the South African firm data source included 100% of own-account workers, the Rwandan data source 33%, and the Ghanaian data source 0.5%. The large variety of data sources used likely accounts for the large differences in coverage of the smallest firms. In countries where the surveys used have very good coverage of own-account workers (South Africa, for example), we expect their estimates to be accurate and similar to those in the IPUMS/UN and the ILO data. However, in countries where own-account worker coverage is low, Ghana for example, we expect their number to overestimate the true mean firm size.

Poschke (2018) uses both the Global Entrepreneurship Monitor (GEM) and the Bureau van Dijk's Amadeus dataset to conduct his analysis. The discussion here is limited to the GEM, because that was made publicly available, and Amadeus was not. The GEM surveys are nationally representative surveys conducted on individual entrepreneurs (GEM, 2023). Poschke (2018) uses data on all non-agricultural firms in the economy, rather than just manufacturing, which was the only sector covered by the Bento and Restuccia

(2017) data. Bento and Restuccia's (2021) paper, which includes firms in the services sector, is more comparable to the Poschke (2018) data from the GEM. Poschke (2018) produces one estimate per country for the 2000 to 2005 period.

An important characteristic of the GEM data, noted by Poschke (2018), is that it does not cover publicly listed firms with dispersed ownership. However, this problem may not be limited to publicly listed firms and would likely also exclude privately held firms with dispersed ownership, as an individual may not consider themselves an entrepreneur if they own a small stake in a private company. This is why Poschke (2018) also uses data from the Bureau van Dijk's Amadeus dataset. GEM likely does better in estimating mean firm size in developing countries where there are few large firms. However, the coverage of developing countries in the dataset is low, as for example there are only 2 African countries and 6 countries from South America and the Caribbean.

The lack of firms with dispersed ownership has the opposite effect in the GEM data compared to the IPUMS/UN data. In the GEM case, this results in an underestimation of the mean firm size, as the GEM is a survey of entrepreneurs (i.e. firm owners). Thus, it is not the case, as in census data, that the employees are present, and the owners are not. In the census data, these missing firms result in an overestimation of mean firm size, because we still count the employees of these firms. But in GEM the entire firm, both owner and employee, is missing. Therefore, if large firms are entirely missing in GEM, mean firm size will be underestimated.

3.5. Cross-Country Coverage

Having explained the limitations of the various sources of data we use to estimate mean firm size we now discuss the coverage of each data source. Table 1 below shows the percentage of countries by continent in each dataset, both at the country level and at the country-year level. The final column shows the true share of countries in each continent. The IPUMS/UN and ILO datasets come closest to a global country distribution. Bento and Restuccia (2017) and particularly Poschke (2018) have very low developing country representation- African countries make up only 14.18% and 4.55% of all countries in these papers respectively. African countries make up 30% of the countries in the ILO data, meaning that African countries are in fact slightly overrepresented in the ILO data. But low observations per African country mitigates this, with 14% of all country-year observations being African. 20.61% of all countries are African in IPUMS/UN. The IPUMS/UN and ILO data overrepresent the Americas in observation terms but underrepresent them in country terms. This is also the case in Bento and Restuccia (2017) and Poschke (2018). Bento and Restuccia (2017) and Poschke (2018) overrepresent Europe, while all datasets underrepresent Oceania.

Table 1: Country-Year Observation Continental Representation

	ILO (Obs)	IPUMS/UN (Obs)	ILO (Country)	IPUMS/UN (Country)	B&R	GEM	World
Africa	30.00%	20.61%	13.98%	17.48%	14.18%	4.55%	24.00%
Americas	19.23%	21.37%	33.39%	28.94%	18.66%	18.18%	22.80%
Asia	27.69%	25.95%	26.66%	23.50%	27.61%	20.45%	20.40%
Europe	16.92%	26.72%	24.85%	25.50%	32.84%	52.27%	21.20%
Oceania	6.15%	5.34%	1.12%	4.58%	6.72%	4.55%	11.60%
Total	130	131	1159	349	134	44	250

Source: ILO(2023), MPS (2023) and UNSD (2023), Bento and Restuccia (2017), Poschke (2018). (Obs) refers to individual country-year observations. (Country) refers to individual country representation. Included in 'country' are territories that would be regarded as overseas territories of other countries. The World Column calculated using data from the United Nations Statistics Division, and includes overseas territories.

3.6. Summary

In this section we have demonstrated that the IPUMS/UN and ILO data are likely to give the most accurate indicator of mean firm size when including own account workers as firms. We believe this to be the case, because these datasets ensure that we include the smallest firms in the firm size distribution. Further, we believe that the denominator in our measure is not too different to the true number of firms. This is because the number of large publicly owned firms or firms with dispersed ownership is small compared to the number of own-account workers. Despite difficulties in identifying the correct 'number of firms' in the denominator, the bias is likely to be smaller than the bias in Bento and Restuccia (2017;2021) and Poschke (2018). In Bento and Restuccia (2017;2021), missing own-account workers are likely to have a substantial effect on the final estimate, as many firms, by the authors' own definition, are missing, particularly in developing countries. The global coverage in the GEM data is so limited that its usefulness is much more limited than the IPUMS/UN and ILO datasets. Thus, we think there is a strong argument that the IPUMS/UN data and the ILO data, are the most accurate and useful datasets for measuring firm size when including the smallest firms, as Bento and Restuccia (2017) suggest is necessary. We now proceed with a description of mean firm size in the different data sources.

4. Estimates of Mean Firm size

In this section we use the datasets described above to show the average mean firm size and the median mean firm size. This entails estimating mean firm size for each country- period using the formula from equation (1) above and estimating the average and median of the estimated mean firm sizes across all country- period observations.

4.1. Mean Firm Size including Own-Account Workers

Table 2 below shows the average and median of all the mean firm size estimates from each of the datasets across all observations from the relevant dataset using equation (1) above, which includes own account workers as firms. Poschke (2018) includes all non-agricultural data, while the three other sources include only manufacturing. Three datasets have an average mean firm size of around 12- Bento and Restuccia (2017), Poschke (2018) using the GEM data and IPUMS/UN. The ILO is smaller than all other datasets, with an average mean firm size of 8.54. Given our priors that the Bento and Restuccia (2017) data under-captures the smallest firms compared to population census data, this is surprising. However, the median mean firm size in the Bento and Restuccia (2017) data is around 3 persons engaged higher, or around 50% higher, than the IPUMS/UN and ILO datasets, which are more similar to each other. Mean firm size in sub-Saharan Africa is much smaller in all datasets except for the GEM from Poschke (2018). This suggests a positive relationship between mean firm size and economic development.

Table 2: Descriptive Statistics from Datasets

Full Datasets	Average Mean Firm Size	Median Mean Firm Size	Number of Observations	Number of Countries
IPUMS/UN	11.59	6.27	349	131
Bento and Restuccia (2017)	11.90	8.92	134	134
GEM from Poschke (2018)	11.76	8.19	44	44
ILO	8.54	5.07	1159	130
Sub-Saharan Africa				
IPUMS	3.44	1.94	52	24
Bento & Restuccia (2017)	5.87	5.93	14	14
GEM from Poschke (2018)	13.18	13.18	2	2
ILO	3.67	2.04	137	35

Source: MPS (2023) & UNSD (2023), ILO (2023), Bento and Restuccia (2017), GEM in Poschke (2018).

Differences in the coverage of countries across the different datasets may drive some of the surprising results in Table 2. Therefore, Table 3 below reports the mean and median differences for common country-year observations between IPUMS and the other datasets. The percentage differences, absolute differences and absolute percentage differences are also included. Estimates from the Bento and Restuccia (2017) data and the GEM data do not have a specific year, as the authors used a single estimate drawn from several years. Therefore, we compare them to the IPUMS estimates that are closest to median year in the time period used by Bento and Restuccia (2017) and Poschke (2018). A maximum of 10 years is allowed between the IPUMS/UN and the other estimate for them to be compared.

Compared to the IPUMS data, the Bento and Restuccia (2017) data produces larger estimates of mean firm size and GEM smaller estimates. The ILO estimates are most similar to the IPUMS/UN estimates but are slightly smaller. The absolute differences shown in the second half of Table 3 show that the ILO is in fact much more similar to the IPUMS/UN data than Bento and Restuccia (2017) and the GEM data. When not using absolute differences, the similarity between the ILO and IPUMS estimates and the differences between IPUMS and both GEM and the Bento and Restuccia (2017) data was masked by the fact that not all estimates differ to the IPUMS/UN estimates in the same direction. The correlation coefficients are further evidence that the ILO data is very similar to the IPUMS/UN data and that the other two datasets produce very different estimates of mean firm size compared to IPUMS when comparing the same countries and time periods.

Table 3: Differences and Percentage Differences between Datasets

	Mean Difference	Median Difference	Mean Percentage Difference	Median Percentage Difference	Correlation Coefficient	Observations
B&R & IPUMS	-2.51	1.08	50.74	20.45	0.22	91
GEM & IPUMS	-1.43	-2.93	12.41	-23.91	0.12	36
ILO (5yr Avg) & IPUMS	-0.86	-0.14	-3.75	-3.59	0.83	124
Absolute Differences and Absolute Percentage Differences between Datasets						
	Mean Difference	Median Difference	Mean Percentage Difference	Median Percentage Difference	Correlation Coefficient	Observations
B&R & IPUMS	9.24	3.78	83.04	45.68	0.22	91
GEM & IPUMS	9.71	6.76	76.22	52.50	0.12	36
ILO (5yr Avg) & IPUMS	2.85	1.39	26.76	21.07	0.83	124

Source: MPS (2023) & UNSD (2023), ILO (2023), Bento and Restuccia (2017), GEM in Poschke (2018). In all cases the IPUMS mean firm size is subtracted from the other dataset. The 5yr average in the 3rd set of results calculated as an average of all observations within predefined 5 year periods starting 1960 and ending 2020.

The results we have discussed are in line with the predictions made in the data description section. We suggested above that Bento and Restuccia (2017) were likely to overestimate mean firm size that includes own account workers, since some of their data sources exclude many of these types of workers and our analysis reveals their data indeed produce the largest estimates of mean firm size.

We noted that the IPUMS/UN estimates may also be upward biased because firms with dispersed shareholder ownership may be excluded from the denominator of our mean firm size measure. However, the larger Bento and Restuccia (2017) estimates suggest Bento and Restuccia (2017,2021) overestimate mean firm size by more. Further, we expected Bento & Restuccia (2017) to overestimate mean firm size by more in countries where their coverage of own-account workers was low, but to be similar to our estimates where their coverage was good. This appears the case, as in South Africa where their coverage was good, they find a mean firm size of 9.6 which is similar to our estimate of 8 in 2007. In Ghana, where their coverage was poor, their estimate of 9.3 is much larger than our estimate in 2010 of 1.5. The smaller estimates from GEM are also unsurprising as we expected them to underestimate mean firm size because firms with dispersed shareholder ownership and their engaged persons are likely entirely excluded. The similarity between the ILO and IPUMS/UN estimates is also unsurprising, because of the similarity in census and household survey data.

4.2. Mean Firm Size excluding Own-Account Workers

The second measure of firm size we discussed above excludes own account workers. It may be incorrect to think of own account workers as firms. Unfortunately, we find our estimates of this measure of firm size to be implausibly large. In the IPUMS/UN data, the average mean firm size when excluding own-account workers is 58 persons engaged, compared to 10 when they are included. Some large outliers do skew the mean, but the median nevertheless increases from 6 to 26. In the ILO data, the average mean firm size increases from 8 to 28, and the median from 5 to 19.

A few examples show the unreliability of this measure. In India in 2004, excluding own-account workers increases the mean firm size from 3 to 43 in the IPUMS/UN data, whilst it increases from 2 to 38 in the ILO data. The Philippines is a more extreme example, as excluding own-account workers increases the mean firm size from 18 to 321 in the IPUMS/UN data in 2010 but from 5 to 30 in the ILO data from 2012. Bento and Restuccia's (2017) estimate is 13. In the ILO data, the mean firm size in Ethiopia in 2021 increases from 2 to 142, and in Japan in 2015 it increases from 30 to 135 in IPUMS and 25 to 136 in the ILO data around the same time. In general, excluding own-account workers results in implausibly large mean firm size estimates. Furthermore, excluding own-account workers increases mean firm size by different magnitudes across countries. For example, in Greece in 2001 the IPUMS/UN estimate only increases from 4 to 7, while in Botswana in 2019 the ILO estimate only increases from 2.4 to 2.8. These IPUMS/UN and ILO increases are different enough that the correlation coefficient between the two measures is only 0.5 in both the IPUMS/UN and ILO data, much lower than when including own account workers.

These results suggest that population censuses, and to a lesser extent the household surveys underlying the ILO aggregates, do not accurately distinguish between own account and employer self-employed. This is an important result for researchers wanting to compare own account and employer self-employed over time or across countries using population census data and links to research on the comparability of employment estimates from population census data (Gaddis et. al. 2023). This does not matter for our first firm size measure because both categories of self-employed are included. But it clearly does matter when we try to exclude own account workers and end up with too few employers. As a result of the lack of reliability of this second measure we do not use it in our main analysis, which we discuss in the next section.

5. Analysis

Having shown how we can estimate mean firm size using population census data and ILO aggregates derived from household surveys, which cover a long time period and many countries, and how these estimates compare to those from two other papers using alternative sources of data, we now provide evidence on the relationship between these measures of firm size and economic development. The basic specification is shown in Equation 7 below:

$$\ln(Mean\ Firm\ Size_{it}) = \beta_0 + \beta_1 \ln(Real\ GDP\ per\ capita_{it}) + \varepsilon_{it} \qquad (7)$$

$Mean\ Firm\ Size_{it}$ represents mean firm size defined by persons engaged in country i at time t. β_1 represents the elasticity of mean firm size with respect to the Real GDP per capita. We run this regression separately for the different sources of data on mean firm size and estimate the elasticity of mean firm size with respect to GDP per capita using OLS, as in Bento and Restuccia (2017) and Poschke (2018). Common samples across the datasets are also used to explore whether sample composition drives any coefficient differences. coefficient differences.

The IPUMS/UN and ILO datasets are panels. We can thus investigate whether changes in GDP per capita are correlated with changes in mean firm size, controlling for time-invariant country effects denoted by a_i. This specification is shown in Equation 8 below.

$$\ln(Mean\ Firm\ Size_{it}) = \beta_0 + \beta_1 \ln(Real\ GDP\ per\ capita_{it}) + a_i + \varepsilon_{it} \qquad (8)$$

5.1. Results

The results of estimating equation (7) are shown in Table 4 below. All coefficients are positive and statistically significant, showing that firm size is positively correlated with GDP per capita. Thus, our first important finding is that the positive elasticity found by Bento and Restuccia (2017) and Poschke (2018) holds up when using more comparable sources of data over a much longer time period. Both Alfaro, Charlton and Kanczuk (2009) and Bollard, Klenow and Li (2014) found negative elasticities, so our results imply this is likely an artefact of excluding small firms from developing countries.

However, there are some important differences in the magnitude of the coefficients across the different data sources. The ILO data has the largest coefficient - a 1% increase in the GDP per capita of a country is associated with a 0.68% increase in the mean firm size. This is followed by an elasticity of 0.52 in the IPUMS/UN data. The elasticity using the GEM data is slightly smaller at around 0.46. Bento and Restuccia's (2017) elasticity of 0.32 is substantially lower.[5] Thus, the elasticities when using the ILO and IPUMS/UN data are larger than those estimated in Bento and Restuccia (2017) and Poschke (2018).

Table 4 also shows that the ILO has a much larger R squared than the other three data sources, including IPUMS. Measurement error in the dependent variable raises the error variance and lowers the R squared, all else equal. IPUMS and the ILO data have mean firm size constructed in an identical manner, the only difference being that IPUMS is derived almost entirely from population censuses whereas ILO is from household surveys. Thus, the large difference in the R squared in the IPUMS and ILO regressions is suggestive evidence that the categorisation of employment types that underlies our mean firm size measure is more reliably captured in household surveys than population censuses, as one might expect.

[5] It should be noted that the coefficients found here differ slightly from those reported in Bento and Restuccia (2017) and Poschke (2018). With a similar regression of mean firm size on real GDP per capita, Bento and Restuccia (2017) find a coefficient of 0.29. This slightly different coefficient is likely due to a difference in sample composition. Bento and Restuccia (2017) supplement their GDP (PPP) per capita data from the PENN with other sources to ensure maximum coverage, which we do not do. Poschke (2018), when regressing the mean firm size on real GDP (PPP) per worker, finds a coefficient of 0.718. However, using an earlier working paper version of the Poschke (2018) paper, Poschke (2014), and direct correspondence with Poschke himself, Bento and Restuccia (2017) state that the data in Poschke (2018) implies an elasticity of mean firm size with respect to GDP per capita of 0.45. The use of real GDP (PPP) per worker by Poschke (2018) thus results in a larger coefficient.

Table 4: Full Sample Regressions

VARIABLES	(1) MFS IPUMS	(2) MFS B&R	(3) MFS GEM	(4) MFS ILO
Log Real GDPpc	0.517***	0.320***	0.463***	0.683***
	(0.0332)	(0.0417)	(0.117)	(0.0187)
Constant	-2.821***	-0.821**	-2.526**	-4.718***
	(0.305)	(0.400)	(1.206)	(0.178)
Observations	332	112	44	1.016
R-squared	0.364	0.351	0.231	0.543

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Data Source: MPS (2023) & UNSD (2023), Bento and Restuccia (2017), GEM via Poschke (2018), ILO(2023). MFS stands for Mean Firm Size. B&R - Bento and Restuccia (2017). GEM - Global Entrepreneurship Monitor from Poschke (2017). ILO - International Labour Organisation. Real GDP(PPP) per capita data taken from the PENN World Tables V10.0

To test whether the coefficients from the difference data sources are statistically different from each other, the IPUMS/UN dataset was combined into single datasets with each of the other datasets. A regression with a dummy variable indicating the other dataset and an interaction term with the log of GDP per capita was used. The statistical significance of the coefficient on the interaction term is a test of whether there are differences. Bento and Restuccia's (2017) coefficient is significantly smaller than the IPUMS/UN coefficient at the 1% level. The ILO coefficient is significantly larger than IPUMS/UN at the 1% level. The GEM coefficient is insignificantly different from the IPUMS/UN coefficient.

In Figure 1 below, the regression results from Table 4 are presented in graphical format. The smaller slope of the Bento and Restuccia (2017) line of best fit is clear. Figure 1C shows the lack of low- and middle-income countries in the GEM dataset. Some outliers in the IPUMS/UN dataset are made evident in Figure 1A.[6] In Figure 2, we show only the sub-Saharan African countries, together with the regression line estimated from all countries in each dataset. In both the IPUMS/UN and ILO data, sub-Saharan African countries with low GDP per capita have a mean firm size than is lower than predicted by the regression line.

[6] Most of the outliers originate in Eastern Europe- Ukraine, Belarus, Romania and Slovakia. The other clear outlier, Kazakhstan, is also relatively close geographically.

Figure 1: Mean Firm Size by GDP per capita over datasets

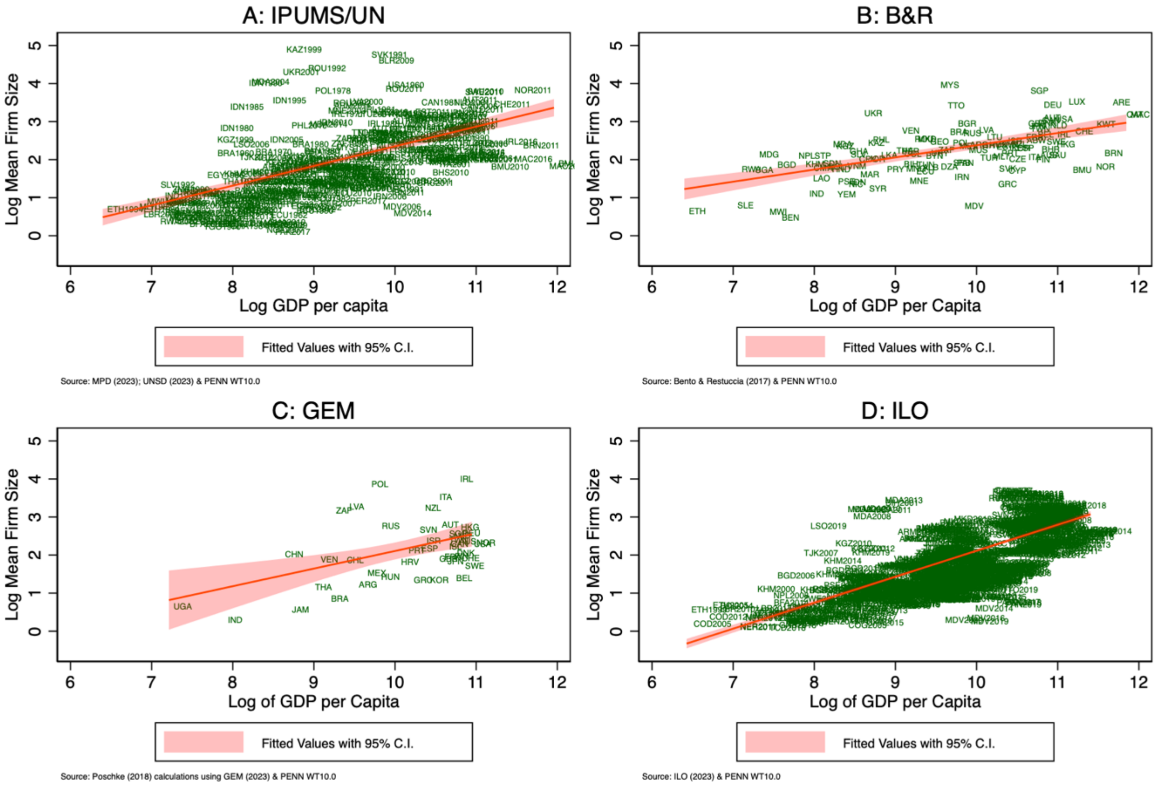
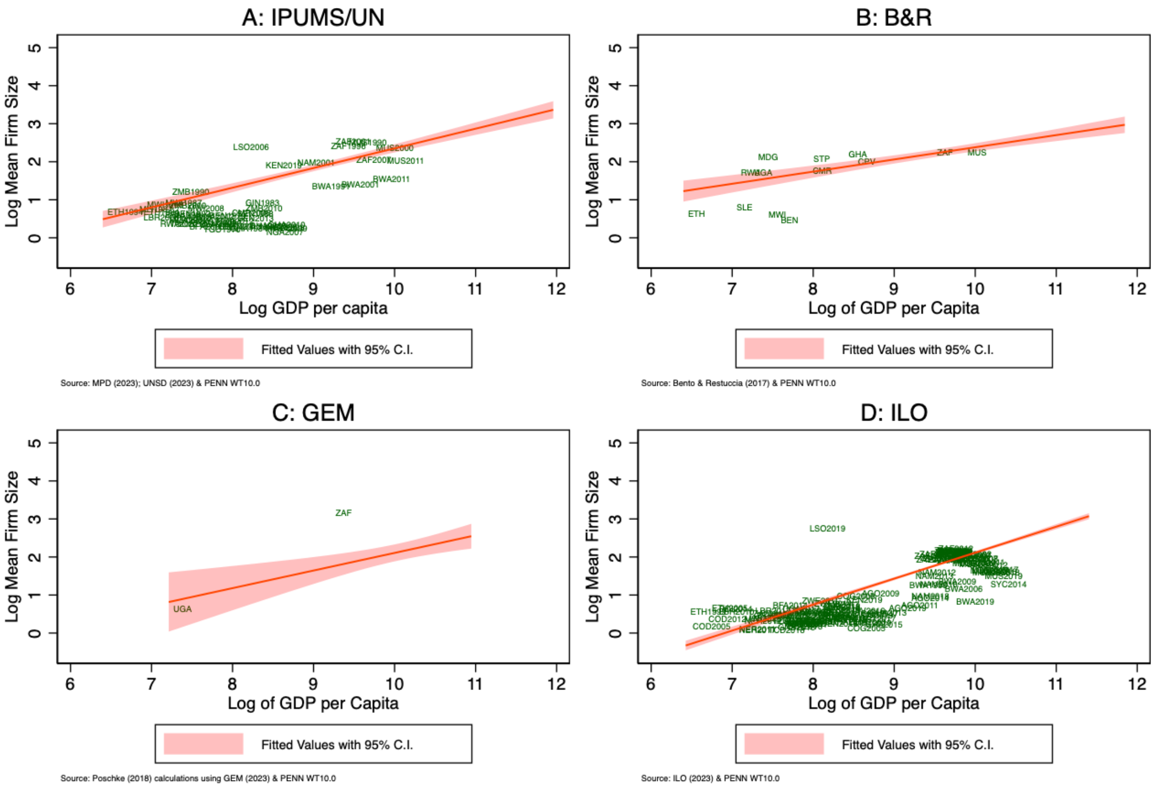


Figure 2: Mean Firm Size by GDPpc over datasets for SSA



To determine whether sample composition drives differences in the coefficients we run regressions comparing IPUMS to each of the other data sources with comparable samples. The small overlap of country- time period observations means the regression sample sizes are a lot smaller.[7] Comparing Bento and Restuccia (2017) and the IPUMS/UN data (columns 1 and 2), the IPUMS/UN data still exhibits a larger coefficient of 0.468 compared to Bento and Restuccia's (2017) 0.28. This difference is statistically significant at the 5% level. IPUMS/UN and GEM again have similar coefficients when using a comparable sample. The ILO coefficient of 0.66 is still larger than, but now closer to the IPUMS/UN coefficient of 0.58 and this difference is now not statistically significant. Although the samples are much smaller the results are broadly the same, implying that sample composition is not driving any of the differences in coefficient sizes between the data sources that we showed in Table 4.

Table 5: Comparable Sample Regressions

	(1)	(2)	(3)	(4)	(5)	(6)
Mean Firm Size	B&R	IPUMS/UN	GEM	IPUMS	ILO 5Y	IPUMS/UN 5Y
Comparison Type	To IPUMS	To B&R	To IPUMS	To GEM	To IPUMS	To ILO
Log Real GDPpc	0.282***	0.468***	0.525***	0.592***	0.660***	0.584***
	(0.0510)	(0.0732)	(0.125)	(0.0776)	(0.0541)	(0.0607)
Constant	-0.484	-2.331***	-3.166**	-3.565***	-4.504***	-3.670***
	(0.488)	(0.723)	(1.269)	(0.768)	(0.511)	(0.583)
Observations	89	89	36	36	103	103
R-squared	0.261	0.298	0.258	0.482	0.559	0.436

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Data Source: MPS (2023) & UNSD (2023), Bento and Restuccia (2017), GEM via Poschke (2017), ILO(2023). Column 5 and Column 6 are regressions run with half-decade and decade averages of each variable. MFS stands for Mean Firm Size. B&R – Bento and Restuccia (2017). GEM – Global Entrepreneurship Monitor from Poschke (2018). ILO – International Labour Organisation. Comparison Type refers to the comparison sample that is being used, i.e. TO IPUMS/UN means all observations present in the relevant dataset and IPUMS/UN are used in the regression. Real GDP (PPP) per capita data taken from the PENN World Tables V10.0

[7] Once again, we compare the IPUMS/UN estimates closest to the median year in the time period used by Bento and Restuccia (2017) and Poschke (2018). For the ILO and IPUMS/UN data, we use 5 year averages.

5.2. Explanations for the Difference in Coefficient Sizes across data sources

Measurement error is one possible explanation for differences in the coefficients between Bento and Restuccia (2017) and the IPUMS/UN dataset. Developing countries have larger shares of own account workers, and so Bento and Restuccia (2017) underestimating the number of own-account workers may bias the coefficient on mean firm size downwards, because the mean firm size in less developed countries are not as small as they should be relative to developed countries. This thus results in a 'flatter' regression line. We can investigate this by regressing the difference across datasets in the log mean firm size in each country by real GDP per capita. This is equivalent to plotting the log of the ratio between each country-year observation across the datasets. This is done for all datasets in the figures below, with the regression results presented in Table 6.

Figure 3: Log Ratio of Datasets to IPUMS/UN by GDPpc

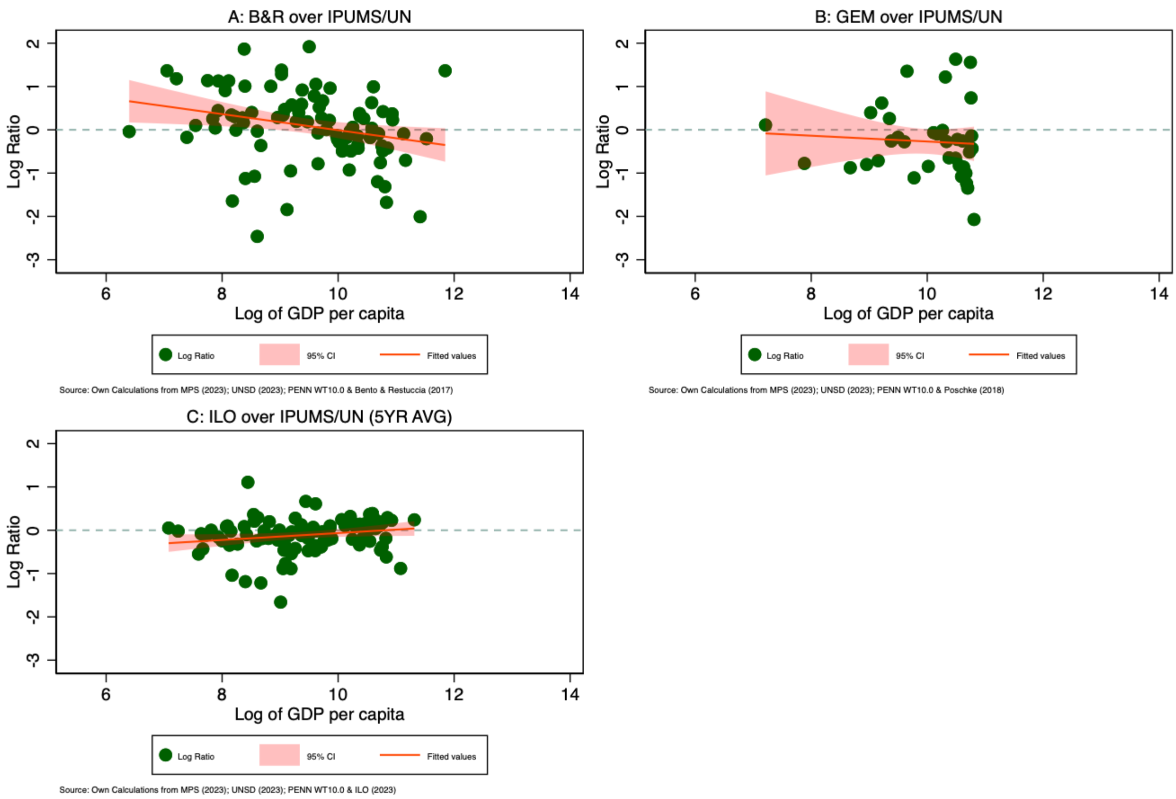


Table 6: MFS Difference to IPUMS/UN by GDPpc Regressions

VARIABLES	(1) B&R	(2) GEM	(3) ILO
Log Real GDPpc	-0.186**	-0.0670	0.0788**
	(0.0783)	(0.138)	(0.0377)
Constant	1.847**	0.400	-0.854**
	(0.757)	(1.324)	(0.366)
Observations	89	36	103
R-squared	0.067	0.005	0.037

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Source: MPS (2023) & UNSD (2023), GEM used in Poschke (2018), ILO(2023). ILO regression done using 5 year averages. Real GDP (PPP) per capita taken from PENN World Tables V10.0. Dependent Variables are all the log ratio of the mean in firm size in the datasets to the IPUMS/UN dataset.

First, we consider the IPUMS/UN and Bento and Restuccia (2017) data. There is a negative and statistically significant relationship between the log ratio of the mean firm sizes and log GDP per capita. Most of the ratios are positive, indicating that most of the Bento and Restuccia (2017) mean firm size estimates are larger, as we noted above. However, as the log of GDP per capita increases, this log ratio decreases. This indicates that the difference between the log mean firm size estimates decreases with log GDP per capita. Therefore, the Bento and Restuccia (2017) estimates are more similar to IPUMS the higher is real GDP per capita. In less developed countries, estimates from Bento and Restuccia (2017) are larger than those in the IPUMS/UN dataset.

No relationship between the log ratio of mean firm size and log GDP per capita is found for the GEM dataset. However, the figure confirms that most GEM estimates are smaller than the IPUMS/UN estimates across the GDP per capita distribution. The GEM estimates are constructed using non-agricultural data and this may result in smaller firm size estimates in GEM. There is a positive and significant relationship between the log ratio of the ILO mean firm size to the IPUMS/UN estimates and GDP per capita. The ILO data is generally smaller than the IPUMS/UN data. This means that the ILO data is smaller by less in developed countries than in developing countries. However, this positive relationship is also driven by some larger ILO estimates in developed countries. These differences explain the larger coefficient on real GDP per capita in the ILO data, as they induce a steeper regression curve.

5.3. Robustness Checks

The robustness of our main results can be checked with three alternative ways of using the data. First, we average across decade instead of using individual year observations for mean firm size. For example, in the ILO data, we may have 10 observations between 2010 and 2020. This is rarer in the IPUMS/UN data. Nevertheless, for example, we have two observations for Australia between 2010 and 2020 and for Canada between 2000 and 2010. We can, therefore, average across decades and then conduct the same analysis. We do this for two reasons. Averaging may reduce measurement error in individual years and reduce any attenuation bias. It also enables us to conduct analysis in a manner more similar to Bento and Restuccia (2017;2021) and Poschke (2018).

Our second check is to change the sample from the manufacturing sector to the non-agricultural sector in the IPUMS/UN and ILO data and add in services to the Bento and Restuccia (2017) data. The GEM data from Poschke is all sectors except agriculture, and thus adjusting the other three data sources to include the same or similar sectors increases comparability. Third, because of the lack of firms with dispersed ownership in the population census and ILO data we adjust these data using WBES data to include firms with dispersed ownership.

The detailed results of all three robustness checks are not reported here but are included in Appendix A, B and C. Averaging over decades does not substantively change our results, although the IPUMS/UN and ILO coefficients are slightly more similar to each other. When using all sectors except agriculture the estimates from the other data sources become more similar to GEM, implying that the lower coefficient in GEM found in our main results is due to it including all non-agriculture and not just manufacturing.

Our third robustness check is to adjust the IPUMS/UN and ILO estimates using the WBES. As discussed above, we believe the IPUMS/UN and ILO data likely overestimate mean firm size because these data sources exclude dispersed shareholder firms and partnerships from the denominator, since no shareholder (and possibly partner) is likely to report themselves as an employer. We therefore estimate the number of these types of firms using the WBES for a limited number of countries and use them to adjust our measure of mean firm size. Several criteria are used to identify these firms in the WBES. First, using the firm category variable, firms who are either publicly listed or privately owned with traded shares can be identified. For the private firms, extra criteria must be met. The firm must either be a part of a larger firm or have an owner that does not own the whole firm or have more than 90% foreign ownership or where the owner is not the top manager. For partnerships, the firm must fall into the category of “Partnership or Limited Partnership” and its largest owner must own less than 100%. We then adjust our mean firm size estimate by adding these firms to the denominator. This adjustment is likely not perfect because some WBES exclude some regions. Thus, where this this is the case, such as in Ghana and South Africa, we increase the denominator by too little and thus still overestimate mean firm size.

Because our adjustment increases the denominator, the WBES-adjusted estimates of mean firm size in both the IPUMS/UN and ILO datasets decline by 2.64 and 0.63 persons engaged in the IPUMS/UN and ILO data respectively, or by 21% and 11%. The median differences are smaller - 0.32 and 0.05 persons engaged- suggesting that a few large differences skew the mean. When only adjusting for shareholder firms, the median difference is only 0.12 and 0.02 persons engaged in the IPUMS/UN and ILO data respectively. Table C3 shows these differences are positively related to GDP per capita. This indicates that adjustments matter more in more developed countries. This is unsurprising as developed countries are likely to have more listed and large firms with dispersed ownership.

Table C4 shows that these adjustments make only a small difference to the estimated relationship between mean firm size and GDP per capita. The coefficient on log GDP per capita drops by 0.09 in the IPUMS/UN data and 0.06 in the ILO data when both shareholder firms and partnerships are adjusted for. When only shareholder firms are adjusted for, both the IPUMS/UN and ILO coefficients only decrease by 0.05. Thus, although the IPUMS/UN and ILO coefficients in the main analysis are upwardly biased, the impact on our estimates is small.

5.4. Fixed Effects Regression Results

Thus far we have shown that using more comparable data than previous research still results in a strong and positive elasticity of mean firm size with respect to GDP per capita, and that these results hold up after correcting for various potential biases. One concern is that we have treated the data as one large cross section. But the IPUMS/UN and ILO data each contain multiple observations for around 120 countries, meaning we have an unbalanced panel and can use fixed effects regressions to control for time invariant country characteristics. The results from the fixed effects regression are presented in Table 7 below. The coefficients on GDP per capita decline very substantially and are not statistically significant. GDP per capita within countries is strongly correlated over time and likely to be subject to measurement error, which can lead to measurement error bias and attenuation. This is one possible explanation for our results. But it is also possible that in many countries that have grown in recent times firm size has not increased, which would be an interesting complication for Jensen's (2022) set of "macro-development stylised facts". This is a perhaps surprising result. We found elasticities using the more reliable and comparable mean firm size measures that were much larger than those using less comparable data, and yet our elasticities go to zero once we control for time invariant country differences.

We noted above that Chen used the OECD Structural Business Statistics to investigate the relationship between GDP per capita and the thickness of the right tail of the firm size distribution, a related question to the one we ask. The OECD data also had multiple observations per country, and Chen (2022) used the fixed effects estimator to control for time invariant country characteristics, finding that these were not very different from his OLS results. Our result is quite different. This may be because the OECD data used by Chen (2022) excludes many developing countries.

Table 7: Fixed Effects Regressions

VARIABLES	(1) IPUMS/UN	(2) ILO
Log Real GDPpc	-0.0110	0.0434
	(0.0962)	(0.0636)
Constant	1.991**	1.387**
	(0.876)	(0.607)
Observations	332	1016
R-squared	0.000	0.004
Number of countries	121	118

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Data Source: MPS (2023) & UNSD (2023), ILO(2023). ILO – International Labour Organisation. Real GDP (PPP) per capita data taken from the PENN World Tables V10.

6. Conclusion

The recent literature on the importance of distortions in explaining differing levels of economic development generates predictions of a positive relationship between mean firm size and GDP per capita. Earlier work examining this relationship initially found a negative relationship (Alfaro, Charlton and Kanczuk ,2009; Bollard, Klenow and Li, 2014). Bento and Restuccia (2017) argued that this prior work had used data that was unsuitable by only including formal sector firms, which excluded most firms in developing countries. Bento and Restuccia (2017) obtained firm census and survey data for 134 countries to estimate what they argued were comparable measures of mean firm size. But Kerr and McDougall (2020) argued that these were not as comparable as Bento and Restuccia (2017) suggested, because at least some data sources had very different coverage of the smallest firms, particularly own account workers.

In this paper we have used the insight from Salas-Fumás and Sanchez-Asin (2019) that household survey or population census data can be used to estimate mean firm size even without a question on the size of the firm. We used ILO household survey data aggregates and population census microdata from IPUMS supplemented with UN population census aggregates to estimate mean firm size. We argued that these measures are more comparable than those estimated by Bento and Restuccia (2017) and are superior to those from Poschke (2018) because they cover a much larger number of countries. We found that mean firm size is smaller in both these data sources than in Bento and Restuccia (2017) and Poschke (2018). We also showed that this is not driven by differences in the samples of countries in IPUMS or ILO compared to Bento and Restuccia (2017) and Poschke (2018).

We used the mean firm size estimates derived from in the more comparable ILO and IPUMS data to confirm that there is a positive relationship between mean firm size and GDP per capita, or what Jensen (2022) calls a macro-development stylised fact. We found that this cross-sectional relationship is stronger than found by Bento and Restuccia (2017) or Poschke (2018). We showed that part of the explanation for this finding is that mean firm size estimates in Bento and Restuccia (2017) or Poschke (2018) are systematically larger than those derived from ILO and IPUMS in lower income countries.

The IPUMS and ILO data have the additional advantage of multiple observations per country for different time periods. We could thus estimate regressions controlling for time invariant country characteristics, finding that the elasticity of mean firm size with respect to GDP per capita is very close to zero and not statistically significant in both data sources. It is possible that this result is driven by measurement error in GDP per capita, a common concern when using panel data. Our results suggest that within country growth may not result be correlated with growth in the mean size of firms, even though this relationship is clear in the cross section. These results suggest that the experiences of rich countries, which clearly have experienced both growth and increases in the size of firms, may not have been replicated in low- and middle-income countries, and thus calls into question one of Jensen's (2022) macro-development stylised facts.

Reference List

Alfaro, L. Charlton, A. and Kanczuk, F. (2009) 'Plant size distribution and cross-country income differences'. *NBER International seminar on macroeconomics*. 5(1), pp. 243-272. Chicago, IL: The University of Chicago Press. Available at: https://doi.org/10.1086/596010.

Bento, P. and Restuccia, D. (2017) 'Misallocation, Establishment Size, and Productivity', *American Economic Journal: Macroeconomics*, 9(3), pp. 267–303. Available at: https://doi.org/10.1257/mac.20150281.

Bento, P. and Restuccia, D. (2021) 'On average establishment size across sectors and countries', *Journal of Monetary Economics*, 117, pp. 220–242. Available at: https://doi.org/10.1016/j.jmoneco.2020.01.001.

Bollard, A., Klenow, P.J. and Li, H. (2014). Entry Costs Rise with Development. Working Paper No. 518. Stanford Centre for International Development. Stanford, California. Available at: https://kingcenter.stanford.edu/sites/g/files/sbiybj16611/files/media/file/518wp_0.pdf. (Accessed: 13 November 2023).

Chen, Z. (2022) *Economic Growth and the Rise of Large Firms*. Job Market Paper. Princeton University. Available at: https://zhangc-econ.github.io/files/TailGrowth.pdf (Accessed: 13 February 2023).

Feenstra, R., Inklaar, R. and Timmer, M. 'PENN World Tables V10.0'. Dataset. University of Groningen. Groningen. Available at: https://www.rug.nl/ggdc/productivity/pwt/pwt-releases/pwt100. (Accessed: 31 May 2023).

Fonseca, R. and Utrero, N. (2017) 'Financial markets and firm size: The role of employment protection laws and barriers to entrepreneurship', *Applied Economics*, 49(26), pp. 2515–2531. Available at: https://doi.org/10.1080/00036846.2016.1243209.

García-Santana, M. and Ramos, R. (2015) 'Distortions and the size distribution of plants: evidence from cross-country data', *SERIEs*, 6(3), pp. 279–312. Available at: https://doi.org/10.1007/s13209-015-0129-y.

Global Entrepreneurship Monitor Consortium. 'Global Entrepreneurship Monitor'. Dataset. Available at: https://www.gemconsortium.org/data/sets?id=aps. (Accessed: 31 May 2023).

Hsieh, C.-T. and Klenow, P.J. (2009) 'Misallocation and Manufacturing TFP in China and India *', *Quarterly Journal of Economics*, 124(4), pp. 1403–1448. Available at: https://doi.org/10.1162/qjec.2009.124.4.1403.

International Labour Organization (2023a) 'Labour Force Statistics'. Dataset. Geneva. Available at: https://ilostat.ilo.org/data/ (Accessed: 31 May 2023).

International Labour Organisation (2023b) `ILOSTAT Database Description: Labour Force Statistics. Available: https://ilostat.ilo.org/resources/concepts-and-definitions/description-labour-force-statistics/. (Accessed: 20 November 2023).

Jensen, A. (2022). Employment structure and the rise of the modern tax system. *American Economic Review*, 112(1), pp. 213-34. Available at: https://doi.org/10.1257/aer.20191528

Kerr, A. and McDougall, B. (2020) *What is a firm census in a developing country? An answer from Ghana*. SALDRU Working Paper No. 262. Cape Town. Available at: https://opensaldru.uct.ac.za/bitstream/handle/11090/978/2020_262_Saldruwp.pdf?sequence=1. (Accessed: 31 May 2023).

Kumar, K.B., Rajan, R.G. and Zingales, L. (2001) *What Determines Firm Size?* 496. Chicago. Available at: https://ssrn.com/abstract=170349 (Accessed: 13 February 2023).

Minnesota Population Centre (2023) 'Integrated Public Use Microdata Series, International'. Minneapolis. Available at: https://doi.org/10.18128/D020.V7.3 (Accessed: 31 May 2023).

La Porta, R. and Shleifer, A. (2014) 'Informality and Development', *Journal of Economic Perspectives*, 28(3), pp. 109–126. Available at: https://doi.org/10.1257/jep.28.3.109.

Poschke, M. (2018) 'The Firm Size Distribution across Countries and Skill-Biased Change in Entrepreneurial Technology', *American Economic Journal: Macroeconomics*, 10(3), pp. 1–41. Available at: https://doi.org/10.1257/mac.20140181.

Restuccia, D. and Rogerson, R. 2008. 'Policy distortions and aggregate productivity with heterogeneous establishments', *Review of Economic Dynamics*, 11(4), pp.707-720. Available at: https://doi.org/10.1016/j.red.2008.05.002

Salas-Fumás, V. and Sanchez-Asin, J.J. (2019) 'Social Capital, Trust and Entrepreneurial Productivity', *The Manchester School*, 87(5), pp. 607–639. Available at: https://doi.org/10.1111/manc.12259.

Schoellman, T., Lu, J. and Donovan, K. (2017) *Firm Size Distributions and Cross-Country Labor Market Outcomes*. Society for Economic Dynamics. Meeting Papers No. 1571. Available at: https://ideas.repec.org/p/red/sed017/1571.html (Accessed: 31 May 2023).

United Nations Statistics Division (2023) 'Population Censuses' Dataset'. Dataset. New York. Available at: http://data.un.org/Data.aspx?d=POP&f=tableCode:323 (Accessed: 31 May 2023).

Alfaro, L. Charlton, A. and Kanczuk, F. (2009) 'Plant size distribution and cross-country income differences'. *NBER International seminar on macroeconomics*. 5(1), pp. 243-272. Chicago, IL: The University of Chicago Press. Available at: https://doi.org/10.1086/596010.

Bento, P. and Restuccia, D. (2017) 'Misallocation, Establishment Size, and Productivity', *American Economic Journal: Macroeconomics*, 9(3), pp. 267–303. Available at: https://doi.org/10.1257/mac.20150281.

Bento, P. and Restuccia, D. (2021) 'On average establishment size across sectors and countries', *Journal of Monetary Economics*, 117, pp. 220–242. Available at: https://doi.org/10.1016/j.jmoneco.2020.01.001.

Bollard, A., Klenow, P.J. and Li, H. (2014). Entry Costs Rise with Development. Working Paper No. 518. Stanford Centre for International Development. Stanford, California. Available at: https://kingcenter.stanford.edu/sites/g/files/sbiybj16611/files/media/file/518wp_0.pdf. (Accessed: 13 November 2023).

Chen, Z. (2022) *Economic Growth and the Rise of Large Firms*. Job Market Paper. Princeton University. Available at: https://zhangc-econ.github.io/files/TailGrowth.pdf (Accessed: 13 February 2023).

Feenstra, R., Inklaar, R. and Timmer, M. 'PENN World Tables V10.0'. Dataset. University of Groningen. Groningen. Available at: https://www.rug.nl/ggdc/productivity/pwt/pwt-releases/pwt100. (Accessed: 31 May 2023).

Fonseca, R. and Utrero, N. (2017) 'Financial markets and firm size: The role of employment protection laws and barriers to entrepreneurship', *Applied Economics*, 49(26), pp. 2515–2531. Available at: https://doi.org/10.1080/00036846.2016.1243209.

Gaddis, I., Oseni, G., Palacios-Lopez, A., & Pieters, J. (2023). Who is employed? Evidence from sub-Saharan Africa on redefining employment. *Journal of African Economies*, 32(2), 151-174.

García-Santana, M. and Ramos, R. (2015) 'Distortions and the size distribution of plants: evidence from cross-country data', *SERIEs*, 6(3), pp. 279–312. Available at: https://doi.org/10.1007/s13209-015-0129-y.

Global Entrepreneurship Monitor Consortium. 'Global Entrepreneurship Monitor'. Dataset. Available at: https://www.gemconsortium.org/data/sets?id=aps. (Accessed: 31 May 2023).

Hsieh, C.-T. and Klenow, P.J. (2009) 'Misallocation and Manufacturing TFP in China and India *', *Quarterly Journal of Economics*, 124(4), pp. 1403–1448. Available at: https://doi.org/10.1162/qjec.2009.124.4.1403.

International Labour Organization (2023a) 'Labour Force Statistics'. Dataset. Geneva. Available at: https://ilostat.ilo.org/data/ (Accessed: 31 May 2023).

International Labour Organisation (2023b) `ILOSTAT Database Description: Labour Force Statistics. Available: https://ilostat.ilo.org/resources/concepts-and-definitions/description-labour-force-statistics/. (Accessed: 20 November 2023).

Jensen, A. (2022). Employment structure and the rise of the modern tax system. *American Economic Review*, 112(1), pp. 213-34. Available at: https://doi.org/10.1257/aer.20191528

Kerr, A. and McDougall, B. (2020) *What is a firm census in a developing country? An answer from Ghana*. SALDRU Working Paper No. 262. Cape Town. Available at: https://opensaldru.uct.ac.za/bitstream/handle/11090/978/2020_262_Saldruwp.pdf?sequence=1 . (Accessed: 31 May 2023).

Kumar, K.B., Rajan, R.G. and Zingales, L. (2001) *What Determines Firm Size?* 496. Chicago. Available at: https://ssrn.com/abstract=170349 (Accessed: 13 February 2023).

Minnesota Population Centre (2023) 'Integrated Public Use Microdata Series, International'. Minneapolis. Available at: https://doi.org/10.18128/D020.V7.3 (Accessed: 31 May 2023).

La Porta, R. and Shleifer, A. (2014) 'Informality and Development', *Journal of Economic Perspectives*, 28(3), pp. 109–126. Available at: https://doi.org/10.1257/jep.28.3.109.

Poschke, M. (2018) 'The Firm Size Distribution across Countries and Skill-Biased Change in Entrepreneurial Technology', *American Economic Journal: Macroeconomics*, 10(3), pp. 1–41. Available at: https://doi.org/10.1257/mac.20140181.

Restuccia, D. and Rogerson, R. 2008. 'Policy distortions and aggregate productivity with heterogeneous establishments', *Review of Economic Dynamics*, 11(4), pp.707-720. Available at: https://doi.org/10.1016/j.red.2008.05.002

Salas-Fumás, V. and Sanchez-Asin, J.J. (2019) 'Social Capital, Trust and Entrepreneurial Productivity', *The Manchester School*, 87(5), pp. 607–639. Available at: https://doi.org/10.1111/manc.12259.

Schoellman, T., Lu, J. and Donovan, K. (2017) *Firm Size Distributions and Cross-Country Labor Market Outcomes*. Society for Economic Dynamics. Meeting Papers No. 1571. Available at: https://ideas.repec.org/p/red/sed017/1571.html (Accessed: 31 May 2023).

United Nations Statistics Division (2023) 'Population Censuses' Dataset'. Dataset. New York. Available at: http://data.un.org/Data.aspx?d=POP&f=tableCode:323 (Accessed: 31 May 2023).

Appendix A : Decade Analysis

Table A1: IPUMS/UN Decade Regressions

Mean Firm Size	(1) IPUMS/ UN	(2) IPUMS/ UN	(3) IPUMS/ UN	(4) IPUMS/ UN	(5) IPUMS/ UN	(6) IPUMS/ UN	(7) IPUMS/ UN
Decade	1960's	1970's	1980's	1990's	2000's	2010's	All
Log GDP pc	1.096***	0.908***	0.658***	0.585***	0.525***	0.642***	0.562***
	(0.220)	(0.124)	(0.0864)	(0.0650)	(0.0571)	(0.0741)	(0.0327)
Constant	-7.792***	-6.180***	-3.910***	-3.224***	-2.983***	-4.209***	-3.249***
	(1.897)	(1.028)	(0.771)	(0.547)	(0.530)	(0.722)	(0.297)
Observations	14	28	35	53	92	81	303
R-squared	0.695	0.582	0.558	0.349	0.375	0.451	0.407

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Column 7 includes all decades in the regression. Data Source: MPS (2023) & UN (2023). GDP(PPP) per capita data taken from the PENN World Tables V10.0. Mean Firm Size is averaged across the decade for where there are >1 observations. GDP per capita is averaged across the whole decade.

Table A2: ILO Decade Regressions

Mean Firm Size	(1) ILO	(2) ILO	(3) ILO	(4) ILO
Decade	1990's	2000's	2010's	All
Log Real GDP	0.514***	0.541***	0.619***	0.597***
	(0.130)	(0.0613)	(0.0408)	(0.0344)
Constant	-3.201**	-3.251***	-4.169***	-3.877***
	(1.156)	(0.564)	(0.367)	(0.313)
Observations	11	71	116	200
R-squared	0.764	0.469	0.590	0.546

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 Column 4 contains all decades. Data Source: ILO(2023). GDP(PPP) per capita data taken from the PENN World Tables V10.0

Table A3: Comparable Regressions – All Decades

	(1)	(2)
Mean Firm Size	IPUMS/UN	ILO
Log Real GDP	0.570***	0.622***
	(0.0572)	(0.0491)
Constant	-3.553***	-4.165***
	(0.551)	(0.465)
Observations	122	122
R-squared	0.433	0.540

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Data Source: MPS (2023) & UN (2023) and ILO(2023). GDP(PPP) per capita data taken from the PENN World Tables V10.0

Table A4: Fixed Effects Regressions: All Decades

	(1)	(2)
Mean Firm Size	IPUMS/UN	ILO
Log Real GDPpc	-0.0363	0.0299
	(0.0822)	(0.0676)
Constant	2.207***	1.299**
	(0.750)	(0.618)
Observations	303	200
R-squared	0.002	0.004
Number of countries	121	118

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Data Source: MPS (2023) & UN (2023) and ILO(2023). GDP(PPP) per capita data taken from the PENN World Tables V10.0

Appendix B: Non-Agricultural Analysis

Table B1: Full Sample Regressions – Non-Agricultural

	(1)	(2)	(3)	(4)
Comparison Type	Full	Full	Full	Full
VARIABLES	MFS IPUMS	MFS B&R	MFS GEM	MFS ILO
Log Real GDPpc	0.403***	0.314***	0.473***	0.455***
	(0.0211)	(0.0369)	(0.115)	(0.0117)
Constant	-1.969***	-1.741***	-2.617**	-2.748***
	(0.199)	(0.355)	(1.190)	(0.115)
Observations	365	91	44	1158
R-squared	0.377	0.450	0.239	0.540

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Data Source: MPS (2023) & UN (2023), Bento and Restuccia (2017), GEM via Poschke (2018), ILO(2023). MFS stands for Mean Firm Size. B&R - Bento and Restuccia (2017). GEM - Global Entrepreneurship Monitor from Poschke (2018). ILO - International Labour Organisation. GDP(PPP) per capita data taken from the PENN World Tables V10.0

Table B2: Comparable Sample Regressions – Non-Agricultural

	(5)	(6)	(7)	(8)	(11)	(12)
Mean Firm Size	B&R	IPUMS	GEM	IPUMS/UN	ILO 5YR AVG	IPUMS/UN 5YR AVG
Comparison Type	To IPUMS	To B&R	To IPUMS	To GEM	To IPUMS	To ILO
Log Real GDPpc	0.292***	0.427***	0.529***	0.418***	0.461***	0.440***
	(0.0420)	(0.0530)	(0.120)	(0.0616)	(0.0324)	(0.0404)
Constant	-1.548***	-2.268***	-3.213**	-2.210***	-2.820***	-2.474***
	(0.403)	(0.530)	(1.229)	(0.621)	(0.315)	(0.399)
Observations	75	75	37	37	116	116
R-squared	0.400	0.453	0.267	0.420	0.641	0.466

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Data Source: MPS (2023) & UN (2023), Bento and Restuccia (2017), GEM via Poschke (2018), ILO(2023). Column 11 and Column 12 are regressions run with half-decade and decade averages of each variable. B&R - Bento and Restuccia (2017). GEM - Global Entrepreneurship Monitor from Poschke (2018). ILO - International Labour Organisation. Comparison Type refers to the comparison sample that is being used, i.e. TO IPUMS means all observations present in the relevant dataset and IPUMS are used in the regression. GDP per capita data taken from the PENN World Tables V10.0

Table B3: Fixed Effects Regressions – Non-Agricultural

VARIABLES	(1) MFS IPUMS/UN	(3) MFS ILO
Log Real GDP	-0.0522	0.0548
	(0.0810)	(0.0361)
Constant	2.164***	1.105***
	(0.735)	(0.347)
Observations	365	1158
R-squared	0.005	0.017
Number of country	124	133

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Data Source: MPS (2023) & UN (2023), Bento and Restuccia (2017), GEM via Poschke (2017), ILO(2023). GDP per capita data taken from the PENN World Tables V10.0.

Appendix C: WBES Adjusted Results

Table C1: Mean Firm Size with WBES Adjustments

Dataset	IPUMS/UNIP			ILO		
Adjustment	Mean	Median	Countries	Mean	Median	Countries
Original	12.43	5.41	101	5.67	3.35	106
Dispersed Shareholders	10.46	5.25	101	5.25	3.33	106
DS + Partnerships	9.79	5.11	101	5.03	3.32	106

Data Source: MPS (2023) & UNSD (2023) and ILO (2023). Dispersed shareholders: All firms categorised as having dispersed shareholder ownership added to the denominator. DS + Partnerships: All firms having dispersed shareholder ownership plus firms categorised as partnerships added to the denominator.

Table C2: Mean Firm Size with WBES Adjustments

Dataset	IPUMS/UNIP		ILO	
Adjustment	Mean	Median	Mean	Median
Dispersed Shareholders	1.97	0.12	0.42	0.02
DS + Partnerships	2.64	0.32	0.63	0.05

Data Source: MPS (2023) & UNSD (2023) and ILO (2023). Dispersed shareholders: All firms categorised as having dispersed shareholder ownership added to the denominator. DS + Partnerships: All firms having dispersed shareholder ownership plus firms categorised as partnerships added to the denominator.

Table C3: Regression of the Ratio between Original & Adjusted on GDPpc

	(1)	(2)	(3)	(4)
Adjustment Ratio	DS	DS	DS & P	DS & P
Dataset	IPUMS/UN	ILO	IPUMS/UN	ILO
Log Real GDPpc	0.0529***	0.0463***	0.0851***	0.0662***
	(0.0115)	(0.0105)	(0.0138)	(0.0134)
Constant	-0.411***	-0.449***	-0.657***	-0.602***
	(0.101)	(0.0951)	(0.121)	(0.118)
Observations				
R-squared	98	94	98	94

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Data Source: MPS (2023) & UNSD (2023) , ILO (2023) and PENN WT10.0.. Dependent variables is the log of the ratio of the adjusted variable to the initial variable. DP: All firms categorised as having dispersed shareholder ownership added to the denominator. DS + P: All firms having dispersed shareholder ownership plus firms categorised as partnerships added to the denominator.

Table C4: WBES Adjusted Regressions

Mean Firm Size Adjustment	(1) IPUMS/UN Original	(2) IPUMS/UN DS	(3) IPUMS/UN DS & P	(4) ILO Original	(5) ILO DS	(6) ILO DS & P
Log Real GDPpc	0.604***	0.551***	0.519***	0.565***	0.519***	0.499***
	(0.0648)	(0.0592)	(0.0606)	(0.0566)	(0.0544)	(0.0552)
Constant	-3.618***	-3.207***	-2.961***	-3.658***	-3.209***	-3.057***
	(0.616)	(0.567)	(0.575)	(0.494)	(0.474)	(0.480)
Observations	98	98	98	94	94	94
R-squared	0.341	0.331	0.308	0.448	0.444	0.427

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Data Sources: MPS (2023) & UNSD (2023), ILO (2023) & PENN WT10.0. DP: All firms categorised as having dispersed shareholder ownership added to the denominator. DS + P: All firms having dispersed shareholder ownership plus firms categorised as partnerships added to the denominator.

About DataFirst

DataFirst is a research data service dedicated to giving open access to data from South Africa and other African countries.

We promote high quality research by providing the essential Open Research Data infrastructure for discovering and accessing data and by developing skills among prospective users, particularly in South Africa.

We undertake research on the quality and usability of national data and encourage data usage and data sharing.

DataFirst

www.datafirst.uct.ac.za
Level 3, School of Economics Building,
Middle Campus, University of Cape Town
Tel: +27 (0)21 650 5708
info@data1st.org / support@data1st.org

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD