University of Toronto
Department of Economics

Working Paper 517

# Misallocation, Establishment Size, and Productivity

By Pedro Bento and Diego Restuccia

July 25, 2014

# Misallocation, Establishment Size, and Productivity[†]

Pedro Bento
West Virginia University[*]

Diego Restuccia
University of Toronto[**]

July 2014

Abstract

We construct a new dataset using census, survey, and registry data from hundreds of sources to document a clear positive relationship between aggregate productivity and average establishment size for manufacturing establishments across 124 countries. We rationalize this relationship using a standard model of reallocation among production units that features endogenous entry and productivity investment. The model connects small operational scales in poor countries to the prevalence in these countries of correlated distortions (the elasticity between wedges and establishment productivity). The model also rationalizes the low establishment-level productivity and aggregate investment found in poor countries. A calibrated version of the model implies that when correlated distortions change from 0.13 in the U.S. to 0.56 in India, establishment size and productivity fall by a factor of six. These substantial size and productivity losses are large compared to the existing literature and more in line with actual data for the differences in size and productivity between India and the United States.

Keywords: misallocation, establishment size, productivity, investment, idiosyncratic distortions.
JEL codes: O1, O4.

[†]We thank Margarida Duarte and Xiaodong Zhu for helpful comments. All errors are our own. Restuccia gratefully acknowledges the financial support from the Social Sciences and Humanities Research Council of Canada.

[*]West Virginia University, College of Business and Economics, Department of Economics, 1601 University Ave., PO Box 6025, Morgantown, WV 26506-6025, USA. E-mail: pedro.bento@mail.wvu.edu.

[**]Department of Economics, University of Toronto, 150 St. George Street, Toronto, ON M5S 3G7, Canada. E-mail: diego.restuccia@utoronto.ca.

# 1 Introduction

A consensus view in the literature has emerged where the large variations in income per capita across countries are mostly explained by differences in total factor productivity (TFP). A potential explanation for these productivity differences is the (mis)allocation of factors of production among heterogeneous production units that differs across countries. An important finding in the empirical literature on misallocation is that not only is there evidence of large productivity effects from misallocation across existing production units (e.g. Hsieh and Klenow, 2009), but also that there are substantial differences in existing establishment-level productivity (e.g., Hsieh and Klenow, 2009; Pagés-Serra, 2010; Gal, 2013). What explains the differences in establishment-level and aggregate productivity across countries? We address this question through a standard model of reallocation with heterogeneous production units that features an investment decision on establishment-level productivity and an entry decision. The model connects two pieces of empirical evidence which we document—the prevalence of correlated distortions and the smaller operational scale of production units in poor countries—to explain low establishment-level productivity that contributes to low aggregate productivity in addition to static misallocation.

Evidence of the relationship between development and establishment size has been both sparce and inconclusive due to the lack of standardized size data for a large group of countries.[1] We address this by constructing a standardized database on establishment and firm sizes based on individual-country data from manufacturing censuses, surveys, and registries. Using hundreds of separate sources, we have assembled data for 124 countries with comparable employment-size data. In contrast to Alfaro et al. (2009) and Bollard et al. (2014), who use international data plagued by cross-country differences in the size of sampled firms, we show that average establishment size is strongly positively correlated with GDP per capita. For instance, whereas average establishment size in U.S. manufacturing is 22 workers, in Benin and Sierra Leone it

[1] Poschke (2014) reports a positive relationship between size and development, while Alfaro et al. (2009) and Bollard et al. (2014) find the opposite relationship. We discuss this further in Section 2.

is about 2 workers, an 11-fold difference. As a summary measure of the effect of development on size, we compute the income elasticity of establishment size to be 0.26. Large differences in operational scales are also found in other sectors such as agriculture from Census data where the operational scale of farms in rich countries is 34 times that of poor countries (see for instance Adamopoulos and Restuccia, 2014). Our data confirms the finding in Poschke (2014) where two separate sources of representative survey data for small and large establishments are shown to present a similar positive income elasticity of size across 43 countries.

We consider a standard model of heterogeneous production units that builds from Lucas (1978). For comparability, the set up follows closely the monopolistic competition framework used in the empirical analysis of Hsieh and Klenow (2009). The basic framework is extended along two important dimensions in order to address differences in entry and establishment-level productivity. We incorporate an endogenous entry decision of establishments as well as an endogenous investment decision on establishment-level productivity by entrants. In order to keep the analysis tractable and without loss of generality we abstract from establishment heterogeneity. In the theory, ex-ante identical establishments make the same investment decision for productivity and hence are ex-post identical. The theory connects policy distortions, institutions, and frictions that create misallocation to establishment-level investment. The key emphasis in the model is the extent to which distortions effectively penalize more productive relative to less productive establishments—what Restuccia and Rogerson (2008) call correlated idiosyncratic distortions. In the model, we show there is a strong connection between the extent of correlated distortions and establishment-level productivity and the mass of entrants in the economy. These effects work to lower establishment size, establishment-level productivity, and aggregate productivity. By staying close to Hsieh and Klenow, we are able to explicitly separate the effects of misallocation working through the investment and entry channels emphasized in this paper, from those working through firms' output decisions analyzed in Hsieh and Klenow.

We calibrate a benchmark economy to U.S. data and show that reasonable variations in the extent of correlated distortions have substantial negative effects on establishment size, establishment-level productivity, and aggregate output per capita. In order to provide a quantitative assessment of this channel in explaining productivity differences across countries, we first document evidence from cross-country micro data for the elasticity between distortions (wedges) and establishment productivity, using establishment-level data from the World Bank's Enterprise Surveys. We show that the elasticity of distortions with respect to productivity in the micro data is strongly negatively related with both average establishment size and GDP per capita across 53 countries. We then evaluate the quantitative implications of the observed measures of correlated distortions in the model on establishment size and productivity. Our results are striking. Compared to the calibrated U.S. benchmark economy, increasing correlated distortions to 0.56 (the elasticity between wedges and establishment productivity found for India) generates a reduction in establishment size to 3.4 workers versus the 22 workers in the U.S. economy (a 6-fold difference), and a similar reduction in establishment-level productivity. The 1.6-fold difference in aggregate productivity between the U.S. and India generated by static misallocation in Hsieh and Klenow (2009) is also amplified in this paper into a 6-fold difference. These are large size and productivity losses compared to the existing literature in misallocation and development and are closer in line to the observed size and productivity levels in India.

We emphasize that with no endogenous investment in productivity, the model would imply no differences in establishment size and productivity. As a result, this channel not only amplifies the losses in output and productivity from misallocation, but also rationalizes the impact of distortions on establishment size as observed in the cross-country data. See for instance the related work of Hopenhayn (2013) emphasizing the potential importance of establishment size in generating substantial differences in productivity from specific distortionary policies. To the extent that misallocation is reduced within a country over time, the model also contributes to understanding trends in establishment size. In the United States, for example, Poschke (2014) reports a doubling of average firm size since the early twentieth century, while Ziebarth (2013) and Hsieh and Klenow (2009) suggest a significant reduction in misallocation in the U.S. over the same time period.

Our paper is closely related to the broad literature on misallocation and productivity.[2] In particular, within this literature we relate to papers studying the impact of policies and institutions generating misallocation that also induce disincentives for establishments to invest in productivity. Early examples of this literature are Restuccia (2013a) and Bello et al. (2011) with more elaborate analysis in Ranasinghe (2014), Bhattacharya et al. (2013), Gabler and Poschke (2013), and Da-Rocha, et al. (2014), among others. Our modeling of the mechanism through which misallocation affects productivity investment is similar to both Bhattacharya et al. (2013) and Hsieh and Klenow (2014). The primary contribution in this paper is to isolate the effect of misallocation working through this mechanism, disentangling it from both the pure output effects of misallocation (i.e., those estimated in Hsieh and Klenow, 2009) and the effects of various other distortions present in Hsieh and Klenow (2014) and Bhattacharya et al. (2013).

In generating differences in establishment size, our work is closely related to the seminal work of Lucas (1978) who showed that an elasticity of substitution less than one may be needed in the production function between capital and labor in order to rationalize the larger operational scales in rich countries. In our framework, even with Cobb-Douglas technology, establishment size can vary with correlated distortions. The view that differences in size across countries can arise from distortions shares with the work of Guner et al. (2008) who emphasize size-dependent distortions, i.e., distortions such as a taxes and regulations that apply only beyond a threshold size in terms of the number of workers in the firm. We differ from Guner et al. in that in our framework any correlated distortion causes productivity at the establishment level to drop for all establishments, adding to the potential static misallocation effects typically emphasized in the literature. For this reason, the size and productivity impact of correlated distortions in our framework are orders of magnitude larger than those emphasized in Guner et al. (2008). The literature has also explored many specific policies thought to explain income differences across

[2]See for instance the surveys of this literature in Restuccia and Rogerson (2013), Restuccia (2013b), and Hopenhayn (2013).

countries such as firing costs, entry costs, or average tax rates that lead to larger establishment sizes in poor countries (e.g., Hopenhayn and Rogerson, 1993; Barseghyan and DiCecio, 2011; Moscoso Boedo and Mukoyama, 2012; among others). To the extent that poor countries have both harmful policies *and* correlated distortions, this paper helps to rationalize why establishments are smaller in countries with higher average costs of doing business.

The paper is organized as follows. In the next section, we present the facts from our constructed dataset of 124 countries to establish that establishment size increases substantially with the level of development across countries. Section 3 presents the model and characterizes the qualitative implications. In Section 4, we calibrate the model to data for the United States and show the quantitative implications of the model for hypothetical variations in the extent of correlated distortions. We then construct and document measures of correlated distortions across countries and assess their potential to generate differences in size and productivity. We also discuss our results for reasonable extensions in the model and reasonable variations in key parameter values. We conclude in Section 5.

## 2 Average Establishment Size Across Countries

We construct a dataset for the average employment size of manufacturing establishments and firms across countries between 2000 and 2012 using hundreds of economic censuses and surveys which use comprehensive business registries to create sampling frames.[3] We include all countries with publicly available data representative of all manufacturing establishments or firms.[4] Not included in the data are businesses without a fixed location. Businesses operating out of households are generally included only if signs are posted on the premises.[5]

[3]In the appendix we provide greater detail about how we construct the dataset, a list of the countries included, and a list of the sources we use for each country.

[4]The dataset also includes all territories such as French Guiana, Hong Kong, and Puerto Rico. We use the word 'country' solely for ease of exposition.

[5]The sole exception to this rule is the United States. Although U.S. employer data uses a standard definition of 'establishment,' the data for non-employers (i.e., self-employed) includes businesses with no fixed location

Our standardized definition of size is the average number of persons engaged per establishment (defined as a physical location where economic activity takes place), but different countries report statistics using different definitions. Some countries report the number of active establishments and firms (where a firm is a collection of establishments under common ownership), while other countries report only one or the other. Three definitions of employment are also used: persons engaged (paid and unpaid), average number of employees (including full-time and part-time paid workers), and full-time equivalent employees. In the appendix we explain in detail how we use these data to construct our standardized measure.

It is worth digressing at this point to note that accounting for establishments without paid employees is crucial when investigating differences in establishment size across rich and poor countries, as these establishments (which often employ unpaid family members) account for a significant portion of establishments in poor countries. In Sierra Leone, for example, 83 percent of establishments have no paid employees, and in Ghana, unpaid workers account for almost half of the manufacturing workforce. As a result, excluding non-employer establishments would generate a highly distorted picture of cross-country establishment size differences.

In our final dataset, persons engaged per establishment is averaged over all years for each of the 124 countries.[6] Table 1 reports some descriptive statistics concerning average establishment size, GDP per capita (relative to the U.S.), and population.[7]

Figure 1 shows average establishment size for 124 countries in relation to GDP per capita. Both variables are in logs and reported relative to the U.S. (this is also the case for all subsequent

like food trucks or sub-contractors in construction. Our focus on manufacturing should prevent this from being an issue, but our reported employment size for the U.S. may as a result be slightly biased downwards.

[6]Although size data is available for Norfolk Island, it has been dropped for lack of any reliable measure of GDP per capita.

[7]GDP per capita (adjusted for purchasing power parity, PPP) is from Penn World Table v. 8.0 for 98 countries, the IMF's World Economic Outlook 2013 for 7 countries, the CIA World Factbook for 14 countries. For four countries (actually overseas departments of France), GDP per capita is from France's National Institute of Statistics and Economic Studies and is made relative to the U.S. GDP per capita using market exchange rates. GDP per capita for Âland Islands is from Statistics and Research Âland, and adjusted for purchasing power parity using Finland's PPP exchange rate from Penn World Table v. 8.0. Population data is from Penn World Table v. 8.0 (98 countries), the World Bank's World Development Indicators (21 countries), the CIA World Factbook (4 countries), and Statistics and Research Âland (for Âland Islands).

Table 1: Descriptive Statistics

| | Mean | Median | Poorest Decile | Richest Decile |
|---|---|---|---|---|
| Establishment Size | 12 | 9 | 6 | 19 |
| GDP per capita (thousands) | 18 | 13 | 1.3 | 54 |
| Population (millions) | 32 | 6 | 30 | 25 |

Notes: 'Poorest' and 'richest' refer to GDP per capita. Data from multiple sources, see text for details.

figures). The data clearly show a positive correlation between average establishment size and GDP per capita. In particular, the elasticity of establishment size with respect to GDP per capita is 0.26. Figure 2 shows that the correlation between size and income is even stronger if we omit small countries with populations less than half of one million. In this case, the elasticity rises to 0.33. Recently developed models linking market size and markups predict both GDP per capita and establishment size should increase with population, suggesting that the relationship illustrated in Figures 1 and 2 could be explained by differences in population size across countries.[8] But Figure 3 shows that establishment size seems to be unrelated to population.[9]

To confirm that the observed relationship between establishment size and GDP per capita is not being driven by how the establishment size data is constructed, we separately test the relationship between size and GDP for persons engaged per establishment, persons engaged per firm, employees per establishment, and employees per firm, using only the source data for each country. The corresponding elasticities are all positive and of comparable magnitude: 0.40, 0.36, 0.20, and 0.15.[10]

We now compare the implications of our data relative to the existing work in the literature.

[8]Melitz and Ottaviano (2008) and Desmet and Parente (2012), for example, each develop models in which larger populations can lead to both higher output per capita and larger establishments.

[9]The regression coefficient (standard error) in Figure 1 is 0.26 (0.05), and in Figure 2 it is 0.33 (0.05). In Figure 3 it is an insignificant 0.01 (0.03).

[10]The corresponding number of observations in each regression are 53, 41, 42, and 49.

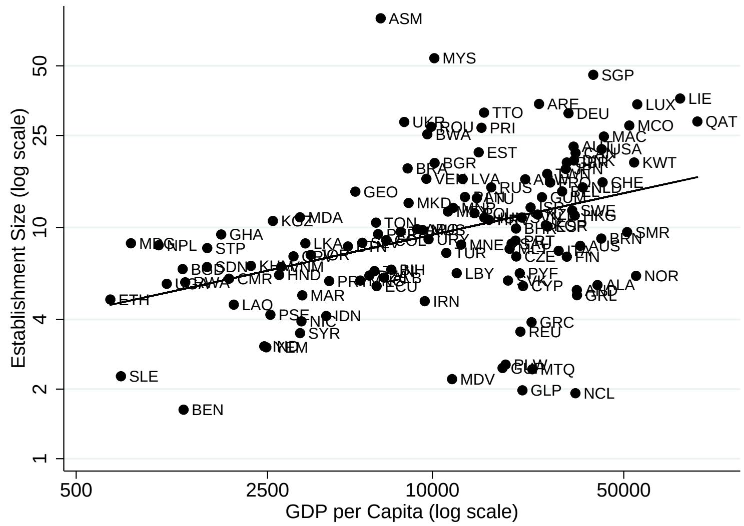


Figure 1: Establishment Size and GDP per Capita

A widely cited reference for the relationship between firm size and income is Alfaro et al. (2009). They use Dun & Bradstreet's WorldBase data (DB) to document a negative relationship between firm size and income per capita across 79 countries. More recently, Bollard et al. (2014) report the same negative relationship using data from the United Nations Industrial Development Organization's (UNIDO) Industrial Statistics Database for 72 countries. These observations are in direct contrast to those just documented from our data. They are also opposite to the relationship found for specific sectors such as agriculture where census data indicates much smaller farm size operation in poor countries relative to rich (see for instance Adamopoulos and Restuccia, 2014). To understand Alfaro et al., it useful to first emphasize that DB is comprised of business data aggregated from multiple sources that is typically used to provide credit and market assessment services. Key here is that DB has sparse coverage of small firms in poor countries relative to rich countries, with no attempt to make the data representative of all establishments. As a result, when calculating average establishment size in poor countries, the under-representation of small firms biases the average upwards. In a sense, Alfaro et al. are comparing average size across most establishments in rich countries with the average size of only large establishments in poor countries. The UNIDO data used by Bollard et al. similarly include countries with uncomparable populations of establishments, with some

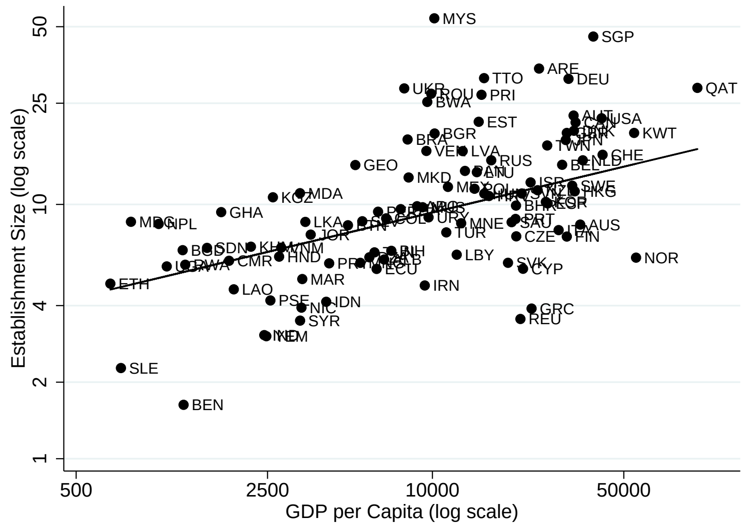


Figure 2: Establishment Size and GDP per Capita (small countries removed)

countries reporting data for all establishments and other countries reporting data only for larger establishments.[11] More importantly, our data contains information for 54 countries from Alfaro et al.'s sample and 55 countries from Bollard et al.'s sample, and the result of a positive relationship between establishment size and development is even stronger in these subsamples than for all 124 countries.

Poschke (2014) finds that firm and establishment size are strongly increasing in development, consistent with our evidence. Poschke uses two datasets, one from Global Entrepreneurship Monitor for small and medium firms in 47 countries, and one from Amadeus for large firms in 34 countries. Unlike Alfaro et al. (2009) and Bollard et al. (2014), the survey data used in Poschke is constructed in such a way as to be representative of all firms for each country. Although his sample of countries is smaller, Poschke is able to show that size is increasing in development across multiple sectors of the economy.

Comparing the results of our analysis with those of Alfaro et al. (2009), Bollard et al. (2014), and Poschke (2014) makes it clear that analyzing standardized, representative size data, especially with respect to the smallest establishments in poor countries, is crucial to obtaining clear

[11]For this reason, some of the countries used in Bollard et al. (2014) have been excluded from our dataset.

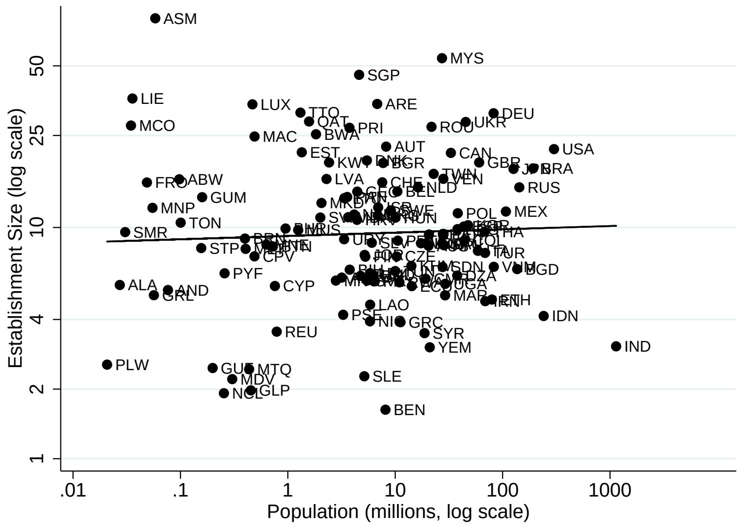


Figure 3: Establishment Size and Population

evidence of the relationship between development and average employment size of establishments across countries.

## 3 The Model

Consider an economy where time is discrete and indexed by $t$. A representative final-good firm uses a variety of imperfectly substitutable inputs from intermediate-good firms to produce the final consumption good.[12] There is a stand-in household endowed with a continuum of members (of measure one), each supplying one unit of labor each period. Each household member can either operate a firm or work for a firm. Household members are free to start an intermediate-good firm, but must make an investment decision determining firm-level productivity upon entry, and forgo the market wage for the life of the firm.[13] Firms face output distortions which may be correlated with firm-level productivity. Entrants take policy distortions into account when investing in productivity. We assume an exogenous probability of exit and, as a result,

[12]Throughout we use 'firm' and 'establishment' interchangeably.

[13]This specification of operating costs is consistent with Bollard et al., 2014, who argue using time-series data that entry and operating costs should scale up with secular development.

there is ongoing entry and exit in steady state. We study the decentralized equilibrium of the economy in which firms take the wage, the interest rate, and the size of the economy as given, and free entry ensures the value of entry is driven to zero. We then consider how the extent of correlated distortions affects the number of firms, investment, and aggregate output.[14] We begin by describing the environment in more detail.

## 3.1 Environment

The representative final-good firm produces output using a variety of inputs from intermediate-good firms according to the following production function;

$$Y = \left( \int_0^N y_i^{\frac{\sigma-1}{\sigma}} di \right)^{\frac{\sigma}{\sigma-1}},$$

where $N$ is the number of intermediate-good firms, $y_i$ the demand for input $i$, and $\sigma$ the constant elasticity of substitution between varieties.

Each intermediate-good firm has access to the following production function;

$$y = s\ell,$$

where $s$ is productivity and $\ell$ is labor. Each entrant chooses $s$ by incurring a cost equal to $c_S Y s^{\theta}$ upon entry. At the end of each period after entry, each intermediate-good firm faces an exogenous probability of exit equal to $\lambda$.

Output distortions are such that each firm retains a fraction (1-$\tau$) of its output, and we assume $\tau$ depends on firm-level productivity as follows;

$$(1 - \tau_i) = \left( \frac{s_i}{\bar{s}} \right)^{-\gamma},$$

[14] We refer to the 'measure' of firms as the 'number' of firms for ease of exposition.

where $\bar{s}$ is the average level of productivity across all firms, and the parameter $\gamma$ is the elasticity of a firm's distortion with respect to its productivity. Abstracting from any heterogeneity in productivity and distortions allows us to isolate the effects of $\gamma$ working through each firm's investment decision, without entangling them with the effects of misallocation on each firm's output decision. Below we show that the effects of misallocation working through each channel (investment and output decisions) are multiplicative and can be combined to estimate the total effect of misallocation. Given our assumptions, each firm will be identical in equilibrium, and all entrants will choose to continue operating.

Household members are indifferent between operating and working for a firm. We abstract from the household's intertemporal consumption decision and simply assume an exogenous interest rate $R$.

## 3.2 Equilibrium

We focus on the steady-state decentralized equilibrium of the economy in which prices and allocations are constant. A *steady-state decentralized equilibrium* is defined as a wage rate $w$, intermediate-good price $P$, firm-level productivity $s$, output $y$, labor demand $\ell$, profits $\pi$, number of firms $N$, and aggregate output $Y$, such that;

*(i)* given $P$, the final-good firm demands intermediate-good inputs to maximize profits in each period,

*(ii)* given $w$, $R$, and $Y$, intermediate-good producers choose labor to maximize per-period profits,

*(iii)* given $w$, $R$, and $Y$, entrants choose productivity to maximize the expected present value of lifetime profits,

*(iv)* free entry ensures the expected present value of lifetime profits for an entrant is equal

to the optimal productivity investment plus the expected present value of the owner's forgone wages (zero profit condition),

($v$) markets clear, i.e., the supply of labor (equal to one) is equal to the number of firms plus the quantity of productive labor demanded by firms.

The final-good firm takes input prices as given and maximizes profits in each period, generating the following inverted demand function for each input $i$;

$$P_i = Y^{\frac{1}{\sigma}} y_i^{\frac{-1}{\sigma}}.$$

Profits in each period for an incumbent firm-$i$ are therefore;

$$\pi_i = (1-\tau_i) Y^{\frac{1}{\sigma}} y_i^{\frac{\sigma-1}{\sigma}} - w\ell_i, \quad \text{where} \quad y_i = s_i \ell_i.$$

Firms choose labor to maximize profits each period, generating the following demand for labor and optimal output;

$$\ell_i = \frac{(1-\tau_i)^\sigma Y s_i^{\sigma-1}}{w^\sigma} \left(\frac{\sigma-1}{\sigma}\right)^\sigma,$$

$$y_i = \frac{(1-\tau_i)^\sigma Y s_i^{\sigma}}{w^\sigma} \left(\frac{\sigma-1}{\sigma}\right)^\sigma.$$

Per-period profits for firm-$i$, given $s_i$, are therefore;

$$\pi_i = \frac{(1-\tau_i)^\sigma Y s_i^{\sigma-1} (\sigma-1)^{\sigma-1}}{w^{\sigma-1} \sigma^\sigma}. \tag{1}$$

Combining $y_i$ above with the final-good production function results in the following expression;[15]

$$w^\sigma = \left(NE\left[s^{\sigma-1}(1-\tau)^{\sigma-1}\right]\right)^{\frac{\sigma}{\sigma-1}} \left(\frac{\sigma-1}{\sigma}\right)^\sigma. \tag{2}$$

[15]Although there is no heterogeneity in productivity or distortions, we nevertheless continue (for a time) as if heterogeneity exists across firms in order to show that without investment in productivity, the effects of misallocation in this economy reduce to those in Hsieh and Klenow (2009).

Labor-market clearing results in;

$$1 = N\left(E[\ell] + 1\right),$$

or

$$w^{\sigma} = \frac{N}{1-N} Y E\left[s^{\sigma-1}(1-\tau)^{\sigma}\right]\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}. \tag{3}$$

Combining equations (2) and (3) and rearranging results in expressions for aggregate output and the wage rate, both as functions of the average tax rate, average firm-level productivity, and the number of firms;

$$Y = N^{\frac{1}{\sigma-1}}(1-N)\frac{E\left[s^{\sigma-1}(1-\tau)^{\sigma-1}\right]^{\frac{\sigma}{\sigma-1}}}{E\left[s^{\sigma-1}(1-\tau)^{\sigma}\right]}, \tag{4}$$

$$w = \frac{Y}{1-N}\left(\frac{\sigma-1}{\sigma}\right)\left(\frac{E\left[s^{\sigma-1}(1-\tau)^{\sigma}\right]}{E\left[s^{\sigma-1}(1-\tau)^{\sigma-1}\right]}\right). \tag{5}$$

We digress here to note that aggregate output in equation (4) can be rewritten as;

$$Y = N^{\frac{1}{\sigma-1}}(1-N)E\left[s^{\sigma-1}\left(\frac{\overline{MRPL}}{MRPL_i}\right)^{\sigma-1}\right]^{\frac{1}{\sigma-1}}, \tag{6}$$

where a firm's revenue marginal product of labor and the average revenue marginal product of labor are defined as in Hsieh and Klenow (2009);

$$MRPL_i = \frac{P_i y_i}{\ell_i} = \left(\frac{\sigma}{\sigma-1}\right)\frac{w}{(1-\tau_i)},$$

$$\overline{MRPL} = E\left[MRPL_i^{-1} \cdot \frac{P_i y_i}{E[P_{i'} y_{i'}]}\right]^{-1}.$$

Equation (6) makes clear that if firm-level productivity were exogenous, removing misallocation (that is, setting $MRPL_i = \overline{MRPL},\ \forall i$) would have the same effect on aggregate output as in Hsieh and Klenow (2009), as long as the number of firms $N$ is not affected. To see that $N$ is indeed unaffected if productivity is exogenous, we use equations (1) and (3) to derive the expected per-period profits of an entrant;

$$E[\pi] = \frac{Y(\sigma-1)^{\sigma-1}}{w^{\sigma-1}\sigma^{\sigma}} E\left[s^{\sigma-1}(1-\tau)^{\sigma}\right] = w \cdot \frac{1-N}{N(\sigma-1)}. \tag{7}$$

As long as the opportunity cost of operating a firm scales up with the wage and is otherwise unaffected by misallocation, equation (7) shows that the number of firms is independent of the extent of misallocation.

We now take into account that $1-\tau_i$ is equal to;

$$1-\tau_i = \left(\frac{s_i}{\bar{s}}\right)^{-\gamma}.$$

The value of entering for some firm $i$ can now be expressed as;

$$v_i = \frac{\pi_i - w}{1-\rho} - c_S Y s_i^{\theta},$$

or

$$v_i = \frac{Y(\sigma-1)^{\sigma-1}}{(1-\rho)w^{\sigma-1}\sigma^{\sigma}} \frac{s_i^{\sigma(1-\gamma)-1}}{\bar{s}^{-\sigma\gamma}} - \frac{w}{1-\rho} - c_S Y s_i^{\theta}, \tag{8}$$

where $\rho \equiv \frac{1-\lambda}{1+R}$, $\lambda$ is the probability of firm death, and $R$ is the real interest rate.

An entering firm will choose its productivity to maximize expected discounted profits, resulting in the following condition;

$$c_S Y s_i^{\theta} = \frac{Y(\sigma-1)^{\sigma-1}}{(1-\rho)w^{\sigma-1}\sigma^{\sigma}} \frac{s_i^{\sigma(1-\gamma)-1}}{\bar{s}^{-\sigma\gamma}} \frac{[\sigma(1-\gamma)-1]}{\theta},$$

or (given that $s_i = \bar{s}$, $\forall i$ in equilibrium)

$$c_S Y s^{\theta} = \frac{E[\pi]}{(1-\rho)} \frac{[\sigma(1-\gamma)-1]}{\theta}. \tag{9}$$

Free entry guarantees the value of entry will be zero in equilibrium, resulting in the following free-entry condition;

$$\frac{E[\pi][\theta + 1 - \sigma(1-\gamma)]}{\theta} = w. \tag{10}$$

Using equations (7), (9), and (10), we can now solve for firm-level productivity and the number of firms in a stationary equilibrium;

$$s = \Delta \left( \frac{[\sigma(1-\gamma) - 1][\theta\sigma + 1 - \sigma(1-\gamma)]}{[\theta + 1 - \sigma(1-\gamma)]} \right)^{\frac{1}{\theta}}, \tag{11}$$

$$N = \frac{[\theta + 1 - \sigma(1-\gamma)]}{\theta\sigma + 1 - \sigma(1-\gamma)}. \tag{12}$$

$$\text{where} \quad \Delta \equiv \left( \frac{1}{\theta(1-\rho)\sigma c_S} \right)^{\frac{1}{\theta}}.$$

Equations (11) and (12) show that productivity is decreasing in the elasticity of distortions with respect to productivity ($\gamma$), whereas the number of firms is increasing in this elasticity. To gain intuition for these results, it is useful to first combine equations (2), (7), and (9) to obtain an optimal investment condition expressing the optimal productivity $s$ as a function of both $\gamma$ and the number of firms $N$;

$$s = \Delta \left( \frac{[\sigma(1-\gamma) - 1]}{N} \right)^{\frac{1}{\theta}}. \tag{13}$$

We now use equations (7) and (8) to create an alternative free-entry condition in order to show the value of $s$ necessary to ensure the value of entry is equal to zero, given $N$;

$$s = \Delta\theta^{\frac{1}{\theta}} \left( \frac{1}{N} - \frac{\sigma - 1}{1 - N} \right)^{\frac{1}{\theta}}. \tag{14}$$

Both of the above conditions are represented in Figure 4, illustrating how $s$ and $N$ are obtained in equilibrium. As the elasticity of distortions with respect to productivity ($\gamma$) is increased, the incentive to invest in productivity is reduced, shifting the investment curve down. This reduction in the investment of entrants increases the value of entry, given $N$. Free entry thus encourages a movement down and to the right along the free entry curve until the increase in the number of firms brings the value of entry back down to zero.

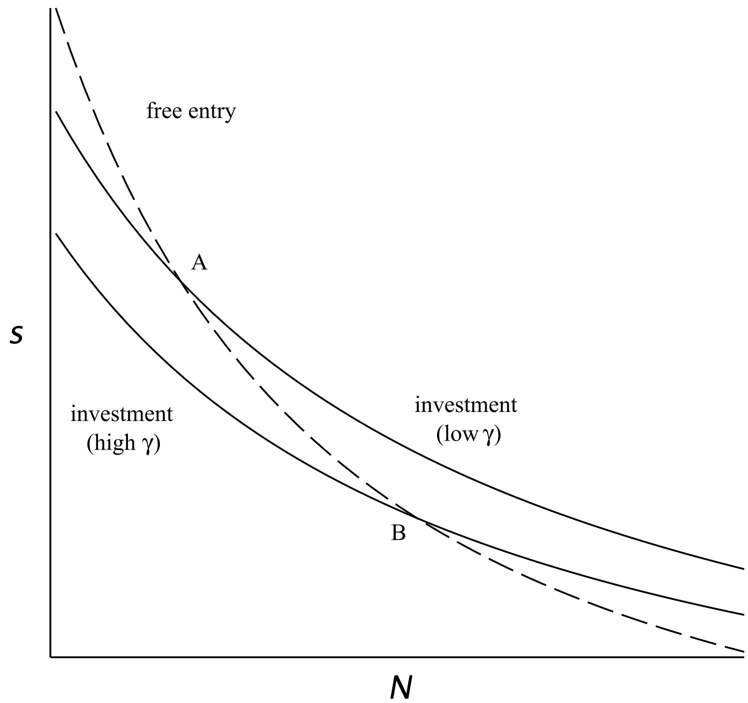


Figure 4: Firm-level Productivity ($s$) and the Number of Firms ($N$) in Equilibrium

We can now express aggregate output as;

$$Y = sN^{\frac{1}{\sigma-1}}(1-N), \tag{15}$$

given $s$ and $N$ above. Aggregate investment in productivity, as a fraction of output, is;

$$\lambda N c_S s^{\theta} = \frac{\lambda[\sigma(1-\gamma)-1]}{\sigma\theta(1-\rho)}. \tag{16}$$

If establishment-level productivity is exogenous, the effects of misallocation would reduce to those estimated by Hsieh and Klenow (2009) as emphasized in equation (6). Further, if greater misallocation simply implied more dispersion in random idiosyncratic distortions (uncorrelated with productivity), then equations (7) and (9) together imply that investment in productivity would be unaffected by misallocation. This reinforces the finding of Restuccia and Rogerson

(2008) that simple random dispersion in idiosyncratic distortions cannot explain much variation in aggregate TFP, and that the strength of correlated distortions (in this paper, $\gamma$) is what generates the large potential impact from misallocation. In Restuccia and Rogerson more correlated distortions cause a reallocation of resources, increasing the output of unproductive establishments marginally while disproportionately decreasing the output of productive establishments (while leaving the equilibrium number of establishments unchanged). By incorporating endogenous entry and investment, we show in this paper that more correlated distortions also reduce the marginal benefit of investing in productivity, thus reducing establishment-level productivity and decreasing average employment across all establishments.

# 4 Quantitative Analysis

In this section, we calibrate the model to U.S. data and show the quantitative implications for establishment size, productivity, and aggregate output of hypothetical variations in the degree of correlated distortions across countries. We then use establishment-level data to estimate empirically the extent of correlated distortions across countries and their implications for cross-country variations in establishment size, productivity, and output. We end the section with a discussion of these results for variations in the model setup as well as some robustness checks on parameter values.

## 4.1 Calibration

We calibrate the model to data for the United States in order to quantify the cross-country effects of correlated distortions on average establishment size, productivity, and aggregate output. These effects depend on three key parameters in our model:

- $\theta$: the elasticity of the investment cost function with respect to establishment productivity,

- $\gamma$: the elasticity of gross distortions $(1-\tau)^{-1}$ with respect to productivity,

- $\sigma$: the elasticity of substitution between varieties.

In order to keep a close tie with the literature for comparison, we follow Hsieh and Klenow (2009, 2014) in setting $\sigma = 3$. For the U.S., Hsieh and Klenow (2014) report $\gamma_{US} = 0.13$. Given $\sigma$ and $\gamma_{US}$, equation (12) implies that we can obtain a value for $\theta$ by targeting the average size of manufacturing establishments in the United States of 21.8 workers. The result is $\theta = 1.78$. We note that this value for $\theta$ is close to the value of 2.01 estimated using trade data in Rubini (2014). From equation (16), $\theta = 1.78$ also implies a share of aggregate output invested in productivity of 21 percent.[16] This number is higher than McGrattan and Prescott's (2010) estimate that 13.5 percent of U.S. GDP is invested in technology and plant-specific intangible capital, but we note that the R&D intensity of U.S. manufacturing (which can be interpreted as a proxy for total investment in productivity) is substantially higher than that of the entire U.S. economy.[17]

With these calibrated parameter values, we quantify how average establishment size and output per capita (TFP) changes with the extent of correlated distortions, i.e., when $\gamma$ is increased above the U.S. level (keeping other parameters constant). We report these implications in Table 2. The main finding is that the model implies large variations in average establishment size and output per capita across economies with different correlated distortions.

For instance, an economy with $\gamma = 0.4$ features an average establishment size that is 21 percent of that in the United States. This economy also features an establishment-level productivity that is 28 percent of the benchmark due to lower investment in productivity (only 50 percent of that in the benchmark). As a result, output per capita is 50 percent of the benchmark economy. These are large differences in size and productivity compared to the findings in

[16] To calculate the investment share of output predicted by the model, we assume $\lambda = 0.1$ and $R = 0.05$.

[17] In addition, Baldwin et al. (2009) report Canadian manufacturing's intangible investment as a share of value added is 1.42 times larger than the investment share of the entire economy. Assuming the same inflator for U.S. manufacturing generates an investment share of 19 percent, very close to our result.

Table 2: Model Results across Correlated Distortions $\gamma$

| $\gamma$ | Establishment Size | Establishment Productivity | Productivity Investment | Relative Output |
|---|---|---|---|---|
| 0.13 ($\gamma_{US}$) | 22 | 1 | 0.21 | 1 |
| 0.2 | 10 | 0.61 | 0.18 | 0.84 |
| 0.3 | 6.2 | 0.40 | 0.14 | 0.66 |
| 0.4 | 4.6 | 0.28 | 0.10 | 0.50 |
| 0.5 | 3.8 | 0.19 | 0.07 | 0.36 |
| 0.56 ($\gamma_{India}$) | 3.4 | 0.14 | 0.04 | 0.27 |

Notes: Columns report equilibrium values of average establishment size ($1/N$), establishment productivity ($s$), aggregate productivity investment over output, and aggregate output ($Y$). Results in columns 2 and 4 are reported relative to the benchmark U.S. economy.

the broad literature on misallocation. The $\gamma$'s in Table 2 are hypothetical, but the range is plausible. As a point of reference, consider that Hsieh and Klenow (2014) report $\gamma = 0.56$ for India. Given this value for $\gamma$, the model predicts an average establishment size of 3.4 workers, close to the value of 3 workers found in the data. The model also predicts India should have an aggregate output of about 27 percent of the U.S. level, but note that this estimates the impact of $\gamma$ working solely through investment and entry. If we combine the impact of $\gamma = 0.56$ in this paper with the additional output effects of misallocation in India reported in Hsieh and Klenow (2009), this implies that reducing the extent of misallocation in India to that of the U.S. would increase Indian TFP by a factor of six.[18]

## 4.2 Correlated Distortions

The calibrated model shows how correlated distortions encourage smaller establishments, lower aggregate output, and lower investment in productivity. In this section, we provide systematic evidence that the productivity elasticity of distortions is indeed higher in poor countries. We

[18] Hsieh and Klenow (2009) estimate that the greater misallocation present in India is responsible for a 1.6-fold difference in TFP between India and the United States. Combining this estimate with ours results in a predicted TFP for India of 17 percent relative to the U.S., a 6-fold difference.

then provide evidence consistent with the mechanism highlighted in Section 3, using cross-country R&D data to show that aggregate investment in R&D is increasing in development, average establishment size, and the extent of correlated distortions.

Our measure of correlated distortions is constructed using establishment-level data from the World Bank's Enterprise Surveys. Enterprise Surveys is an ongoing project of the World Bank to collect establishment-level data from mostly low and middle-income countries through face-to-face surveys. The dataset contains standardized information about sales, intermediate purchases, inputs, and a host of other variables for establishments in over 100 countries for at least one year since 2002. In each country, between 150 and 1000s of establishments have been surveyed, and efforts have been made to make these samples representative of the population of establishments with at least five employees.[19] Importantly for our purposes, manufacturing establishments are classified into fifteen industries. From this dataset, we use observations containing values for industry classification, sales, number of employees, total wage bill, and purchases of materials and intermediate goods, for countries with establishment size data in Section 2.

We back out our measure of establishment-level distortions and productivity for each establishment within a country-industry following Hsieh and Klenow (2009), except that we do not use capital data (more on this below). Abstracting from capital allows us to increase the number of usable countries by 40 percent, as a large number of establishments in the Surveys do not report capital. From Section 3, labor productivity for some establishment $i$ is;

$$\frac{P_i y_i}{\ell_i} = \frac{w}{(1-\tau_i)}\left(\frac{\sigma}{\sigma-1}\right) \propto \frac{1}{(1-\tau_i)},$$

where $P_i y_i$ is an establishment's value added (sales minus intermediate inputs) and $\ell_i$ is employment.[20] As in Hsieh and Klenow, we remove the constant in the above expression by using

[19] Given the absence of very small establishments in the Enterprise Surveys data, we need to assume (as we do in Section 3) that the elasticity of distortions with respect to productivity is constant.

[20] Following Hsieh and Klenow (2009), we use an establishment's total wage bill (including benefits) in our

labor productivity *relative to* the weighted average of labor productivity across all establishments within the same industry.[21] We infer an establishment's productivity $s_i$ by exploiting the following relationship;

$$s_i = \frac{y_i}{\ell_i} \propto \frac{(P_i y_i)^{\frac{\sigma}{\sigma-1}}}{\ell_i}.$$

With our measures of distortions and productivity in hand, we then do a simple OLS regression of logged distortions on logged productivity to obtain each country's productivity elasticity of distortions.[22] Some countries have data for two or even three years, so we average elasticities over all years, weighting by the number of observations in each year. We obtain elasticities for 87 countries, 52 of which have establishment-size data.[23] Among these 52 countries, elasticities range from 0.22 to 0.70, averaging 0.52. Among all 87 countries the average elasticity remains 0.52. It is reassuring to note that our computed elasticity for India is 0.56, the same value Hsieh and Klenow (2014) obtain using much more comprehensive micro data. To check the sensitivity of our measures to abstracting from capital, we also calculate elasticities using Hsieh and Klenow's (2009) TFPR and TFPQ as our measures of distortions and productivity. Among the 37 countries which satisfy the criteria above, the average elasticity is 0.53. If we recalculate these elasticities abstracting from capital data (but only using observations that report capital) we find the same average, and the correlation between the two measures is 0.88.[24]

Figures 5 and 6 show how average establishment size and GDP per capita are related to the productivity elasticity of distortions in 53 countries (the elasticity for the U.S., 0.13, is taken

computations instead of employment in order to control for differences in human capital across establishments.

[21]More precisely, we measure the distortion faced by firm $i$ as;

$$\frac{P_i y_i}{\ell_i} \cdot E\left[\left(\frac{P_{i'} y_{i'}}{\ell_{i'}}\right)^{-1} \cdot \frac{P_{i'} y_{i'}}{E[P_{i'} y_{i'}]}\right].$$

Productivity $s_i$ is similarly measured relative to $E(s_{i'}^{\sigma-1})^{\frac{1}{\sigma-1}}$.

[22]Before doing the regressions, we first trim the 1 percent tails of both distortions and productivity for each country to remove outliers. We then recalculate the averages as above.

[23]We do not use countries with fewer than 100 observations. Over the 52 countries with size data, we use a total of 32,561 establishment-level observations in our regressions.

[24]This is consistent with Gal (2013, Table 9), who calculates both labor productivity and TFPR for firms in a handful of OECD countries and reports correlations between the two statistics ranging from 0.8 to 0.9.

from Hsieh and Klenow, 2014).[25] The data show a clear link between the elasticity and both average size and GDP per capita, consistent with the model.

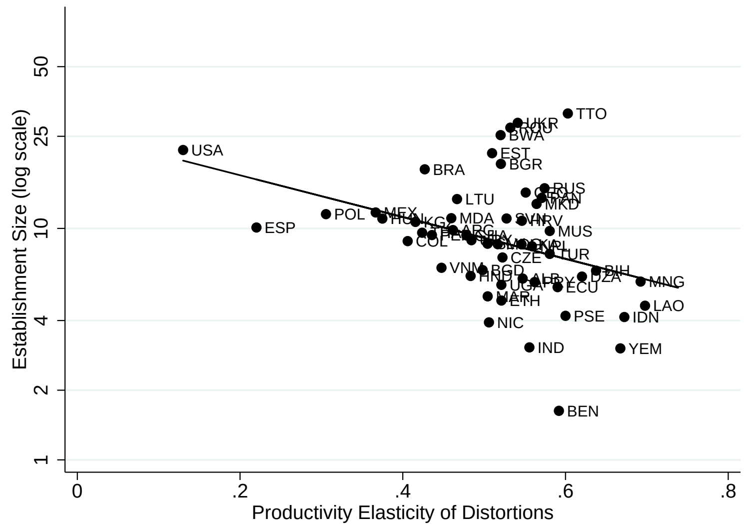


Figure 5: Establishment Size and Correlated Distortions

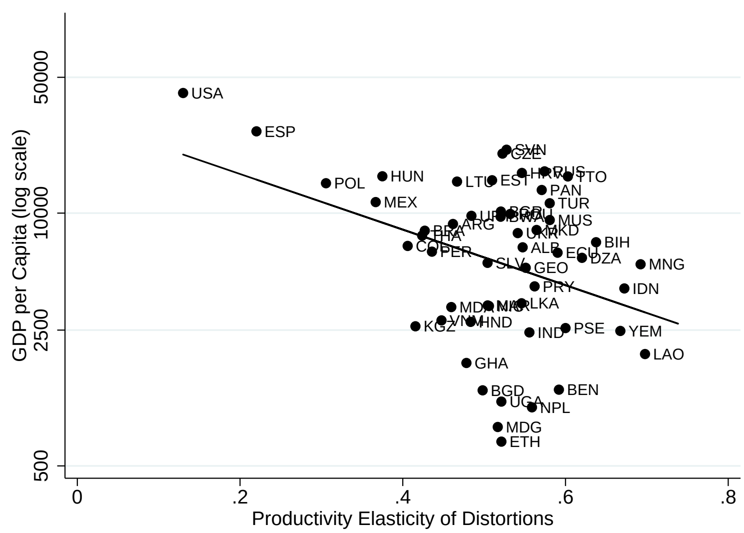


Figure 6: GDP per Capita and Correlated Distortions

In the model developed in Section 3, the mechanism through which correlated distortions reduce establishment size is the disincentive to invest in productivity. As a consequence, the model

[25]The regression coefficients (standard errors) in Figures 5 and 6 are -2.08 (0.53) and -3.29 (0.83).

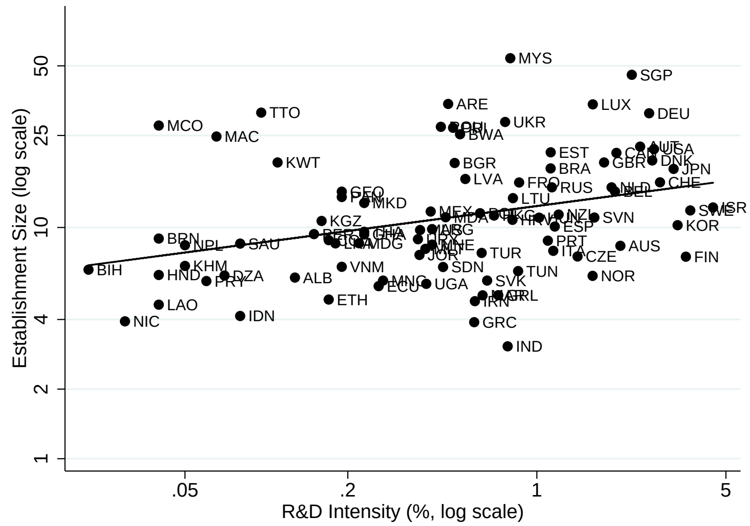


Figure 7: Establishment Size and R&D Intensity

also predicts the share of output invested in productivity should be lower in economies with high $\gamma$. Broad measures of investment in intangible capital have not yet been collected for a large number of countries, but R&D intensity (one significant component of intangible capital investment) should provide a fair proxy. Figure 7 shows how establishment size across countries varies with R&D intensity, while Figure 8 shows how R&D intensity is related to $\gamma$.[26] Again using India as a point of reference, the model predicts an investment share in India about 19 percent of the U.S. level. In the data, India's R&D intensity is 30 percent of the U.S. level. It is important to note here that various other policies that might be correlated with high $\gamma$'s (i.e., higher average distortions, higher entry costs, or a higher average cost of productivity investment $c_S$) might also affect aggregate productivity, but would *not* affect the investment share of aggregate output in the model.

[26]The regression coefficients (standard errors) in Figures 7 and 8 are 0.15 (0.05) and -5.01 (0.98). R&D data is taken from UNESCO, and is calculated as (logged) total investment in R&D as a share of GDP, relative to the U.S.

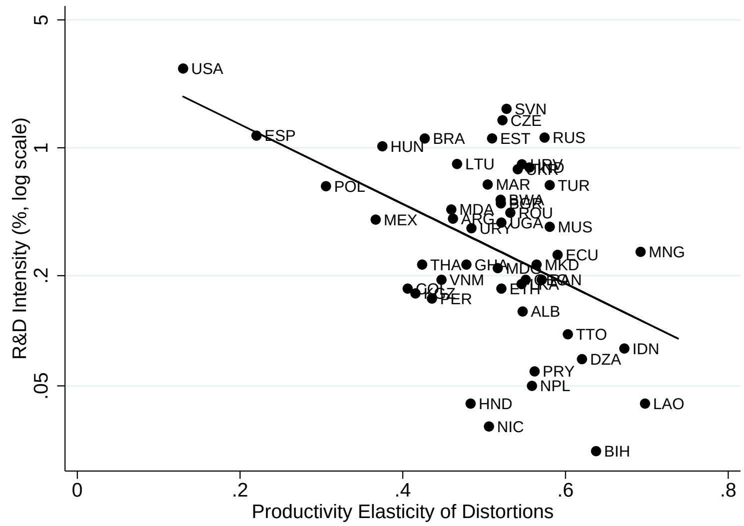


Figure 8: R&D Intensity and Correlated Distortions

## 4.3 Discussion

Here we discuss our main results for reasonable extensions of the model and different values of key parameters.

**Model Extensions** In our baseline specification, the implied elasticity of establishment size with respect to GDP per capita is 0.65, and this elasticity is decreasing in $\theta$.[27] In the baseline specification, the only cost of entry is the labor cost associated with the entrepreneur. We extend the model to allow for a fixed cost of entry by assuming, in addition to the owner's forgone wages, a fixed cost of entry in units of goods $c_F \cdot Y$. We calibrate the economy to the same targets as in the baseline. The calibrated $\theta$ increases to 1.84 (1.78 in the baseline) but the impact of correlated distortions $\gamma$ on output per capita $Y$ and the number of firms $N$ is reduced only marginally. For instance, with $c_F = 2$, we find that for an economy with India's $\gamma$, output is 31 percent of that in the United States and establishment size is 4.1 (27 percent and 3.4 in the baseline). The implied cross-country income elasticity of establishment size is 0.45 (0.65 in the

[27] The elasticity of establishment size with respect to GDP is decreasing in $\theta$ because a higher $\theta$ dampens the impact of $\gamma$ on establishment size more than it dampens $\gamma$'s impact on aggregate output.

baseline). We note that $c_F = 2$ is very high relative to regulatory costs reported in the World Bank's Doing Business Surveys. We have also explored incorporating capital into the model. Introducing capital affects the results of the model because forgone wages of entrepreneurs is a key cost of creating a firm (the only cost in the baseline). Since the wage is a lower fraction of output per worker than in a model without capital, adding capital is just a way of effectively lowering the fixed cost being used in our calibration, implying a lower calibrated $\theta$ and a bigger impact of correlated distortions on size and productivity. The implied cross-country income elasticity of establishment size is 0.44. We have also explored a version with heterogeneity and have verified that our main results are robust to this extension as well.

**Robustness on $\theta$** A critical parameter in determining the impact of correlated distortions on establishment size and output is the elasticity of investment in productivity $\theta$. We explore variations in the cross country implications of the model for two alternative specifications of $\theta$, 1.7 and 2. Recall that in the benchmark calibration $\theta = 1.78$ to match an average establishment size in the US of 21.8 workers. This implies an investment to GDP ratio of 21 percent. When $\theta = 1.7$, establishment size in the benchmark economy is 39 workers with an investment to GDP ratio of 22 percent, whereas when $\theta = 2$, establishment size is 11 workers with an investment to GDP ratio of 19 percent. We report the results of variations in correlated distortions on establishment size, productivity, investment, and aggregate output in Table 3. While the cross-country implications of the model are sensitive to the value of $\theta$, for an empirically reasonable range, the model implies substantial negative effects of correlated distortions on establishment size and productivity.

**Robustness on $\sigma$** As in Hsieh and Klenow (2009), the impact of misallocation is increasing in $\sigma$, the elasticity of substitution between differentiated varieties. But assuming different values for $\sigma$ is made complicated by the fact that $\sigma$ also affects average establishment size. Given our calibration strategy, then, assuming a higher value for $\sigma$ requires a higher value for $\theta$ in order

Table 3: Model Results across Correlated Distortions $\gamma$

| $\gamma$ | Establishment Size | Establishment Productivity | Productivity Investment | Relative Output |
|---|---|---|---|---|
| | | $\theta = 1.7$ | | |
| 0.13 ($\gamma_{US}$) | 39 | 1 | 0.22 | 1 |
| 0.2 | 12 | 0.47 | 0.19 | 0.79 |
| 0.3 | 6.7 | 0.28 | 0.15 | 0.60 |
| 0.4 | 4.8 | 0.19 | 0.11 | 0.45 |
| 0.5 | 3.8 | 0.13 | 0.07 | 0.31 |
| 0.56 ($\gamma_{India}$) | 3.5 | 0.09 | 0.04 | 0.23 |
| | | $\theta = 2$ | | |
| 0.13 ($\gamma_{US}$) | 11 | 1 | 0.19 | 1 |
| 0.2 | 7.7 | 0.77 | 0.16 | 0.89 |
| 0.3 | 5.4 | 0.57 | 0.13 | 0.74 |
| 0.4 | 4.3 | 0.44 | 0.09 | 0.60 |
| 0.5 | 3.7 | 0.32 | 0.06 | 0.44 |
| 0.56 ($\gamma_{India}$) | 3.4 | 0.24 | 0.04 | 0.34 |

Notes: Columns report equilibrium values of average establishment size ($1/N$), establishment productivity ($s$), aggregate productivity investment over output, and aggregate output ($Y$). Results in columns 2 and 4 are reported relative to an economy with $\gamma = \gamma_{US}$.

to match the average size of U.S. manufacturing establishments, which will *decrease* the impact of $\gamma$. In addition, assuming a higher value for $\sigma$ generally increases the estimates of $\gamma$ we obtain in Section 4.2, which will tend to *increase* the implied impact of observed misallocation. In Table 4 we report the estimated impact of misallocation in India relative to the United States with $\sigma = 3$ and with $\sigma = 5$. The net result is that the effect of misallocation in India on most variables is dampened when $\sigma$ is increased. With $\sigma = 5$, the model predicts a 4-fold difference in establishment size (instead of 6), and a 3-fold difference in establishment-level productivity (instead of 6). The effect on aggregate investment actually increases, while the effect on aggregate output is reduced. When we combine our predicted impact on India's TFP with that reported by Hsieh and Klenow (2009) for $\sigma = 5$, we find that the total predicted difference in TFP between India and the United States is reduced from a factor of six to a

Table 4: Model Results for Different $\sigma$'s

| $\sigma$ ($\theta$, $\gamma_{India}$) | Establishment Size | Establishment Productivity | Productivity Investment | Relative Output |
|---|---|---|---|---|
| $\sigma = 3$ ($\theta = 1.78$, $\gamma_{India} = 0.56$) | 0.16 | 0.14 | 0.20 | 0.27 |
| $\sigma = 5$ ($\theta = 4.15$, $\gamma_{India} = 0.77$) | 0.24 | 0.33 | 0.04 | 0.40 |

Notes: Columns report equilibrium values of average establishment size ($1/N$), establishment productivity ($s$), aggregate productivity investment over output, and aggregate output ($Y$). Results are reported relative to an economy with $\gamma = \gamma_{US}$ and $size = size_{US}$, with $\theta$ and $\gamma_{India}$ recalculated for each value of $\sigma$ (see text for details).

factor of five.

# 5 Conclusion

Using a unique dataset of manufacturing establishments we construct from hundreds of sources, we documented a strong positive association between average establishment size and GDP per capita. The cross-country income elasticity of establishment size is 0.26 in our sample of 124 countries. We considered an otherwise standard model of establishments with endogenous entry and investment in establishment productivity. We showed that a reasonably calibrated version of the model generates substantial differences in establishment size and productivity across countries. These differences arise in the presence of correlated distortions which we construct and document across countries. Overall, the analysis in this paper puts us closer to understanding the patterns in operational scale and productivity observed across countries and over time.

Our analysis has abstracted from many factors which may be worth exploring further. For example, we have abstracted from different forms of entry and operation costs that seem to hinder the operation of firms in many poor countries. We have also abstracted from establishment heterogeneity which may interact with the different forms of entry costs. We leave a detailed exploration of these factors for future research.

# A Establishment Size Data

We describe in more detail how we construct the establishment size data. Our standardized definition of establishment size is the number of persons engaged per manufacturing establishment. Persons engaged is defined as the average number of persons working for an establishment, both paid and unpaid. A manufacturing establishment is defined as a physical location where the primary activity is manufacturing. Households are counted as an establishment only if signs are posted on the property indicating commercial activity. Not all countries report persons engaged or the number of establishments, so we also use data on the number of paid employees, the number of full-time equivalent employees, and the number of firms (collections of one or more establishments under common ownership).

The source data for each country is from economic censuses, as well as surveys which use comprehensive business registries to create sampling frames.[28] We use all publicly available data for the years 2000 through 2012.[29] In an effort to maintain consistency across countries, we do not use data unless efforts were made by a statistical agency to make the data representative of an economy's entire population of manufacturing establishments. We exclude any data collected without accounting for small establishments, except in cases where only establishments without paid employees are excluded. In the later case, we use U.S. data to adjust measured establishment size (this is the case for six countries). Further, we include data for any country that excludes establishments with low revenue, as long as the revenue threshold is lower than the country's GDP per capita (this is the case for four countries). Two countries (Algeria and Honduras) do not report employment, but do report the distribution of establishments across multiple employment tranches. In these two cases we estimate total employment by using an average employment within each tranche consistent with data in comparable countries.[30] We

[28]For some countries data is from EUROSTAT or OECD's Structural Business Statistics, but we check each country's methodology to confirm the consistency of definitions.

[29]In some cases countries have published only press releases or bulletins describing the census data. We include these countries when the data meets our criteria.

[30]We assume average employment within a tranche to be one third of the distance from the lower to the upper threshold. For the last open-ended tranche (for example, 200 or more employees) we assume an average

are left with 124 countries with useable data for at least one year, with an average of five years per country.[31] Table 5 reports the total number of countries reporting each variable for at least one year, as well as the total number of poor countries and the total number of rich countries (defined as having GDP per capita below and above the median) doing the same.

Table 5: Sample of Countries

| Variable | Total Number of Countries | Number of Poor Countries | Number of Rich Countries |
|---|---|---|---|
| persons engaged | 86 | 44 | 42 |
|   engaged and employees | 45 | 15 | 30 |
| employees | 82 | 33 | 49 |
| full-time equivalents | 27 | 3 | 24 |
| establishments | 77 | 47 | 30 |
|   establishments and firms | 12 | 3 | 9 |
| firms | 59 | 18 | 41 |

Note: 'Poor' and 'Rich' refer to countries with GDP per capita below and above the median. Data from multiple sources, see text for details.

We construct our standardized measure of establishment size (persons engaged per establishment) in the following way. First, we regress persons engaged on employees and full-time equivalent employees, as well as on employees and full-time equivalents separately. Second, country-years which report employees or full-time equivalents (or both) but lack measures of persons engaged are then assigned predicted values using the estimated coefficients from the above regressions. Third, predicted values for persons engaged per establishment are then assigned to country-years which report only firm-level data using the estimated coefficient from a regression of persons engaged per establishment on persons engaged per firm. Fourth, to standardize persons engaged per establishment in country-years which exclude non-employer establishments (this is the case for six countries), we multiply these values by a factor equal to the average ratio of persons engaged per establishment to persons engaged per establishment with paid employees across all years in the U.S. (this ratio is 0.51). The same method is used to

employment equal to twice the lower threshold.

[31] Although size data is also available for Norfolk Island, it has been dropped for lack of any reliable measure of GDP per capita.

standardize data for five countries which report statistics for a combination of manufacturing, extraction, and energy (this ratio is 1.14). The results of the four regressions described above are;

- persons engaged = 1.44 · employees − 0.41 · full-time equivalents
- persons engaged = 1.07 · employees
- persons engaged = 1.12 · full-time equivalents
- persons engaged per establishment = 0.88 · persons engaged per firm

In our final dataset, persons engaged per establishment is averaged over all years for each of the 124 countries.

Table 6 lists each country in the final dataset, the number of years for which data is available, and the sources from which data has been collected.

Table 6: List of Countries and Sources

| Country | Code | Years | Sources |
|---|---|---|---|
| Âland Islands | ALA | 9 | Statistics and Research Âland: Statistical Yearbooks of Âland 2006-2010 and 2013, and www.asub.ax |
| Albania | ALB | 8 | Instituti i Statistikave: www.instat.gov.al/en/figures/statistical-databases.aspx |
| Algeria | DZA | 1 | Office National des Statistiques, Alger: Premier recensement économique -2011- Résultats définitifs |
| American Samoa | ASM | 2 | U.S. Census Bureau: County Business Patterns |
| Andorra | AND | 12 | Departament d'Estadística: 2010 Statistical Yearbook, and www.estadistica.ad |
| Argentina | ARG | 1 | Instituto Nacional de Estadística y Censos: 2005 Economic Census |
| Armenia | ARM | 4 | Central Bureau of Statistics: Enterprises in Aruba 2000-2008 |
| Australia | AUS | 5 | Australian Bureau of Statistics: Counts of Australian Businesses 2003-2007, Labour Force Surveys (Quarterly) |
| Austria | AUT | 12 | Statistik Austria: statcube.at, and OECD's SDBS Structural Business Statistics |
| Bahrain | BHR | 2 | Kingdom of Bahrain Central Informatics Organization: Population, Housing, Buildings, Establishments and Agriculture Census |
| Bangladesh | BGD | 1 | Bangladesh Bureau of Statistics: Economic Census 2001 & 2003 |
| Belgium | BEL | 11 | Eurostat, and OECD's SDBS Structural Business Statistics |
| Benin | BEN | 1 | Institut National de la Statistique et de l'Analyse Economique: General Census of Companies, and Les Entreprises Artisanales au Benin |
| Bhutan | BTN | 4 | National Statistics Bureau: Statistical Yearbooks 2010-2013 |
| Bosnia and Herzegovina | BIH | 5 | Institute for Statistics of FB&H: Federation of Bosnia and Herzegovina in Figures 2008, 2010, 2013 |
| Botswana | BWA | 5 | Statistics Botswana: Statistical Yearbook 2010, and www.gov.bw/en/Citizens/Topics/Statistics |
| Brazil | BRA | 7 | Brazilian Institute of Geography and Statistics: Cadastro Central de Empresas |
| Brunei | BRN | 1 | Department of Economic Planning and Development: Brunei Darussalam Statistical Yearbook 2010 |
| Bulgaria | BGR | 12 | Eurostat |
| Cambodia | KHM | 2 | National Institute of Statistics: Economic Census 2011, and Establishment Listing 2009 |
| Cameroon | CMR | 1 | Institut National de la Statistique du Cameroun: Recensement Général des Entreprises 2009 |
| Canada | CAN | 7 | Statistics Canada: CANSIM |
| Cape Verde | CPV | 4 | Instituto Nacional de Estatística: Business Census 2007, and Annual Business Surveys 2008-2009 |
| Columbia | COL | 1 | Encuesta Annual Manufacturera, and www.dane.gov.co |
| Croatia | CRV | 4 | Eurostat |
| Cyprus | CYP | 12 | Eurostat |
| Czech Republic | CZE | 10 | Eurostat, and OECD's SDBS Structural Business Statistics |
| Denmark | DNK | 12 | Eurostat, and OECD's SDBS Structural Business Statistics |

Table 6: List of Countries and Sources

| Country | Code | Years | Sources |
|---|---|---|---|
| Ecuador | ECU | 1 | Instituto Nacional Estadística y Censos: National Economic Census 2010 |
| El Salvador | SLV | 1 | Ministerio de Economica: Tomo I de los VII Censos Económicos Nacionales 2005 |
| Estonia | EST | 12 | Eurostat |
| Ethiopia | ETH | 1 | Central Statistical Agency: Report on Small Scale Manufacturing Industries Survey, and Report on Large and Medium Scale Manufacturing and Electricity Industries Survey |
| Faroe Islands | FRO | 12 | Statistics Faroe Islands: www.hagstova.fo |
| Finland | FIN | 12 | Eurostat |
| French Guiana | GUF | 1 | Institut national de la statistique et des études économiques: Caractéristiques des entreprises et établissements |
| French Polynesia | PYF | 1 | Institut de la Statistique de la Polynésie française: Enquête annuelle d'entreprise 2007 |
| FYR Macedonia | FYR | 5 | State Statistical Office: www.stat.gov.mk |
| Georgia | GEO | 10 | National Statistics Office of Georgia: Statistical Yearbooks 2009-2013 |
| Germany | DEU | 12 | Eurostat, and OECD's SDBS Structural Business Statistics |
| Ghana | GHA | 1 | Ghana Statistical Service: National Industrial Census 2003 |
| Greece | GRC | 6 | Eurostat, and OECD's SDBS Structural Business Statistics |
| Greenland | GRL | 5 | Statistics Greenland: bank.stat.gl |
| Guadeloupe | GLP | 1 | Institut national de la statistique et des études économiques: Caractéristiques des entreprises et établissements |
| Guam | GUM | 4 | U.S. Census Bureau: County Business Patterns |
| Honduras | HND | 1 | Instituto Nacional de Estadística y Censos: Directorio de Establecimientos Económicos |
| Hong Kong | HKG | 13 | Census and Statistics Department: Annual Survey of Industrial Production, and www.statistics.gov.hk |
| Hungary | HUN | 11 | Eurostat, and OECD's SDBS Structural Business Statistics |
| India | IND | 1 | Central Statistics Office: 2005 Economic Census |
| Indonesia | IDN | 3 | Statistics Indonesia: Statistical Yearbook 2013 |
| Iran | IRN | 1 | Statistical Centre of Iran: Statistical Yearbook 1382 |
| Israel | ISR | 9 | Central Bureau of Statistics: www1.cbs.gov.il, Eurostat, and OECD's SDBS Structural Business Statistics |
| Italy | ITA | 12 | Eurostat, and OECD's SDBS Structural Business Statistics |
| Japan | JPN | 3 | Statistics Japan: Establishment and Enterprise Censuses 2001, 2004, 2006 |
| Jordan | JOR | 8 | Department of Statistics: www.dos.gov.jo |
| Korea | KOR | 9 | Statistics Korea: Censuses on Establishments 2007, 2009, 2011, 2012 |
| Kuwait | KWT | 10 | Central Statistical Bureau: Annual Surveys of Establishments 2002-2011 |
| Kyrgyzstan | KGZ | 1 | National Statistical Committee of Kyrgyz Republic: stat.kg |
| Laos | LAO | 1 | Lao Statistics Bureau: Economic Census 2006 |

Table 6: List of Countries and Sources

| Country | Code | Years | Sources |
|---|---|---|---|
| Latvia | LVA | 10 | Central Statistical Bureau of Latvia: www.csb.gov.lv, and Eurostat |
| Libya | LBY | 2 | Bureau of Statistics and Census Libya: bsc.ly |
| Liechtenstein | LIE | 6 | Statistical Office: Statistical Yearbooks 2007/2008, 2009-2012 |
| Lithuania | LTU | 7 | Eurostat |
| Luxembourg | LUX | 12 | Eurostat |
| Macau | MAC | 13 | Statistics and Census Service: www.dsec.gov.mo |
| Madagascar | MDG | 1 | Institut National de la Statistique: Rapport de l'enquete sur les Entreprises a Madagascar |
| Malaysia | MYS | 6 | Department of Statistics Malaysia: Economic Census 2011 Report |
| Maldives | MDV | 1 | Department of National Planning: Economic Survey 2007/2008 |
| Malta | MLT | 7 | Eurostat |
| Martinique | MTQ | 1 | Institut national de la statistique et des études économiques: Caractéristiques des entreprises et établissements |
| Mauritius | MUS | 2 | Statistics Mauritius: Censuses of Economic Activity 2002, 2007, Phases I and II |
| Mexico | MEX | 2 | Instituto Nacional de Estadstica y Geografía: Censos Economicos 2004, 2009 |
| Moldova | MDA | 8 | Statistica Moldovei: www.statistica.md |
| Monaco | MCO | 13 | Monaco Statistics: Observatoire de l'Economie 2012, 2013 |
| Mongolia | MNG | 2 | National Statistical Office of Mongolia: Monthly Bulletins of Statistics 2011, 2012 |
| Montenegro | MNE | 3 | Statistical Office of Montenegro: www.monstat.org |
| Morocco | MAR | 1 | Haut-Commissariat au Plan du Maroc: 2001-2 Economic Census |
| Nepal | NPL | 1 | Central Bureau of Statistics: Census of Manufacturing Establishments 2006/7, Survey of Small Manufacturing 2008/9 |
| Netherlands | NLD | 12 | Eurostat |
| New Caledonia | NCL | 3 | Institut de la Statistique et des Etudes Economique: Tableaux de L'Économie Calédonienne |
| New Zealand | NZL | 13 | Statistics New Zealand: www.stats.govt.nz |
| Nicaragua | NIC | 1 | Instituto Nacional de Información de Desarrollo: Urban Economic Census |
| Norfolk Island | NFK | 1 | Australian Business Statistics: www.ausstats.abs.gov.au |
| Northern Mariana Islands | MNP | 4 | U.S. Census Bureau: County Business Patterns |
| Norway | NOR | 8 | Eurostat, and OECD's SDBS Structural Business Statistics |
| Palau | PLW | 1 | Office of Planning and Statistics: 2012 - 2nd, 3rd Quarters Economic Indicators |
| Palestinian Territories | PSE | 7 | Palestinian Central Bureau of Statistics: Establishment Censuses 2004, 2007, 2012, and Comparison Study on Industrial Activities 1999-2004 |
| Panama | PAN | 1 | Instituto Nacional de Estadística y Censo: Preliminary Results of Economic Census 2012 |
| Paraguay | PRY | 1 | Direccin General de Estadística, Encuestas y Censos: National Economic Census 2011 |

Table 6: List of Countries and Sources

| Country | Code | Years | Sources |
|---|---|---|---|
| Peru | PER | 1 | Instituto Nacional de Estadística e Informática: IV Censo National Economico 2008 |
| Poland | POL | 12 | Central Statistical Office of Poland: Statistical Yearbook 2011, 2012, Eurostat, and OECD's SDBS Structural Business Statistics |
| Portugal | PRT | 11 | Eurostat, and OECD's SDBS Structural Business Statistics |
| Puerto Rico | PRI | 6 | U.S. Census Bureau: County Business Patterns |
| Qatar | QAT | 2 | Ministry of Development Planning and Statistics: Establishment Censuses 2008, 2010 |
| Réunion | REU | 1 | Institut national de la statistique et des études économiques: Caractéristiques des entreprises et établissements |
| Romania | ROU | 12 | Eurostat |
| Russia | RUS | 3 | Federal State Statistics Service: Industry of Russia 2008, 2009, 2011, and Small and Medium Businesses in Russia 2008, 2009, 2011 |
| Rwanda | RWA | 1 | National Institute of Statistics of Rwanda: Establishment Census - 2011 |
| San Marino | SMR | 8 | Ufficio Informatica, Tecnologia, Dati e Statistica: www.statistica.sm |
| São Tomé and Príncipe | STP | 2 | Instituto Nacional de Estatísticas de São Tomé e Príncipe: Business Statistics 2006, 2007 |
| Saudi Arabia | SAU | 1 | Central Department of Statistics and Information: 2010 Economic Census |
| Sierra Leone | SLE | 1 | Statistics Sierra Leone: Report of the Census of Business Establishments 2005 |
| Singapore | SGP | 10 | Department of Statistics Singapore: Census of Manufacturing Activities 2012 |
| Slovak Republic | SVK | 2 | Eurostat |
| Slovenia | SVN | 12 | Eurostat, and OECD's SDBS Structural Business Statistics |
| Spain | ESP | 12 | Eurostat, and OECD's SDBS Structural Business Statistics |
| Sri Lanka | LKA | 1 | Department of Census and Statistics - Sri Lanka: Census of Industry 2003/4 |
| Sudan | SDN | 1 | Central Bureau of Statistics: Statistical Year Book for the Year 2009 |
| Sweden | SWE | 12 | Eurostat, and OECD's SDBS Structural Business Statistics |
| Switzerland | CHE | 3 | Swiss Statistics: www.bfs.admin.ch/bfs/portal/en/index.html |
| Syria | SYR | 4 | Central Bureau of Statistics: www.cbssyr.sy |
| Taiwan | TWN | 3 | National Statistics: Industry, Commerce and Service Censuses 2001, 2006, 2011 |
| Thailand | THA | 2 | National Statistical Office: Industrial Censuses 2007, 2012 |
| Tonga | TON | 7 | Tonga Department of Statistics: Manufacturing Output, Employment and Wages/Salaries 2000-2003, 2001-2005, 2002-2006 |
| Trinidad and Tobago | TTO | 7 | Central Statistical Office: Business Establishments in T & T by Industry Economic Activity 2005-2007 |
| Tunisia | TUN | 12 | Institut National de la Statistique: Statistiques Issues du Répertoire des Entreprises |

Table 6: List of Countries and Sources

| Country | Code | Years | Sources |
|---|---|---|---|
| Turkey | TUR | 8 | Eurostat, and OECD's SDBS Structural Business Statistics |
| Uganda | UGA | 2 | Uganda Bureau of Statistics: Report on the Census of Business Establishments 2010/2011 |
| Ukraine | UKR | 3 | State Statistics Service of Ukraine: www.ukrstat.gov.ua |
| United Arab Emirates | ARE | 1 | National Bureau of Statistics: www.uaestatistics.gov.ae |
| United Kingdom | GBR | 12 | Eurostat, and OECD's SDBS Structural Business Statistics |
| United States | USA | 11 | U.S. Census Bureau: County Business Patterns, and Nonemployer Statistics |
| Uruguay | URY | 9 | Instituto Nacional de Estadística: Anuario Estadístico 2000-2012 |
| Venezuela | VEN | 1 | Instituto Nacional de Estadística: IV Censo Económico |
| Vietnam | VNM | 3 | General Statistics Office: Establishment Censuses 2002, 2007, and 2012 |
| Yemen | YEM | 2 | Central Statistical Organization: Results of Economic Surveys 2005-2006 |

# References


[1] Adamopoulos, Tasso and Diego Restuccia, "The Size Distribution of Farms and International Productivity Differences," *American Economic Review*, 104:6 (2014): 1667-97.

[2] Alfaro, Laura, Andrew Charlton, and Fabio Kanczuk, "Plant-Size Distribution and Cross-Country Income Differences," in *NBER International Seminar on Macroeconomics 2008*, ed. Jeffrey A. Frankel and Christopher Pissarides, (National Bureau of Economic Research, 2009).

[3] Baldwin, John R., Wulong Gu, Amélie Lafrance, and Ryan Macdonald, "Investments in Intangible Assets in Canada: R&D, Innovation, Brand, and Mining, Oil and Gas Exploration Expenditures," *Canadian Productivity Review*, 15-206-X, (Statistics Canada, 2009).

[4] Barseghyan, Levon and Riccardo DiCecio, "Entry Costs, Industry Structure, and Cross-Country Income and TFP Differences," *Journal of Economic Theory*, 146 (2011), 1828-1851.

[5] Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta, "Cross-Country Differences in Productivity: The Role of Allocation and Selection," *American Economic Review*, 103 (2013), 305-1334.

[6] Bhattacharya, Dhritiman, Nezih Guner, and Gustavo Ventura, "Distortions, Endogenous Managerial Skills and Productivity Differences," *Review of Economic Dynamics*, 16 (2013), 11-25.

[7] Bello, Omar, Juan Blyde, and Diego Restuccia, "Venezuela's Growth Experience," *Latin American Journal of Economics*, 48 (2011), 199-226.

[8] Bollard,Albert, Peter J. Klenow, and Huiyu Li, "Entry Costs Rise with Development," *manuscript*, (2014).

[9] Da-Rocha, Jose M., Marina Mendes, and Diego Restuccia, "Policy Distortions and Aggregate Productivity with Endogenous Establishment-Level Productivity," manuscript, (2014).

[10] Desmet, Klaus, and Stephen L. Parente, "The Evolution of Markets and the Revolution of Industry: A Unified Theory of Growth," *Journal of Economic Growth*, 17 (2012), 205-234.

[11] Gabler, Alain and Markus Poschke, "Experimentation by Firms, Distortions, and Aggregate Productivity," *Review of Economic Dynamics*, 16 (2013), 26-38.

[12] Gal, Peter N., “Measuring Total Factor Productivity at the Firm Level using OECD-ORBIS,” *OECD Economics Department Working Paper*, 1049 (2013).

[13] Guner, Nezih, Gustavo Ventura, and Yi Xu, “Macroeconomic Implications of Size-Dependent Policies,” *Review of Economic Dynamics*, 11 (2008), 721-744.

[14] Hopenhayn, Hugo, “Distortions, Firms and Aggregate Productivity,” presentation slides (2013).

[15] Hopenhayn, Hugo, and Richard Rogerson, “Job Turnover and Policy Evaluation: A General Equilibrium Analysis,” *Journal of Political Economy*, 101 (1993), 915-938.

[16] Hsieh, Chang-Tai and Peter J. Klenow, “Misallocation and Manufacturing TFP in China and India,” *Quarterly Journal of Economics*, 124 (2009), 1403-1448.

[17] Hsieh, Chang-Tai and Peter J. Klenow, “The Life Cycle of Plants in India and Mexico,” *Quarterly Journal of Economics*, forthcoming (2014).

[18] McGrattan, Ellen R. and Edward C. Prescott, “Technology Capital and the US Current Account,” *American Economic Review*, 100 (2010), 1493-1522.

[19] Melitz Marc J. and Giancarlo I. P. Ottaviano, “Market Size, Trade, and Productivity,” *Review of Economic Studies*, 75 (2008), 295-316.

[20] Moscoso Boedo, Hernan J., and Toshihiko Mukoyama, “Evaluating the Effects of Entry Regulations and Firing Costs on International Income Differences,” *Journal of Economic Growth*, 17 (2012), 143-170.

[21] Pagés-Serra, Carmen, *The Age of Productivity: Transforming Economies from the Bottom Up*, editor (Palgrave Macmillan, 2010).

[22] Parente, Stephen L. and Edward C. Prescott, *Barriers to Riches*, (MIT Press, 2000).

[23] Poschke, Markus, “The Firm Size Distribution Across Countries and Skill-Biased Change in Entrepreneurial Technology”, *manuscript*, (2014).

[24] Ranasinghe, Ashantha, “Impact of Policy Distortions on Firm-Level Innovation, Productivity Dynamics and TFP,” *Journal of Economic Dynamics and Control*, forthcoming (2014).

[25] Restuccia, Diego and Richard Rogerson, "Policy Distortions and Aggregate Productivity with Heterogeneous Establishments," *Review of Economic Dynamics*, 11 (2008), 707-720.

[26] Restuccia, Diego, "The Latin American Development Problem: An Interpretation," *Economía*, 13 (2013a), 69-100.

[27] Restuccia, Diego, "Factor Misallocation and Development," *The New Palgrave Dictionary in Economics*, Online Edition, (2013b).

[28] Rubini, Loris, "Innovation and the Trade Elasticity," *Journal of Monetary Economics*, 66 (2014), 32-46.

[29] World Bank, The, *Enterprise Surveys*, online (2014).

[30] Ziebarth, Nicolas L., "Are China and India Backward? Evidence From the 19th Century U.S. Census of Manufacturers," *Review of Economic Dynamics*, 16 (2013), 86-99.