

NBER WORKING PAPER SERIES

ON THE MEASURE OF DISTORTIONS

Hugo A. Hopenhayn

Working Paper 20404

<http://www.nber.org/papers/w20404>

NATIONAL BUREAU OF ECONOMIC RESEARCH

1050 Massachusetts Avenue

Cambridge, MA 02138

August 2014

The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2014 by Hugo A. Hopenhayn. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

On the Measure of Distortions
Hugo A. Hopenhayn
NBER Working Paper No. 20404
August 2014
JEL No. O11,O4,O47

ABSTRACT

Abstract The paper considers formally the mapping from distortions in the allocations of resources across firms to aggregate productivity. TFP gaps are characterized as the integral of a strictly concave function with respect to an employment-weighted measure of distortions. Size related distortions are shown to correspond to a mean preserving spread of this measure, explaining the stronger effects on TFP found in the literature. In general, the effect of correlation between distortions and productivity is shown to be ambiguous; conditions are given to determine its sign. An empirical lower bound on distortions based on size distribution of firms is derived and analyzed, revealing that substantial rank reversals in firm size are necessary for distortions to explain large TFP gaps. The effect of curvature on the impact and measurement of distortions is also considered.

Hugo A. Hopenhayn
UCLA
Department of Economics
Los Angeles, CA 90095
and NBER
hopen@econ.ucla.edu

1 Introduction

Among the factors explaining the disparity of aggregate productivity across countries, misallocation of resources across firms has been the focus of recent macro development literature (Alfaro et al. [2007], Banerjee and Duflo [2005], Guner et al. [2008], Hsieh and Klenow [2009b], Restuccia and Rogerson [2008], Buera et al. [2011], Bartelsman et al. [2013].) The key finding of these papers is that institutions and policies preventing the equalization of the marginal value of inputs across firms can potentially generate large losses in aggregate productivity. The benchmark models used in most papers (Hopenhayn [1992], Hopenhayn and Rogerson [1993], Jovanovic [1982], Lucas Jr [1978], Melitz [2003]) share a similar structure. So far, the results have been mainly quantitative and while motivated conjectures about the mapping between distortions and aggregate TFP, there are no theoretical results. The main objective of this paper is to fill this gap, providing a precise characterization of the link between the structure of these inter-firm distortions and aggregate productivity that is summarized in a measure of distortions.

The basic setting considered here has a set of firms producing a homogenous product using labor as the only inputs. Firms produce output with a homogenous production function that exhibits decreasing returns and with an idiosyncratic productivity. The optimal allocation of labor across firms was considered by Lucas Jr [1978]. It leads to an endogenous size distribution of firms after equating marginal product of labor and a simple expression for the aggregate production function. Aggregate productivity in this undistorted economy is a geometric mean of firm level productivities.¹

A series of papers have considered explicit policies or constraints that generate wedges in the allocation of resources across firms² as a source of misallocation. This was followed by very influential papers (see in particular Restuccia and Rogerson [2008], Guner et al. [2008], Hsieh and Klenow [2009a], Bartelsman et al. [2009]) that abstract from policies and considers

¹As shown in Hsieh and Klenow [2009a] there is an equivalence between a model of product differentiation Dixit and Stiglitz [1977], Melitz [2003] with curvature on the demand side and this homogenous good model.

²Barriers to the reallocation of labor resulting from firing costs were first considered in Hopenhayn and Rogerson [1993]. Financial constraints as a source of misallocation in Buera et al. [2011], Caselli and Gennaioli [2013], D’Erasmus and Moscoso Boedo [2012], Jeong and Townsend [2007], Midrigan and Xu [2013], Moll [2012], Quintin [2008], among others. Labor taxes and exemptions by Garicano et al. [2013], Gourio and Roys [2013], among others. Variable markups by Epifani and Gancia [2011], Peters [2013]. Information and learning are considered by Jovanovic [2014], David et al. [2014].

the quantitative effect of hypothetical barriers preventing the reallocation of labor represented by a joint distribution firm specific wedges and productivity. The main results suggested by this literature are: 1) Large distortions can lead to large effects on productivity; 2) More concentrated distortions have larger effects; 3) distortions that result in a reallocation of labor from establishments with higher TFP to those with lower TFP are more detrimental to productivity than those that inefficiently reallocate labor within size classes.

The analysis in these papers is purely quantitative and there is no theoretical analysis establishing these results. To develop some intuition, consider distortions that move employment from a set of firms to another set, starting at the efficient allocation. First note that small changes have a second order effect, regardless of the original size (proportional to TFP) of the firms involved since at the efficient allocation marginal productivities are equalized. Since first order effects are zero, the effects of the reallocation on aggregate TFP must depend on infra-marginal considerations. The first observation is that, fixing the number of workers reallocated if the reallocation involves small firms it will affect a larger proportion of their employment, digging deeper in the infra-marginal distortion. This is more damaging to aggregate TFP than a reallocation involving the same number of workers and firms, but the involved firms are larger (this would include reallocating those workers from large to small firms.) Second, for a given amount of total employment reallocated, the negative effect on productivity depends on the fraction of original efficient employment being reallocated (depth) and not on the specific source and destination. For example taking 10% of employment from one firm with 1000 employees is equivalent to taking 10% employment from 10 firms with 100 employees each.

These two observations lead to the following measure of distortions. Let n_i be the employment of a firm under the efficient allocation and $\theta_i n_i$ its distorted employment. The measure of distortions $N(\theta)$ counts the total fraction of aggregate original employment -regardless of source- that was distorted by θ . This measure is sufficient to derive the effects on aggregate productivity. Moreover, the ratio of TFP in the distorted economy to the undistorted level has a simple representation: $\int \theta^\alpha dN(\theta)$, where α is the degree of decreasing returns faced by firms. As $0 < \alpha < 1$, it follows immediately that mean preserving spread of this measure leads to lower productivity. The notion of a mean preserving spread can be interpreted as putting more employment mass at "larger distortions" and "concentrating" more the distortions. This also explains the quantitative results found in Restuccia and Rogerson [2008].

While the concentration of distortions leads to lower TFP, the correlation of wedges (the marginal product gap) and firm level productivity (or equally, size in the optimal allocation) has ambiguous effects, contrary to conjectures in the literature. The paper provides a necessary and sufficient condition so that increases in the correlation between wedges and productivity lead to decreases in TFP. This condition and examples that follow, suggest that while in highly distorted economies it is likely to hold, under plausible conditions increases in correlation can result in higher aggregate TFP.

As an application of our results, we provide an answer to the following question: given two size distributions of firms F and G , where F corresponds to an undistorted economy and G to a distorted one, what are the minimum set of distortions (in terms of their effect on TFP) that rationalize G ? The characterization reduces the problem to one of assortative matching with the simple solution of setting $\theta(n)$ so that $F(n) = G(\theta(n))$ for all n .³ I then apply this method to obtain lower bounds on distortions for India, China and Mexico taking the US as a benchmark. The effects on productivity according to this lower bound are meager: about 3.5% for India and Mexico and 0.5% for China. I also show that the size distributions generated by the distortions considered in Restuccia and Rogerson [2008] can be also rationalized with distortions that imply much smaller TFP decrease (e.g. 7% instead of 49%.) The difference is explained due to the presence of substantial degree of rank reversals in their simulations as well as in the calculated wedges in Hsieh and Klenow [2009a]. This result also suggests that without explicitly measuring distortions or deriving them from observed policies, there is not much hope of establishing large effects if they are only disciplined by consistency with measured size distributions.

Another important factor in determining the effect of distortions is the assumed curvature in the firm level production function (or demand), for which there is no general consensus. For instance, while Restuccia and Rogerson [2008] and many of the papers that follow take a value $\alpha = 0.85$, the analysis in Hsieh and Klenow [2009a] use an implied value $\alpha = 1/2$. I first establish that for given measure of distortions, the effects on productivity are zero at the extremes $\alpha = 0$ and $\alpha = 1$, so that the impact of α in the calculations is non-monotonic. I then consider the question of curvature in the context of the calculations carried in Hsieh and Klenow [2009a]. There are more subtleties to the analysis as given data on firms inputs and output, the implied productivities and calculated distortions also depend on α . Thus the measure of distortions also varies with α . Surprisingly, the effect of α is

³This is in line with the method used in Alfaro et al. [2007]

perfectly determined going monotonically from no productivity losses when $\alpha = 0$ to maximum losses for $\alpha = 1$. The proof is remarkable as it reduces the TFP ratio to a certainty equivalent and then uses standard analysis of risk aversion.

The paper is organized as follows. Section 2 sets up the benchmark model. Section 3 discusses the distorted economy. Section 4 develops the measure of distortions, derives the main Propositions and discusses its implications. Section 5 derives the comparative statics results on correlation of distortions. Section 6 derives the lower bound of distortions using the size distribution of firms and provides calculations of those bounds for a set of countries. Section 7 considers the role of curvature and Section 8 concludes.

2 Baseline model

This section describes a simple baseline model that will be used throughout the paper. The model is a simplified version of Hopenhayn and Rogerson [1993] that builds on Hopenhayn [1992] but without entry and exit, and closely related to Lucas Jr [1978] and Jovanovic [1982]. The production side of the economy is given by a collection of firms $i = 1, \dots, M$, with production functions

$$y_i = z_i n_i^\alpha,$$

where z_i is an idiosyncratic productivity shock for firm/establishment $i = 1, \dots, n$.⁴ Production displays decreasing returns ($\alpha < 1$) in the only input labor and total endowment in the economy N is supplied inelastically. Firms behave competitively taking prices as given. This economy has a unique competitive equilibrium $(\{n_i\}, w)$, where n_i is the profit maximizing input choice for firm i and labor market clears. The competitive equilibrium is also the solution to the planners problem:

$$\begin{aligned} & \max_{n_i} \sum_i z_i n_i^\alpha \\ & \text{subject to: } \sum n_i \leq N. \end{aligned}$$

The first order conditions for this problem imply that

$$\ln n_i = a_0 + \frac{1}{1 - \alpha} \ln z_i \tag{2.1}$$

⁴We won't make any distinctions here. The data used later is based on establishments.

where a_0 is a constant that depends on α , N and the vector of firm level productivities. Substituting in the production function,

$$\ln y_i = \ln z_i + \alpha \left(a_0 + \frac{1}{1-\alpha} \ln z_i \right) \quad (2.2)$$

$$= \alpha a_0 + \frac{1}{1-\alpha} \ln z_i \quad (2.3)$$

$$= \ln n_i - a_0 (1-\alpha) \quad (2.4)$$

is also proportional to z_i , implying that at the efficient allocation $y_i/n_i = y/n = a_0^{\alpha-1} \equiv a$ for all i .⁵ Finally, using the aggregate resource constraint to substitute for a , it follows that

$$y = \left(\sum_i z_i^{\frac{1}{1-\alpha}} \right)^{1-\alpha} N^\alpha.$$

This is an aggregate production function of the same class as the underlying firm-level production function, with TFP parameter given by $\left(\sum_i z_i^{\frac{1}{1-\alpha}} \right)^{1-\alpha}$.⁶ This technology exhibits decreasing returns in the aggregate, as firms here are treated as a fixed factor. This can be more clearly seen, dividing the first term by $M^{1-\alpha}$

$$y = \left(E z_i^{\frac{1}{1-\alpha}} \right)^{1-\alpha} M^{1-\alpha} N^\alpha. \quad (2.5)$$

This aggregate production function has constant returns to scale in firms and other inputs (in our example, labor), where aggregate TFP is a geometric mean of firm level productivity. Throughout the paper I refer to the term in brackets as TFP.⁷

3 The distorted economy

This section analyzes the consequences of deviations from the optimal allocation of resources across productive units. Figure 3.1 provides a useful picture of the type of distortions that might occur:

The solid line shows an optimal allocation, where $\ln n_i$ is a linear function of $\ln z_i$. The dots represent actual employment.

⁵This would obviously not hold when with a fixed cost in terms of overhead labor, as used in Bartelsman et al. [2013].

⁶Similar aggregation is given in Melitz [2003]

⁷In doing so, I treat firms as a form of capital. For most of the analysis the number of firms is constant, so this distinction is irrelevant.

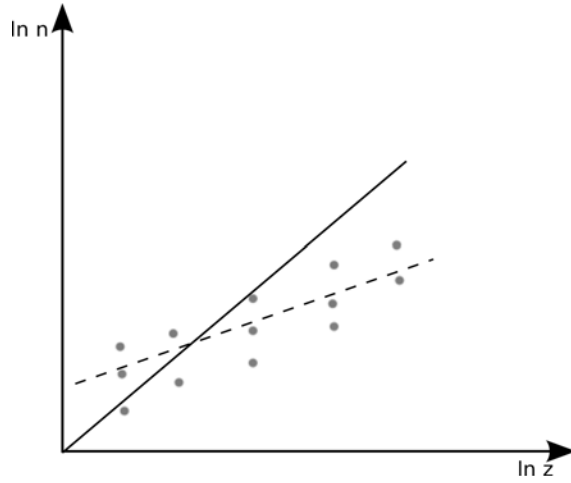


Figure 3.1: Wedges in marginal product

1. n_i not equal for all firms with the same z_i , termed *uncorrelated distortions*;
2. average $\ln n_i(z) \neq a + \frac{1}{1-\alpha}z$, termed *correlated distortions*, in the case of Figure 3.1 it is a distortion that results in reallocation of labor from more to less productive firms.

Both of these distortions result in losses of productivity as marginal product (or the marginal value of labor) is not equated across productive units. As an accounting device and following the literature, it is useful to model these distortions as firm-specific implicit taxes/subsidies that create a wedge between its revenues and output:

$$\begin{aligned} r_i &= (1 - \tau_i) y_i = (1 - \tau_i) z_i n_i^\alpha \\ &= \alpha (z_i (1 - \tau_i))^{\frac{1}{1-\alpha}}, \end{aligned}$$

where α is a constant that depends only on the equilibrium wage. Equilibrium in this economy will be identical in terms of allocations to the equilibrium of an undistorted economy where the distribution of firm productivities is changed to $z_i (1 - \tau_i)$. Total revenues are given by

$$r = N^\alpha M^{1-\alpha} \left(E [z_i (1 - \tau_i)]^{\frac{\alpha}{1-\alpha}} \right)^{1-\alpha} \quad (3.1)$$

and total output

$$y = \int y_i di = \int r_i (1 - \tau_i)^{-1} di$$

$$\frac{y}{r} = \frac{\int r_i (1 - \tau_i)^{-1} di}{\int r_i} = \frac{E(1 - \tau_i)^{-1} (z_i (1 - \tau_i))^{\frac{1}{1-\alpha}}}{E(z_i (1 - \tau_i))^{\frac{1}{1-\alpha}}}. \quad (3.2)$$

Using equations (3.1) and (3.2), it follows that

$$y = N^\alpha M^{1-\alpha} \frac{E(1 - \tau_i)^{-1} (z_i (1 - \tau_i))^{\frac{1}{1-\alpha}}}{\left(E(z_i (1 - \tau_i))^{\frac{1}{1-\alpha}}\right)^\alpha}. \quad (3.3)$$

3.1 Distortions and aggregate productivity: some examples

There is a general consensus in the literature (see for example Restuccia and Rogerson [2008], Guner et al. [2008], Hsieh and Klenow [2009a], Bartelsman et al. [2013]) that correlated distortions that implicitly tax high productivity firms and subsidize low productivity ones are more damaging to aggregate productivity than those that are uncorrelated to size. This section shows that under a homogenous production function, such presumption is not necessarily true.

I will first consider three examples that illustrate some of the key insights.

Example. There are two types of firms with productivities $z_1 = 1$ and $z_2 = 2$. Suppose there are 16 firms of each type, a total labor endowment $N = 2000$ and $\alpha = \frac{1}{2}$. It is easily verified that the optimal allocation requires $n_1 = 25$ and $n_2 = 100$ and total output $y = 400$.

Uncorrelated distortions for low productivity firms. Now suppose that 12 type 1 firms are excluded from production while 4 of them get 100 workers. This gives a feasible set of distortions as it is easily verified that total employment doesn't change and total output $y = 360$.

Uncorrelated distortions for high productivity firms. Assume instead that 3 type 2 firms are excluded from production while one of them gets 400 workers. This does not change aggregate employment and also gives total output $y = 360$.

Correlated distortions. Now assume 12 firms of type 1 are excluded from production while 1 firm of type 2 gets 400 workers. Again, this does not change aggregate employment and also total output $y = 360$.

What do all these examples have in common? In all cases employment in some firms is dropped to zero while for other firms it is multiplied by 4, relative to the efficient allocation. Moreover, in all cases the original amount of employment that is affected by each of these distortions is exactly the same. In the first case, the employment dropped to zero is that of 12 type 1 firms giving a total of $12 \times 25 = 300$. In turn, 4 of these firms had

employment quadrupled so the total employment affected by this distortion is $4 \times 25 = 100$. In the second case, the three firms of type 2 excluded from production represent an original total employment of 300 while the one firm whose employment quadrupled had 100 workers. It is easily verified that the same is true for the last case.

The examples suggest that what matters for total productivity is not what type of firms are hit by each distortion but the total original employment affected by them. In the next section we prove this conjecture, providing a general characterization of distortions.

4 The Measure of Distortions and Aggregate Productivity

For the undistorted economy, employment of each firm $n(z) = az^{\frac{1}{1-\alpha}}$, where a is a constant that depends on the total labor endowment N and the distribution of productivities. I will define a distortion as a ratio θ from actual employment to the undistorted one: $n = \theta n(z)$, where $\theta \geq 0$. (It is easy to see that this is proportional to a wedge $(1 - \tau) = \theta^{1-\alpha}$.) Distortions reallocate resources across firms, so they generate the same level of aggregate employment. This motivates the following definition:

Definition 1. A feasible distortion is a conditional probability distribution $P(\theta|z)$ such that

$$N/M = \int \theta n(e) dP(\theta|e) dG(e),$$

where N is employment allocated to production and M is the number of firms, so N/M is the average size of a firm. Since there is a one to one mapping between z and $n(z)$, it is more convenient to summarize these distortions, after the corresponding change of variables with a joint measure $\mu(\theta, n)$ with mass M and such that:

$$N = \int n d\mu(\theta, n) = \int \theta n d\mu(\theta, n)$$

where n is undistorted employment corresponding to $n(z)$ for some z .

For every $\hat{\theta}$ let

$$N(\hat{\theta}) = \int_{\theta \leq \hat{\theta}} n d\mu.$$

It is easy to see that this defines a measure on θ with the property that $N = \int dN(\theta)$ and by feasibility $N = \int \theta dN(\theta)$. Here $N(\hat{\theta})$ corresponds

to the total original (undistorted) employment affected by a distortion $\theta \leq \hat{\theta}$. Notice that it is silent about the productivity of the firms underlying these distortions.⁸

Consider now total output in the distorted economy:

$$y_d = M \int y(\theta, z) dP(\theta|z) dG(z).$$

Using

$$y(\theta, z) = z(\theta n(z))^\alpha = \theta^\alpha y(z) = \theta^\alpha a n(z)$$

where a is average labor productivity in the undistorted economy, it follows that:

$$\begin{aligned} y_d &= a \int \theta^\alpha n(z) d\mu(\theta, z) \\ &= a \int \theta^\alpha dN(\theta) \end{aligned}$$

Since a is average labor productivity in the undistorted economy, it follows that $y_u = aN$ so the ratio of TFP in the distorted economy to the undistorted one is:

$$\frac{TFP_d}{TFP_u} = \frac{\int \theta^\alpha dN(\theta)}{N} \quad (4.1)$$

which simply corresponds to integrating θ^α with the the normalized measure of distortions, i.e by the corresponding employment weights.⁹ As a corollary to this result, consider two subsets of firms with identical total employment. Assigning the feasible distortions $\theta < 1$ to one group of firms and $2 - \theta$ to the other one gives the same decrease in TFP regardless of the assignment.¹⁰

⁸Note that in the the three examples above this measure is given by:

$$\begin{aligned} N(\theta) &= 300 \text{ for } 0 \leq \theta < 1 \\ &= 1900 \text{ for } 1 \leq \theta < 4 \\ &= 2000 \text{ for } 4 \leq \theta. \end{aligned}$$

⁹A similar result can be obtained considering the the employment weighted distribution of wedges, given the proportionality of θ to $(1 - \tau)^{\frac{1}{1-\alpha}}$.

¹⁰This implies that in the analysis of Restuccia and Rogerson [2008] detailed in Section 5, the effect of correlation would be null if the firm quantiles for the calculation of correlated distortions were picked weighted by employment.

Connection to TFPR

There is a close connection between θ and TFPR. In our benchmark model, TFPR equal labor productivity, y_i/n_i . Let $a = y_i^o/n_i^o = y^o/n$ denote average labor productivity in the optimal allocation, where it is equated across firms. Then $TFPR_i = y_i/n_i = \theta^{\alpha-1} (y^o/n)$, so $\theta = (y_i/n_i)^{1/(\alpha-1)}$. An alternative formula that is closer to the one used in the literature (Hsieh and Klenow [2009a], Bartelsman et al. [2013]) can be derived exploiting this connection between θ and TFPR. Substituting in (4.1) it is easy to show that:

$$\frac{TFP^e}{TFP} = \left(\sum_i \frac{n_i}{n} \left(\frac{y_i/n_i}{y/n} \right)^{\frac{1}{1-\alpha}} \right)^{1-\alpha} \quad (4.2)$$

The corresponding formula for the monopolistic competition case is:

$$\frac{TFP^e}{TFP} = \left(\sum \frac{R_i}{R} \left(\frac{TFPR_i}{TFPR} \right)^{\frac{\alpha}{1-\alpha}} \right)^{\frac{1-\alpha}{\alpha}} \quad (4.3)$$

where R_i is the revenue of firm i , $TFPR_i = R_i / (k_i^{\hat{\alpha}} l_i^{1-\hat{\alpha}})$ and letting R be total revenue, $TFPR = R / (K^{\hat{\alpha}} N^{1-\hat{\alpha}})$.¹¹

4.1 Ordering distortions

Not all measure of distortions correspond to feasible distortions, for they need to be consistent with total employment.

Definition 2. A feasible measure of distortions is a measure $N(\theta)$ that integrates to N and such that

$$\int \theta dN(\theta) = N.$$

It follows immediately that a mean preserving spread of a feasible measure of distortions is also a feasible measure of distortions. Together with equation (4.1) this suggests a very natural order on measures of distortions given by second order stochastic dominance. Indeed, as the function under integration is concave, a mean preserving spread of a measure of distortions gives another

¹¹In the case of monopolistic competition, R_i is proportional to the input aggregator, e.g. $k_i^{\hat{\alpha}} n_i^{1-\hat{\alpha}}$, so if there is only labor R_i is proportional and can be replaced by n_i in the last formula.

measure of distortions with lower associated TFP. The intuition behind this result is that the infra-marginal effect of distortions is stronger the more deep and concentrated they are. The following corollary is a direct consequence of this observation.

Corollary 1. *The effect of uncorrelated distortions on TFP increases with the total employment of the group involved. In particular, holding fixed the number of firms affected, uncorrelated distortions to large firms are more detrimental to TFP than uncorrelated distortions to small firms.*

4.2 Generalization to more inputs

The above analysis generalizes easily to more inputs with a Cobb-Douglas specification. For exposition, we consider here the case of two inputs, n and k and production function $y_i = z_i n_i^\alpha k_i^\beta$. Letting $n(z)$ and $k(z)$ denote the optimal allocation and θ_L and θ_k the corresponding distortions, total output is:

$$y = M \int \theta_L^\alpha \theta_k^\beta z n(z)^\alpha k(z)^\beta dP(\theta_L, \theta_k | z) dG(z)$$

where $K/M = \int \theta_k k(z) dP(\theta_L, \theta_k) dG(z)$ and $N/M = \int \theta_L d(\theta_L, \theta_k) dG(z)$.

Using linearity between $k(z)$ and $n(z)$ it follows that total output

$$\begin{aligned} y &= aM \int \theta_L^\alpha \theta_k^\beta z \times z^{\frac{\alpha+\beta}{1-\beta-\alpha}} dP dG \\ &= a_0 \int \theta_L^\alpha \theta_k^\beta dN(\theta_L, \theta_K) \end{aligned}$$

for some constants a and a_0 that are independent of the θ 's, and consequently

$$\frac{TFP}{TFP_{eff}} = \frac{1}{N} \int \theta_L^\alpha \theta_k^\beta dN(\theta_L, \theta_K).$$

This is again an employment weighted measure integrating $\theta_L^\alpha \theta_K^\beta$, that is in turn homogeneous aggregator of distortions. In the particular case where $\theta_L = \theta_K = \theta$ this aggregator is $\theta^{\alpha+\beta}$ that setting $\alpha = \alpha + \beta$ is the same as the one obtained before. A caveat to this extension is that we are treating total capital as given. This in general is not the case, but as we will see is justified in the analysis of the next section.

4.3 Markups and distortions: an application

In a recent paper, Peters [2013] considers the impact of competition on aggregate TFP. He finds that the TFP gap depends on the marginal distribution of markups only and not the correlation between markups and firm productivity. This can be explained easily using the measure of distortions, as shown in this section.

The model is one of monopolistic competition. The set of products (and firms) is indexed by $[0, 1]$. Preferences are given by $\int \ln y_i di$. Production displays constant returns, $y_i = z_i n_i$, where z_i is the productivity of firm i . Let m_i denote the markup of firm i , so $p_i = (w/z) m_i$. Rearranging, this implies that

$$p_i y_i = p_i z_i n_i = w m_i n_i \quad (4.4)$$

and because of log preferences, this is equated across all products. In absence of distortions, $m_i = m^0$ for all i and thus employment is equated across firms to the aggregate (and average) labor endowment, that I normalize to one. Using (4.4), it follows that $\theta_i = (\int m_i^{-1} di / m_i)$ and $N(d\theta)$ is equal to measure of firms¹² with $\theta_i \leq \theta$ is fully determined by the distribution of markups.

5 Concentration and Correlation

In their seminal paper, Restuccia and Rogerson [2008] examine the potential effects of distortions in an economy similar to the one described here. The specification is similar to the one above with the addition of capital with production function $y_i = z_i n_i^{\tilde{\alpha}_1} k_i^{\tilde{\alpha}_2}$ where $\tilde{\alpha}_1 + \tilde{\alpha}_2 = \alpha < 1$. Restuccia and Rogerson [2008] consider taxes on output, so profits are of the form $(1 - \tau_i) y_i - w l_i - r k_i$, where τ_i denotes a sales tax. In their numerical exercises, τ_i takes two values $\tau_1 > 0 > \tau_2$ applied to two subsets of firms. The subsidy τ_2 is chosen so that total capital remains constant.¹³

The following table is taken from the simulations reported by Restuccia and Rogerson. The first two columns consider the case of uncorrelated distortions, where a fraction of establishments is taxed and the counterpart subsidized at a rate so that total capital stock is unchanged. The second pair of columns consider the case where the $x\%$ most productive establishments are taxed while the counterpart is subsidized, again at a rate that maintains the total capital stock unchanged.

¹²Recall that all firms' employment is the same, normalized to one, in the optimum.

¹³The government budget constraint is satisfied with the addition of lump-sum taxes.

Table 1: Uncorrelated and Correlated Distortions

% Estab. taxed	Uncorrelated		Correlated	
	τ_t		τ_t	
	0.2	0.4	0.2	0.4
90%	0.84	0.74	0.66	0.51
50%	0.96	0.92	0.80	0.69
10%	0.99	0.99	0.92	0.86

There are two distinguishing features of these distorted economies: 1) the larger the share of establishments taxed, the larger the negative effect on TFP and 2) Correlated distortions seem to have a larger effect than uncorrelated ones.¹⁴

To examine the first feature, take an uncorrelated distortion that sets a tax $\tau_t > 0$ to a fraction α of establishments while subsidizing the rest with $\tau_s < 0$. Consider the corresponding distortions $\theta_t < 1$ and $\theta_s > 1$. To preserve total employment (and use of capital) it must be that

$$\theta_s - 1 = \frac{\alpha}{1 - \alpha} (1 - \theta_t). \quad (5.1)$$

The corresponding measure of distortions is: $\{(\alpha N, \theta_t), ((1 - \alpha) N, \theta_s)\}$.¹⁵ An increase in α can be interpreted as a mean preserving spread of the original measure. To see this, take $\alpha' > \alpha$ and new measure $\{(\alpha' N, \theta_t), ((1 - \alpha') N, \theta'_s)\}$. It follows immediately that $\theta'_s > \theta_s$ and that

$$\frac{(1 - \alpha') \theta'_s + (\alpha' - \alpha) \theta_t}{1 - \alpha} = \theta_s.$$

So the new measure can be constructed by taking $(\alpha' - \alpha) N$ from the original mass at θ_s and assigning it to θ_t and mass $(1 - \alpha') N$ and assigning it to θ'_s , which as shown is mean preserving.

Consider now the second feature. Correlated distortions do in fact have larger effects in this setting, but not for the reasons claimed above. Indeed, our examples in Section 3.1 suggest that this need not be true. Even though

¹⁴"A key difference is that in this case the distortion is not to the size distribution of establishments of a given productivity, but rather to the distribution of resources across establishments of varying productivity." (Restuccia and Rogerson [2008])

¹⁵More precisely, it is the product of this measure since the same distortions apply to labor and capital.

correlated distortions move resources from establishments with higher TFP to those with lower ones, at the efficient allocation marginal productivities are equated and so the nature of marginal distortions does not matter. The reason why correlated distortions are more detrimental to productivity in the Restuccia and Rogerson simulations is that they hit a larger fraction of the population and result in a more dispersed measure of distortions. The analysis follows very similar lines to comparative statics with respect to α considered above. To preserve equality of total employment, the following must hold:

$$(\theta_s - 1) = \frac{N_t}{N_s} (1 - \theta_t),$$

where N_t corresponds to the total employment (at the efficient allocation) of establishments taxed and N_s of those subsidized. With correlated distortions where larger establishments are taxed and holding constant α , N_t/N_s will be higher. Following the same logic as above, the corresponding measure of distortions is a mean preserving spread of the distribution of uncorrelated distortions with the same α where N_t/N_s would be lower.

More general analysis

To stay closer to the literature, define a system of distortions by a joint probability distribution $Q(\tau, z)$ where the first marginal Q_τ is the unconditional distribution of (output) wedges and the second marginal Q_z is the distribution of firm productivity. For expositional purposes, suppose both supports are finite so Q is a matrix with typical element q_{ij} and let the elements of the respective supports be ordered, so $\tau_{i+1} > \tau_i$ and $z_{j+1} > z_j$. Following the literature on stochastic orderings (see Epstein and Tanny [1980], Meyer and Strulovici [2013], Tchen [1980]) define an increase in correlation as follows. Take $i = \{i_1, i_2\}$ and $j = \{j_1, j_2\}$ where $i_1 < i_2$ and $j_1 < j_2$. Suppose that $q_{ij} > 0$ for all four pairs i, j in this set. Define a new matrix \tilde{Q} that is identical to Q with the exception that $\tilde{q}_{i_1 j_1} = q_{i_1 j_1} + \varepsilon$, $\tilde{q}_{i_2 j_2} = q_{i_2 j_2} + \varepsilon$ and $\tilde{q}_{i_1 j_2} = q_{i_1 j_2} - \varepsilon$, $\tilde{q}_{i_2 j_1} = q_{i_2 j_1} - \varepsilon$. This *correlation increasing transformation* is represented in Figure 5.1 for points $\{a = (z_1, \tau_1), b = (z_2, \tau_1), c = (z_1, \tau_2), d = (z_2, \tau_2)\}$. The arrows indicate a shift of ε probability from b to a and c to d , a local increase in association between τ and z . For joint distributions with finite support this defines a bivariate *correlation* (more precisely *concordance*) ordering, where $\tilde{Q} \succeq Q$ if and only if the former can be obtained by a finite number of correlation

increasing transformations of the latter.¹⁶

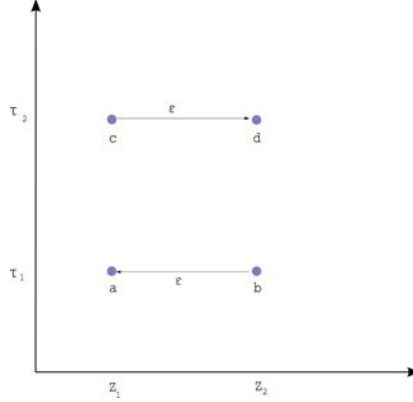


Figure 5.1: Increase in correlation

To evaluate the effect of increased correlation on TFP, it is useful to examine its effect on the measure of distortions. With the aid of Figure 5.1, it is useful to decompose this effect in two steps. Let θ_1 and θ_2 be the distortions corresponding to τ_1 and τ_2 respectively, where it easily follows that $\theta_2/\theta_1 = ((1 - \tau_2)/(1 - \tau_1))^{1/1-\alpha} < 1$. Let $n_2 > n_1$ be the efficient employment levels associated to z_1 and z_2 , respectively. As a first step, the shift of ε probability from b to a and from c to d imply a decrease of $\delta = (n_2 - n_1)\varepsilon$ in the employment measure of θ_1 and a corresponding increase in employment associated to θ_2 . But given that $\theta_2 < \theta_1$, this shift implies a reduction in total employment. As wages decrease to reestablish the equilibrium, all θ 's increase proportionately. The effect of positively correlated distortions (again, adhering to the standard definition this means larger wedges for higher levels of productivity) is to move employment from higher to lower levels of θ while increasing proportionately employment elsewhere. Consider the case of a finite set of distortions $\theta_1 > \theta_2 > \dots > \theta_{N-1} > \theta_N$ with corresponding employment measures n_1, \dots, n_N . An increase in correlation can be represented by a new vector $n' = n_1, n_2, \dots, n_k - \delta, \dots, n_m + \delta, \dots, n_N$. As indicated above at the original θ 's total employment would be short of the endowment, so a rescaling is necessary. More specifically, the new vector

¹⁶For this definition, see Epstein and Tanny [1980] and Tchen [1980]. The order is easily extended to all integrable functions by taking limits. It coincides with the ordering induced by using supermodular functions as a basis for comparison and the bivariate ordering of the corresponding cdf's $F_Q(\tau, z) \leq F_{\tilde{Q}}(\tau, z)$ for all (τ, z) . It is also implied that the correlation between τ and z is greater for \tilde{Q} .

$\theta'_j = \theta_j / \sum \theta_i n'_i = \theta_j / (\sum \theta_i n_i + (\theta_m - \theta_k) \delta)$ and the tfp ratio given by:

$$\sum_j \left(\frac{\theta_j}{\sum \theta_i n_i + (\theta_m - \theta_k) \delta} \right)^\alpha n'_j$$

Taking the derivative with respect to δ evaluated at $\delta = 0$ and using $\sum \theta_i n_i = 1$ gives:

$$\partial TFP_{ratio} / \partial \delta = \theta_m^\alpha - \theta_k^\alpha - (\theta_m - \theta_k) \alpha \sum_i \theta_i^\alpha n_i \quad (5.2)$$

This proves the following:

Proposition 1. *Take an initial joint distribution $Q(\tau, z)$ of wedges with finite support. Let $\tilde{Q}(\tau, z)$ be derived from Q with a correlation increasing transformation on $(\tau_1, z_1) \times (\tau_2, z_2)$ where $\tau_1 < \tau_2$ and $z_1 < z_2$. The TFP ratio of \tilde{Q} is lower than the TFP ratio of Q if and only if*

$$-(\theta_1^\alpha - \theta_2^\alpha) + (\theta_1 - \theta_2) \alpha \sum_i \theta_i^\alpha n_i \leq 0 \quad (5.3)$$

where $\theta_i = (1 - \tau_i)^{1/1-\alpha}$.

Inequality 5.3 highlights two countervailing forces. On the one hand, there is the shift of employment away from higher to lower θ with a marginal negative effect on TFP of the order $\theta_1^\alpha - \theta_2^\alpha$, represented by the first term. On the other hand, this shift of employment to lower θ frees up employment that is reallocated uniformly across distortions in proportion to their current employment. The second term can be written as $(\theta_1 - \theta_2) \sum_i \alpha \theta_i^{\alpha-1} (\theta_i n_i)$. The first term in brackets represent employment resources freed up. The second term is the marginal effect on distortions (weighted by employment) and is proportional to the marginal product of labor being reallocated.¹⁷

Corollary 2. *Under the assumptions of Proposition 1, a sufficient condition for the increase in correlation to lower TFP is that $\theta_1 \leq 1$.*

Proof. Since $\theta_1 > \theta_2$, by concavity it follows that $\theta_1^\alpha - \theta_2^\alpha \geq \alpha \theta_1^{\alpha-1} (\theta_1 - \theta_2)$. Given $\theta_1 \leq 1$ and $\sum_i \theta_i^\alpha n_i < 1$, it follows that $\theta_1^\alpha - \theta_2^\alpha \geq \alpha (\theta_1 - \theta_2) > (\theta_1 - \theta_2) \alpha \sum_i \theta_i^\alpha n_i$. Applying Proposition 1, the Corollary is proved. \square

¹⁷While this Proposition considers a local change between two pair of distortions, an arbitrary directional derivative can be obtained aggregating local changes. An interesting case occurs when (θ, n) is joint-log normally distributed, where the two effects described above cancel out exactly. This is the reason why the degree of correlation of distortions has no impact in that case, as Hsieh and Klenow [2009a] find.

While Corollary 2 gives sufficient conditions for an increase in the correlation of distortions to lower TFP, it is of limited use since it requires that $\theta_1 \leq 1$ so it is restricted to shifts in correlation between two distortions with above average wedges. Indeed, starting from a minimally distorted economy (e.g. $\sum_i \theta_i^\alpha n_i$ close to one), applying a similar argument it is easy to show that if $\theta_2 > 1$ (i.e. taking points with wedges below average) an increase in correlation will actually raise TFP. The following examples show the plausibility of both cases.

Example 1. $\alpha = 0.5$, $N = 2$, $\theta_1 = 2$, $n_1 = 0.5$, $\theta_2 = 0$, $n_2 = 0.5$. This could be the result of a baseline uncorrelated distortion where a share of firms comprising 50% of employment face a positive wedge that essentially shuts them down. All employment is absorbed by the remaining firms, doubling their employment relative to the undistorted case. These distortions lower TFP by 29.3%. Substituting in equation (5.2), $\partial TFP_{ratio}/\partial \delta = -0.707$. More generally, whenever $\theta_2 = 0$ this will be negative since $\partial TFP_{ratio}/\partial \delta = -\theta_1^\alpha + \alpha \theta_1 (\theta_1^\alpha n_1) = \theta_1^\alpha (\alpha - 1)$ where the equality follows from $\theta_1 n_1 = 1$. It is clear in this case that making wedges more positively correlated (i.e. increasing n_1) lowers TFP while making them negatively correlated (increasing n_2) increases TFP.

The previous example is in line with the prevailing view that correlated wedges are more damaging to TFP. The next example shows the opposite can also occur.

Example 2. $\alpha = 0.5$, $N = 2$, $\theta_1 = 3$, $n_1 = 0.2$, $\theta_2 = 0.5$, $n_2 = 0.8$. This could be the result of a baseline uncorrelated distortion where a share of firms that comprise 80% of employment is subject to a positive wedge in marginal products that decreases their employment by 50% while the remaining firms face no wedge yet the fall in wages increases their employment by 200%. These distortions lower TFP by 8.8%. Substituting in equation (5.2), $\partial TFP_{ratio}/\partial \delta = 0.115$. Hence starting from this baseline and making wedges positively (negatively) correlated with productivity would increase (decrease) TFP. For instance, decreasing n_2 to 0.7 and increasing n_1 to 0.3 -as would result more negatively correlated wage- lowers the TFP gap to 9.25%, an extra 0.425% relative to the baseline. Likewise, increasing the employment share of firms with a positive wedge to 0.9 -a *more positively* correlated wedge- results in a TFP gap of only 6.5%, an improvement of 2.3% relative to the baseline.¹⁸

¹⁸As the shares of employment are modified, the θ 's are rescaled as explained above so the employment resource constraint is satisfied. In the first case, $\theta_2 = 0.4$ and $\theta_1 = 2.4$ while in the second case $\theta_2 = 2/3$ and $\theta_1 = 4$.

While the effect of correlation is in general indeterminate, inequality (5.3) suggests that starting from a high level of initial distortions, i.e. a low TFP ratio $\sum_i \theta_i^\alpha n_i$, an increase in correlation will decrease TFP.

6 Distortions and the size distribution of firms

One feature of underdeveloped economies is a different size distribution, in particular a large fraction of employment in small firms. How much do these differences in the size distribution of firms explain the TFP gap? What is the role played by distortions?

There are obvious identification problems with this approach. Even if the economies compared had the same distribution of firm level productivities, the mapping between distortions and size distribution is not invertible. As an example, the same size distribution for an undistorted economy can be obtained by another one where only the least productive firms produce with a distribution of wedges that generates this size distribution. However, as shown below it is very straightforward to find a lower bound on distortions under some identifying assumptions.¹⁹

Let $Q(\theta, n)$ be a joint measure on (θ, n) . The interpretation is as follows: if the economy were undistorted, the size distribution of firms would be the marginal on n :

$$F(n) = \int_{z \leq n} dQ(\theta, z) \quad (6.1)$$

while the actual size distribution

$$G(n) = \int_{\theta z \leq n} dQ(\theta, z). \quad (6.2)$$

This joint distribution has an associated measure of distortions $N(\theta) = M \int n Q(\theta, dn)$. The lower bound on distortions is obtained as the solution to the following program:

$$\max_{Q(\theta, n)} \int \theta^\alpha n Q(d\theta, dn) \quad (6.3)$$

subject to (6.1) and (6.2), where F is a given distribution of efficient establishment size and G the actual size distribution.

The following Proposition characterizes the solution to this problem, but it is very intuitive. For continuous size distributions the solution is for each

¹⁹A related procedure is used in Alfaro et al. [2007]

n a point mass at $\theta(n)$ so that $F(n) = G(\theta(n)n)$. Letting $h(n) = \theta(n)n$ this function provides an assortative match between efficient and actual firm size by matching the percentiles of the corresponding distributions.

Proposition 2. *Suppose F and G are continuous distributions. The solution to (6.3) is given the the joint measure $Q(\theta, n)$ that puts all mass on the graph of $\theta(n)$ where $F(n) = G(\theta(n)n)$ and $dQ(\theta(n), n) = dF(n)$.*

Proof. We show the solution to the above problem can be cast as an optimal matching problem. Any pair (θ, n) can be also represented by (m, n) where $m = \theta n$. Hence for any joint measure $Q(d\theta, dn)$ there exists a corresponding measure $P(dm, dn)$ with first marginal G and second marginal F . Rewrite (6.3) as:

$$\max_{P(dm, dn)} \int \left(\frac{m}{n}\right)^\alpha n P(dm, dn)$$

subject to:

$$\begin{aligned} G(dm) &= P(dm, N) \\ F(dn) &= P(M, dn) \end{aligned}$$

where N and M are the support of F and G , respectively. This is an assignment problem with match-return function $u(m, n) = m^\alpha n^{1-\alpha}$. Since this function is supermodular, the solution is perfectly assortative matching, so that $m(n)$ satisfies $F(n) = G(m(n)) = G(\theta(n)n)$ where $\theta(n) = m(n)/n$. \square

The above procedure would require to know, in addition to the actual size distribution of firms, the hypothetical size distribution for that economy in absence of distortions. This can be done (with somewhat strong assumptions) by benchmarking that economy with an undistorted one under the following identifying assumptions:

Assumption 1. (a) *The benchmark economy is undistorted.* (b) *The underlying distribution of productivities for both economies is the same.*

The algorithm to derive the bound on distortions can be easily explained as follows. Let F denote the cdf for size distribution in the reference economy and G the one in the presumably distorted one and for now assume that average size \bar{n} is the same in both. For $x \in [0, 1]$ define the corresponding quantiles $n(x)$ and $m(x)$ by $F(n(x)) = x$ and $G(m(x)) = x$. By Proposition 2 we can write:

$$\frac{TFP^d}{TFP^u} = \frac{1}{\bar{n}} \int m(x)^\alpha n(x)^{1-\alpha} dx.$$

Scaling

There are large differences in the average size of firms (or number of firms per capita) across countries and this needs to be taken into account in the calculations. Take for instance the size distributions of India and US as the benchmark economy as shown in the left side of Figure 6.1.²⁰ The average size of a US firm is approximately 270 while in India it is only 50 and that is apparent from the strong stochastic dominance observed in this figure. To isolate the effect of the number of firms from distortions, we focus on the question: how far is the distorted economy from its own frontier? This requires *scaling down* the size distributions of firms in the reference economy.

More precisely, letting \bar{n}_d denote average firm size in the distorted economy and \bar{n}_u in the undistorted one, a firm that is of size n in the undistorted economy (e.g. US) would have been of size γn in the distorted one (e.g. India) in the absence of distortions, where $\gamma = \bar{n}_d/\bar{n}_u$. This adjustment is done in the right panel of Figure 6.1. Once this adjustment is made, it shows that India's size distribution is compressed relative to that of the US.

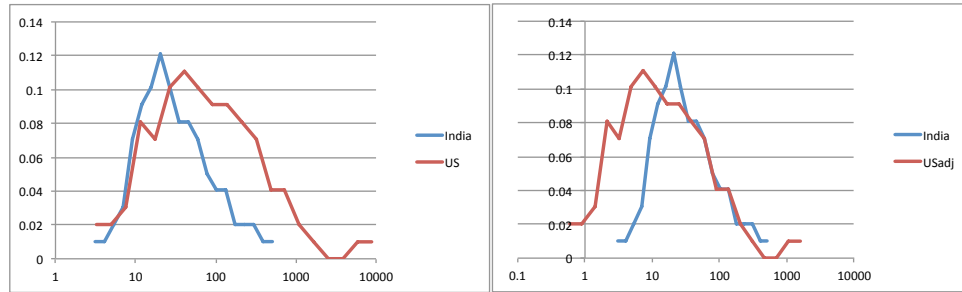


Figure 6.1: Size distribution of Firms: India and US

Let F denote the size distribution of the benchmark economy and G the that of the distorted one. Define the hypothetical size distribution of the distorted economy in absence of distortions \tilde{F} by $\tilde{F}(n) = F(\gamma n)$. Our upper bound is constructed setting $G(m(n)) = \tilde{F}(n)$ and computing:

$$\frac{TFP_d}{TFP_u} = \frac{1}{\bar{n}_d} \int m(x)^\alpha (\gamma n(x))^{1-\alpha} dx.$$

The identifying assumptions are strong, but I believe they can be weakened. In particular, I conjecture that (b) can be weakened to assuming that

²⁰This data has been graciously provided to me by Hsieh and Klenow and corresponds to the one used in their paper. In the case of India, only formal firms are included.

the distribution of productivities of the distorted economy is dominated by that of the distorted one. Differences in productivities would arise from both, difference in underlying productivities together with distortions. A firm could be small in the distorted economy either because its employment is less than optimal or because its productivity is lower. For a given level of employment, the firm's output would be higher in the latter case. This suggests that to explain a lower size distribution in the distorted economy, it is less damaging to the TFP of that economy if this comes from distortions rather than lower distribution of firm productivities. This would imply that assuming both countries have the same distribution of firm productivities provides an upper bound for TFP_d/TFP_u .

6.1 Example: Pareto distributions

Consider two economies a benchmark economy b and a distorted economy d that have Pareto size distributions with parameters (x_b, b) , (x_d, d) , respectively, and mean employments $\bar{n}_b = \frac{bx_b}{b-1}$ and $\bar{n}_d = \frac{dx_d}{d-1}$ satisfying the above assumptions, so $1-F(n) = \left(\frac{n}{x_b}\right)^{-b}$ and $1-G(n) = \left(\frac{n}{x_d}\right)^{-d}$ and $\gamma = \frac{bx_b}{dx_d} \frac{d-1}{b-1}$. Define $\tilde{m}(n)$ by $F(n) = G(\tilde{m}(n))$:

$$\tilde{m}(n) = x_d \left(\frac{n}{x_b}\right)^{b/d}.$$

After some calculations, the bound on relative TFP is :

$$\frac{TFP_d}{TFP_d^c} = \frac{\left(\frac{b}{d} \frac{d-1}{b-1}\right)^\alpha (1-b)}{1 + \frac{b}{d}\alpha - b - \alpha}$$

which is independent of the scaling parameters x_b and x_d . The following table gives the TFP ratios for some values of b and d .

Table 2: TFP ratios for Pareto case

b	d	TFP ratio
1.1	1.2	0.96
	1.5	0.82
	2	0.72
1.5	2	0.98
	3	0.94

6.2 Application: Bounds for some countries

This section computes the TFP bound for three economies: India, China and Mexico.²¹ The benchmark size distribution is the one corresponding to the US. Figure 6.2 provides on the left panel the size distribution of firms for the four countries. The right panel gives the distributions adjusted to the average firm size of the US (the average sizes are as follows: Mexico=15, India=50, US=272, China=558.) It is worth noting that the normalized distribution for China is close to the one for the US, while those of Mexico and India are similar to each other but more compressed than the US.

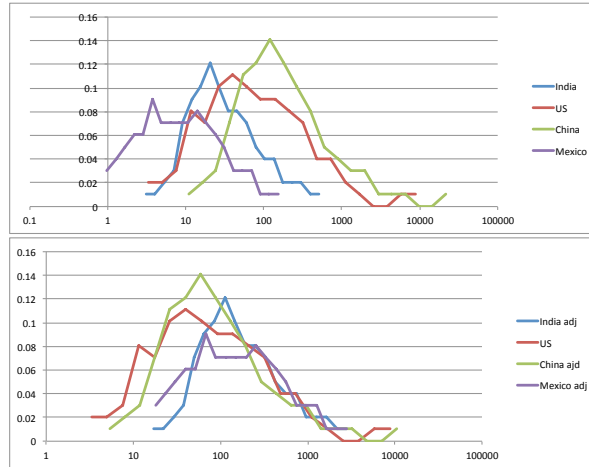


Figure 6.2: Size Distributions: India, China, Mexico and US

Figure 6.3 plots the corresponding measures of distortion. Recall that an undistorted economy corresponds to a point mass measure at one. China's measure is very close to undistorted, while India and Mexico's measures, being very similar to each other, are a clear mean preserving spread of China's. The corresponding TFP ratios are as given in Table 3, giving the above calculation for different values of α . As was apparent from before, China appears as almost undistorted while the TFP losses for India and Mexico are relatively small, especially when taking $\alpha = 0.85$ which is one of the standard value used in the literature. The value of $\alpha = 0.5$ is consistent with the markup values used by Hsieh and Klenow [2009a].

These bounds on distortion losses are very small when compared to what

²¹As mentioned earlier, data for India corresponds to formal firms only. When the calculation is done to include informal firms, result change very little.

Table 3: TFP ratios

	$\alpha = 1/2$	$\alpha = 2/3$	$\alpha = 0.85$
China	0.991	0.992	0.995
India	0.928	0.937	0.964
Mexico	0.931	0.939	0.966
R&R	0.851	0.873	0.929
USA mean	0.655	0.693	0.817

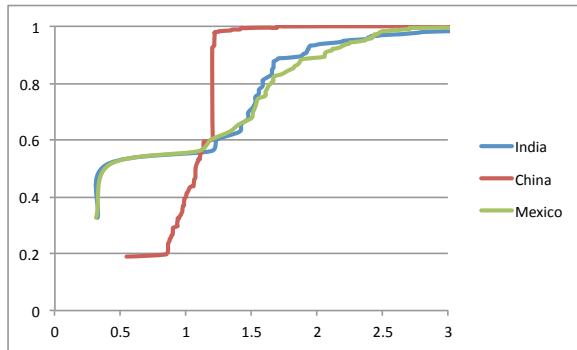


Figure 6.3: Measures of distortion

has been suggested in the literature.²² To put these results in perspective, I consider the most extreme hypothetical case considered in Restuccia and Rogerson [2008], where the top 90% firms are taxed away 40% of their output giving a TFP ratio of 0.51. Using $\alpha = 0.85$ as done in their paper, the implied measure of distortions is $\{(0.033, 90\%), (470, 10\%)\}$ as depicted in the left panel of Figure 6.4.²³ These distortions give rise to a substantial spread in the size distribution and a distance to the US undistorted distribution that is much larger than that of India, as shown in Figure 6.5. Using this size distribution, we can now calculate our lower bound of distortion. The corresponding measure is shown in the right panel of Figure 6.4. It is considerably more dispersed than the one obtained for India but orders of

²²It is worth noticing that Alfaro et al. [2007], using a similar procedure find significant effects. There is a key difference with our calculations in that while in our analysis we exclude the impact of the number of firms from the TFP term (see equation 3.3), they include that term. I believe that to a large extent this explains the difference.

²³The first term of this measure is obtained setting $\theta = (1 - 0.4)^{1/(1-0.85)}$ and the second term is obtained to keep total employment constant, as done in Restuccia and Rogerson [2008].

magnitude less dispersed than the actual measure of distortions as shown in the left panel of the figure. As a consequence, this lower bound on the measure of distortions is considerably less damaging to TFP. The fourth row in Table 3 gives the corresponding TFP ratios. Compare, for example, the bound for $\alpha = 0.85$ that gives a 7% decrease in TFP to the one obtained in Restuccia and Rogerson [2008] which is almost 50%.

Figure 6.4: Measure of Distortions in Restuccia/Rogerson

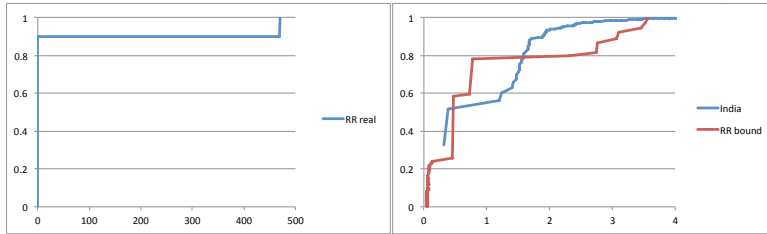
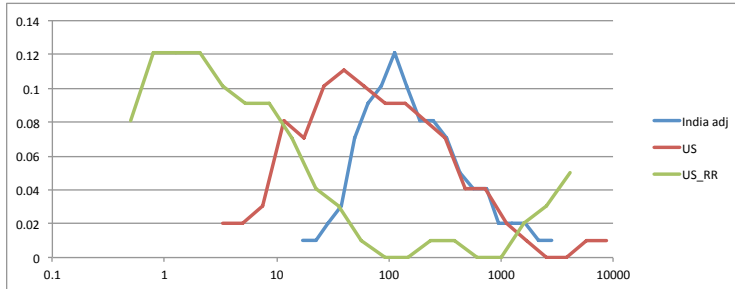


Figure 6.5: Size Distribution with Distortions in Restuccia/Rogerson



Our bounds on distortions are also very small when compared to Hsieh and Klenow [2009a], derived from establishment level data for China and India, that find a TFP ratios in the order of 45%. Table 4 provides some statistics of the dispersion in θ 's found in Hsieh and Klenow [2009a] and in our calculation.²⁴ All measures of dispersion are orders of magnitude higher in their data.²⁵

²⁴The dispersion of $\ln\theta$'s is equivalent to dispersion in $\frac{\ln TFP R}{1-\alpha}$ reported in Hsieh and Klenow [2009a]

²⁵Incidentally, these measures of dispersion are taken across establishments without considering employment weights, which as seen in Section 4 are relevant for the derivation of TFP ratios. Indeed the same establishment level dispersion of $\ln\theta$'s can give rise to very different TFP ratios depending on the employment weights, ranging from zero (all employment on $\theta = 1$) to all employment at the two extremes. In our case, the latter

Table 4: Dispersion of $\ln \theta'$ s

Percentiles	India (94)		China (98)	
	H-K	Bound	H-K	Bound
SD	4.47	0.53	4.93	0.31
75-25	5.4	0.73	6.27	0.36
90-10	10.7	1.14	12.4	0.75

Our method disciplines the calculations of policy distortions with the size distribution of different countries. But at the same time, ours is a lower bound which is only attained when distortions do not lead to rank reversals in firm size. These rank reversals inevitably occur both for the correlated and uncorrelated distortions considered in Restuccia and Rogerson [2008] and in the calculated distortions in Hsieh and Klenow [2009a].²⁶ It follows that *for distortions to really matter quantitatively, they must lead to substantial degree of large rank reversals*. In particular, policies that do not lead to rank reversals will not do the job. Examples of such kinds of policies are widespread, such as tax exemptions as the ones in France discussed in Gourio and Roys [2013] and Garicano et al. [2013]; direct subsidies to small firms; and restrictions to the operation of large retail stores. Indeed these papers find that while these policies have notable impact on the size distribution of firms, they have almost no effect on aggregate TFP.²⁷

7 On the impact of curvature

This section considers the impact of curvature -the degree of decreasing returns at the firm level- on the analysis of distortions. Recall the representation of TFP derived in Section 4

gives approximately 0.86 for India and 0.92 for China. These numbers come down to 0.76 and 0.86 when using $\alpha = 0.5$ as in Hsieh and Klenow [2009a] instead of $\alpha = 0.85$.

²⁶To illustrate this point, consider the extreme case of correlated distortions analyzed in Section 5: a firm with 2 employees in the undistorted economy will have approximately 1,000 in the distorted one and a firm with an original employment of 9,000 employees ends up with less than 300!

²⁷There is a related finding in Guner et al. [2008], that simulate the effect of correlated distortions for OECD countries. They consider labor and capital taxes with exemptions for firms below a size threshold, disciplined to match average OECD firm sizes. While they find considerable impact on capital accumulation, the effects of misallocation on aggregate TFP are minimal. Again, the distortion considered preserves rank.

$$\frac{TFP_d}{TFP_e} = \frac{\int \theta^\alpha dN(\theta)}{N}.$$

By Jensen's inequality, for any $0 < \alpha < 1$ this is strictly less than one. For $\alpha = 1$ it equals one for it must integrate to employment. But also when $\alpha = 0$ this ratio is equal to one. Hence, the relationship between the TFP gap and curvature is not monotonic, for a fixed measure of distortion.

The caveat to this analysis is that we consider fixed the measure of distortions, while this might also be affected by α . This follows from the fact that the optimal distribution of employment across firms is a function of α : as α increases, employment becomes more concentrated in large firms and whether this results in a smaller or larger TFP gap depends on the distribution of distortions. If distortions are obtained as the result of firm level output wedges $(1 - \tau_i)$ as in Restuccia and Rogerson [2008], using equation (3.3) it is straightforward to see that the TFP gap disappears as $\alpha \rightarrow 0$. On the other extreme, the results are ambiguous and might depend on the nature of distortions: with uncorrelated distortions, output will still be concentrated in one firm with highest productivity but if they are positively correlated, it will not.

When distortions are uncovered from the data as in Hsieh and Klenow [2009a] there is an additional reasons why curvature will matter in the calculations: both, the distribution of TFP and implicit distortions (i.e. wedges) depend on α . Interestingly enough, a sharp result emerges in this case as detailed below.

A stylized version of the procedure followed by Hsieh and Klenow [2009a] is as follows. The data consists of establishment levels of inputs and outputs:

$$(n_1, y_1, n_2, y_2, \dots, n_M, y_M)$$

where M is the number of establishments. Using this data and a production function of the form $y_i = z_i n_i^\alpha$, we can solve for the vector of productivities (z_1, z_2, \dots, z_M) and compute the counterfactual efficient level of output.

As shown in Section 2, aggregate TFP in the undistorted economy is:

$$TFP_e = \left(\sum_i z_i^{\frac{1}{1-\alpha}} \right)^{1-\alpha} N^\alpha.$$

Substituting $z_i = y_i/n_i^\alpha$ gives:

$$TFP_e = \left(\sum \left(\frac{y_i}{n_i^\alpha} \right)^{\frac{1}{1-\alpha}} \right)^{1-\alpha}$$

and dividing by actual TFP in this economy y/n^α gives:

$$\frac{TFP_e}{TFP} = \left(\sum \left(\frac{\frac{y_i}{n_i}}{\frac{y}{n^\alpha}} \right)^{\frac{1}{1-\alpha}} \right)^{1-\alpha} = \left(\sum \frac{n_i}{n} LPR_i^{\frac{1}{1-\alpha}} \right)^{1-\alpha} \quad (7.1)$$

where $LPR_i = \frac{y_i/n_i}{y/n}$ stands for labor productivity ratio. From equation (7.1) it follows immediately that:

$$\left(\frac{TFP_e}{TFP} \right)^{\frac{1}{1-\alpha}} = \sum \frac{n_i}{n} LPR_i^{\frac{1}{1-\alpha}}. \quad (7.2)$$

Equation (7.2) expresses the TFP ratio as the certainty equivalent of the lottery $\{(LPR_1, \frac{n_1}{n}), (LPR_2, \frac{n_2}{n}), \dots, (LPR_M, \frac{n_M}{n})\}$ under utility function $u(x) = x^{\frac{1}{1-\alpha}}$. Note precisely because these preferences are *risk loving* they imply a TFP coefficient ratio greater than one. An increase in α implies more risk loving and hence higher TFP_e/TFP , so the TFP gap increases with α . At the extreme, when $\alpha = 0$ utility is linear and there is no TFP gap. In the other extreme, when $\alpha = 1$ and assuming firm M has the highest productivity the $TFP_e/TFP = (\sum (z_i/z_M) (n_i/n))^{-1}$. This proves the following Proposition:

Proposition 3. *When firm level tfp and wedges are obtained from the data as in Hsieh and Klenow [2009a], the ratio TFP/TFP_e decreases with α and it is equal to one (i.e. no gap) at $\alpha = 0$.*

As an example, suppose the economy consists of two firms and $n_1/n = n_2/n = 1/2$. The following table gives the TFP ratios for different levels of curvature and degree of distortions, as measured by the relative average output of the two firms.

Table 5: TFP, Distortions and Curvature

	α	0	0.2	0.5	0.8	0.95
relative y_i/n_i						
0.2	1	1.09	1.28	1.57	1.74	
0.4	1	1.05	1.17	1.39	1.55	
0.6	1	1.02	1.08	1.22	1.35	
0.8	1	1.01	1.02	1.07	1.16	
1	1	1	1	1	1	

It can be seen that TFP is very sensitive to the degree of curvature and as stated in the Proposition increases with α .

8 Final remarks

Distortions that prevent the equalization of marginal products of inputs across firms can be costly. This paper developed a precise characterization of how distortions map into aggregate productivity. More precisely, aggregate TFP is the integral of a strictly concave function of distortions with respect to an input weighted measure of distortions. This establishes a precise ordering of distortions associated to mean preserving spreads. Thus, more concentrated distortions will have more detrimental effects on productivity. In contrast, higher correlation of distortions with firm level productivity is not necessarily more damaging to aggregate productivity, though it is likely to be so for highly distorted economies.

As an application, a lower bound of distortions that is consistent with observed size distribution of firms across countries was developed. This requires a rank-preserving (or monotone) transformation of firm sizes. Any other measure of distortions that generates the same size distribution is a mean preserving spread of this lower bound and thus leads to a larger TFP gap. Applying our bounds to the size distributions of China, India and Mexico (as compared to the United States) delivers very small TFP gaps, orders of magnitude smaller than those reported by Hsieh and Klenow [2009a]. Policies or distortions that lead to large rank reversals are needed for misallocation to have large impact. This observation can provide guidance in future studies in seeking for specific policies or constraints that lead to misallocation. In particular, policies that preserve rankings of firm size such as tax exemptions for firms under some size threshold, subsidies to small firms or restrictions on the size of manufacturing plants or retail stores will not have large effects on TFP, as consistently found in some reported studies.

An important caveat to our analysis, is that it concerns a particular structure, on homogeneous production/revenue functions with the same returns to scale across economies. While restrictive, this has been the class of models used in the growing and influential literature on macro development. A deeper understanding of the determinants of returns to scale and its measurement is clearly an important direction for future research in this area. Relatedly, extending results in this paper to economies with varying returns to scale or non-homogeneous production/revenue functions is also an important direction for future research.

References

- L. Alfaro, A. Charlton, and F. Kanczuk. Firm-size distribution and cross-country income differences. *manuscript, Harvard Business School*, 2007.
- A.V. Banerjee and E. Duflo. Growth theory through the lens of development economics. *Handbook of economic growth*, 1:473–552, 2005.
- E.J. Bartelsman, J.C. Haltiwanger, and S. Scarpetta. Cross-country differences in productivity: the role of allocation and selection. Technical report, National Bureau of Economic Research, 2009.
- Eric Bartelsman, John Haltiwanger, and Stefano Scarpetta. Cross-country differences in productivity: The role of allocation and selection. *American Economic Review*, 103(1):305–34, 2013.
- Francisco J. Buera, Joseph P. Kaboski, and Yongseok Shin. Finance and development: A tale of two sectors. *American Economic Review*, 101(5):1964–2002, 2011. doi: 10.1257/aer.101.5.1964. URL <http://www.aeaweb.org/articles.php?doi=10.1257/aer.101.5.1964>.
- Francesco Caselli and Nicola Gennaioli. Dynastic management. *Economic inquiry*, 51(1):971–996, 2013.
- Joel M David, Hugo A Hopenhayn, Venky Venkateswaran, and NYU Stern. Information, misallocation and aggregate productivity. *NBER Working paper*, 2014.
- Pablo N D’Erasmus and Hernan J Moscoso Boedo. Financial structure, informality and development. *Journal of Monetary Economics*, 59(3):286–302, 2012.
- A.K. Dixit and J.E. Stiglitz. Monopolistic competition and optimum product diversity. *The American Economic Review*, pages 297–308, 1977.
- Paolo Epifani and Gino Gancia. Trade, markup heterogeneity and misallocations. *Journal of International Economics*, 83(1):1–13, 2011.
- Larry G Epstein and Stephen M Tanny. Increasing generalized correlation: a definition and some economic consequences. *Canadian Journal of Economics*, pages 16–34, 1980.
- Luis Garicano, Claire Lelarge, and John Van Reenen. Firm size distortions and the productivity distribution: Evidence from france. Technical report, National Bureau of Economic Research, 2013.

- Francois Gourio and N. Roys. Size-dependent regulations, firm size distribution, and reallocation. In *2013 Meeting Papers*, number 199. Society for Economic Dynamics, 2013.
- N. Guner, G. Ventura, and Y. Xu. Macroeconomic implications of size-dependent policies. *Review of Economic Dynamics*, 11(4):721–744, 2008.
- H. Hopenhayn and R. Rogerson. Job turnover and policy evaluation: A general equilibrium analysis. *The Journal of Political Economy*, 101(5):915–938, 1993.
- H.A. Hopenhayn. Entry, exit, and firm dynamics in long run equilibrium. *Econometrica*, 60(5):1127–1150, 1992.
- C.T. Hsieh and P.J. Klenow. Misallocation and Manufacturing TFP in China and India. *The Quarterly Journal of Economics*, 124(4):1403–1448, 2009a.
- C.T. Hsieh and P.J. Klenow. Misallocation and manufacturing tfp in china and india. *The Quarterly Journal of Economics*, 124(4):1403–1448, 2009b.
- H. Jeong and R.M. Townsend. Sources of TFP growth: occupational choice and financial deepening. *Economic Theory*, 32(1):179–221, 2007.
- B. Jovanovic. Selection and the Evolution of Industry. *Econometrica: Journal of the Econometric Society*, pages 649–670, 1982.
- Boyan Jovanovic. Misallocation and growth. *The American Economic Review*, 104(4):1149–1171, 2014.
- R.E. Lucas Jr. On the size distribution of business firms. *The Bell Journal of Economics*, 9(2):508–523, 1978.
- M.J. Melitz. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6):1695–1725, 2003.
- Margaret Meyer and Bruno Strulovici. The supermodular stochastic ordering. Technical report, Discussion Paper, Center for Mathematical Studies in Economics and Management Science, 2013.
- Virgiliu Midrigan and Daniel Xu. Finance and misallocation: Evidence from plant-level data. *Working paper (forthcoming in the American Economic Review)*, 2013.
- Benjamin Moll. Productivity losses from financial frictions: Can self-financing undo capital misallocation? *Working paper*, 2012.

- Michael Peters. Heterogeneous mark-ups, growth and endogenous misallocation. *Manuscript, Londong School of Economics*, 2013.
- Erwan Quintin. Limited enforcement and the organization of production. *Journal of Macroeconomics*, 30(3):1222–1245, 2008.
- D. Restuccia and R. Rogerson. Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics*, 11(4): 707–720, 2008.
- André H Tchen. Inequalities for distributions with given marginals. *The Annals of Probability*, pages 814–827, 1980.