ASOCIACION ARGENTINA
DE ECONOMIA POLITICA

ANALES | ASOCIACION ARGENTINA DE ECONOMIA POLITICA

LIII Reunión Anual

Noviembre de 2018

ISSN 1852-0022
ISBN 978-987-28590-6-0

Extent and Determinants of Resource Misallocation: A Cross-section Study for Developing Countries

Román Fossati
Matheus Stivali

Extent and Determinants of Resource Misallocation:
A Cross-section Study for Developing Countries

Román Fossati*
Universidad Diego Portales, Chile

Matheus Stivali†
Instituto de Pesquisa Econômica Aplicada, Brazil

22nd May 2018

Abstract

Several factors distorting the allocation of resources across heterogeneous production units can be responsible for differences in total factor productivity (TFP) and output per capita across countries. Identifying and measuring the effect of specific distortions at country level have relevant policy implications. This paper investigates the sources of resource misallocation in developing countries. The main contribution of it is to study the role of particular idiosyncratic distortions not considered in the current literature using comparable methodology and data for a broad set of developing countries. The results indicate that, in contrast with previous evidence documented in the literature, distortions affecting international trade and related to business licensing and operation permits as well as to informality are key factors damaging TFP in the set of developing countries we considered.

JEL Codes: E02, D24, O11, O14, O43
Key words: Total Factor Productivity, Resource-misallocation, Economic Environment.

1 Introduction

Understanding the persistence of disparate levels of per capita output across different countries is a major goal of economics. This work investigates those differences from the perspective of resource misallocation literature, which highlights that idiosyncratic firm-level distortions reduce aggregate productivity (TFP). The work considers countries from two regions, Latin America and Africa, with a twofold objective: to measure the extent of misallocation in each country, and to study the association between misallocation indicators and institutions (business environment).

The workhorse tool to study economic growth and its differences across countries is the neoclassical growth model. Usually, it considers differences in factors of production (capital and labour) with all residual variation accounted as total factor productivity

*Corresponding author, e-mail: <roman.fossati@udp.cl>. Diego Portales University, Avda. Santa Clara 797, Huechuraba, Santiago, 8580000, Chile, (+56) 2213 0122

†e-mail: <matheus.stivali@ipea.gov.br>.

(TFP), a measure of how well an economy use its resources. More complex versions of the model take into account, for example, differences in the quality of labour force (i.e. human capital) and by this way improve model's empirical performance (Mankiw et al., 1992). Moreover, one can include measures of institutional quality in the model as a form to better understand differences in output and the interactions between factor accumulation and productivity. The reasoning of this approach is that good (bad) institutions shape the economic environment and affect decisions on micro level, which builds into aggregate (in)efficiency (Hall and Jones, 1999).

The importance of differences in resources endowment was recently renewed by the consideration that countries with different endowments of low/high skilled labour could choose a technology more appropriate to their circumstances. In this context, skill endowments and technology choices would then explain differences in income *per capita*, (Caselli and Coleman, 2006).

Even though these improvements increased model's explicative power, much of the different income levels is still resting on TFP differences. Therefore, understanding why TFP differs is essential to a better understanding of economic development. One could highlight two lines of research that have been investigating alternative explanations for TFP differences.

One line of research emphasises the role of barriers to technology adoption as the source of TFP disparity. In this framework, a firm needs to invest in technology adoption in order to increase its productivity, but there are barriers (institutional factors related to the economic environment) that make this investment more expensive in developing countries (Parente and Prescott, 1994, 2002). Since the low productivity in this framework is related to the technology adopted by the firms, it can be dubbed *a model of within-firm inefficiency* (Hsieh and Klenow, 2009, p. 1404).

The other line of research considers the heterogeneity of productive units while determining aggregate output. It emphasises that how productive resources are distributed among productive units matter. Considering that firms are heterogeneous, if a country had its resources allocated favouring less efficient firms while another had an unskewed allocation they would have different aggregate TFPs, even if their internal distributions of firm's TFP were the same. Although the heterogeneity of productive units and resource misallocation explanation were raised before (Banerjee and Duflo, 2005; Hopenhayn and Rogerson, 1993, e.g.), two later works shaped the field theoretically (Restuccia and Rogerson, 2008) and empirically (Hsieh and Klenow, 2009).

The factors driving misallocation among firms are usually connected with country's economic environment (regulation, tax system, credit market imperfections, and corruption), thus factors external to the firm. Many of these factors could be labelled as institutional factors, sharing some of the rationale aforementioned used by growth theory when considering institutional quality. Following the name used for the barriers to technology framework, this one could be called *a model of across-firms inefficiency*.

Using this model of across-firms inefficiency, the present work expects to guide policy to increase output and, therefore, welfare in Latin American and African countries. As the study of human capital was important to highlight the role of education for economic development, the study of resource misallocation can draw attention to the importance of reforming some policies and institutions. In developing countries, there are many policies intended to foster business performance, as subsidized credit for small firms, that may be actually harming aggregate productivity. Besides the policies, there are many problems – as corruption, justice performance, and poor infrastructure, for instance – that probably cannot be faced all at the same time. So, sorting these sources of resource misallocation according to their importance may also be a valuable guide to policy.

This work contributes to the field by measuring the extent of misallocation for a broad set of countries using a standardized World Bank's database. Thereafter, data about business environment is used to explore possible sources of resource misallocation.

The paper has more five sections. *Section 2* proceeds a literature review on the field. *Section 3* lays the theoretical groundwork for the empirical analysis, summarizing the ideas developed by Hsieh and Klenow (2009) and the misallocation indicators that arise from that framework. *Section 4* details the data sources used and data preparation, highlighting its limitations. *Section 5* shows the results in two steps, firstly, indicators of misallocation are calculated from firm-level data for each country in order to assess quantitatively the seriousness of resources misallocation within them. Secondly, these indicators are combined with indicators of business environment and then the correlation between the misallocation outcome and the covariates is studied through regression analysis. Finally, *Section 6* compares the results obtained with other results in the literature and offers some conclusions.

2 Literature Review

This section reviews some works from the resource misallocation literature in order to locate the theoretical framework in a broader context. The revision is not exhaustive[1], it focus only on works closely related to the analysis performed here.

For that purpose, it is divided in two subsections. Firstly, general aspects of the literature are presented through the revision of the main works on misallocation and productivity. The last subsection reviews four applied studies on misallocation, summarizing their methodological choices and data sources.

2.1 Approaches to Study Misallocation

As stated by Restuccia and Rogerson (2013), it is possible to identify two approaches to assess misallocation in the literature. The *direct approach* relies on identifying a specific factor that could be a source of misallocation and obtaining an empirical measure for it. Then a model with heterogeneous agents is used to build the aggregate impact of that factor on TFP, an example of this approach is Hopenhayn and Rogerson (1993). In contrast, the *indirect approach* does not try to trace the impact of a specific factor. Instead, it relies on the fact that any factor causing misallocation creates a wedge between market prices and establishment first order conditions. Therefore, this approach focus on identifying that wedges rather than a specific source. Examples of this approach are Restuccia and Rogerson (2008) and Hsieh and Klenow (2009).

Although the direct approach tracks a factor causing misallocation from the start, it has some drawbacks which favours the indirect approach when evaluating a set of possible misallocation determinants. Firstly, the direct approach demands a good measure of the factor assumed to cause misallocation, which often is not available. Secondly,

[1] For comprehensive surveys the reader is referred to Restuccia and Rogerson (2013) and Hopenhayn (2014).

the availability of such good measure generally implies ruling out of the analysis other factors that are also possibly important. In which case, the model cannot say anything about them. Since the goal of this paper is to explore a range of possible determinants, the indirect approach is adopted and the main works using this approach are highlighted on what follows.

Restuccia and Rogerson (2008) can be pointed out as a landmark on the indirect approach, it developed a theoretical model in which two comprehensive sources of misallocation can be identified. This model imposes a microstructure on the neoclassical growth model in order to allow heterogeneous productive units. Each firm has a technology Cobb-Douglas with decreasing returns to scale and a specific TFP, whereas factor shares are common across all establishments. Government intervenes in the economy with policies that create distortions - tax or subsidy - on the output, capital, and labour. These distortions are idiosyncratic to each firm. As a consequence, each firm's conditional factor demand is distorted, a tax (subsidy) will decrease (increase) factor's demand generating an allocation of resources among firms different from that purely driven by market forces. Thence, changing the equilibrium distribution.

The authors considered two possible cases for government's intervention. Firstly, the policy is not correlated with firm's productivity, e.g. the case in which the TFP distribution is homogeneous over a country and the government taxes one region to subsidise the other. It would cause a loss in terms of aggregate productivity. Secondly, the government decides to subsidise firms with low TFP, for instance due to pressure from those establishments, what is shown to cause a greater loss to aggregate productivity. This happens because economy's stock of capital and labour will have a greater share allocated to firms with low TFP (that become larger) at expense of the firms with high TFP (that become smaller), resulting in a lower aggregate TFP.

A second landmark is the work of Hsieh and Klenow (2009) whose model is described on next section. The importance of this work is due to its endeavour to take the aforesaid theoretical concepts to data and evaluate the problem empirically. While the economy's framework has many similarities with that used by Restuccia and Rogerson some modifications were introduced to match theoretical quantities to values observed on firm-level data.

Comparing United States, China, and India they found that substantial part of the difference on TFP among them may be attributed to distortions producing misallocation on the last two. To obtain that result the authors proceeded with a counter-factual exercise in which the resources in China and India were reallocated such that the marginal products were equalized on similar levels to those in the US.

One could point some caveats in the analysis method developed by Hsieh and Klenow. Firstly, it needs to make many parametric assumptions in order to recover information from data, what makes it subordinate to the truthfulness of that assumptions. Moreover, in its original version, the model does not consider misallocation among industries what could be an important situation for some development experiences.

2.2 Cross-country Studies

This subsection reviews four works that study the impact of misallocation in cross-sections of countries. The goal is to map the methodological choices made and the possible gaps that can be addressed by the present work.

The work Kalemli-Ozcan and Sorensen (2012) analysed of the extent of misallocation and its determinants for 10 African countries surveyed by World Bank's Productivity and Investment Climate Survey (PICS)[2]. As a benchmark the authors considered data from some other countries also included on PICS.

Two measures of misallocation are used: Marginal Revenue Product of Capital, and the relative capital distortion; both derived from the original framework of Hsieh and Klenow (2009). However, further simplifying assumptions were made to obtain these indicators. First they consider perfect competition on the most disaggregate level of the economy (or, in terms of the next section, define the parameter $\sigma \equiv \infty$ in the industry sector), moreover they abandon the procedure introduced by Hsieh and Klenow (2009) of estimating the capital shares for each industry from a relatively undistorted economy, i.e. adopting US economy as a benchmark, and simply fix the parameter $\alpha = 1/3$ for all industries. Even though these assumptions greatly facilitate the empirical work, they also imply giving up on much of the compelling microstructure used originally and produce some extreme cases[3]. From the total observations, authors select those with enough information to obtain the indicators and exclude those reporting strange or clearly wrong values for some key variables (e.g. firm age) what reduces their sample.

They found significant levels of capital misallocation for the selected African countries, whereas this was not a severe problem in the benchmark developed countries. Analysing misallocation determinants on country level, they calculate the correlation between the country-average misallocation indicator and country measures of institutional quality. The latter are from International Country Risk Guide (ICRG) and World Bank's Doing Business. They found positive correlation between the time demanded to register property and misallocation, as well negative correlation between expropriation risk and misallocation. On the firm level, they study misallocation determinants estimating an OLS regression of the misallocation indicator against variables related to economic environment from the PICS data. They found that "Limited Access to Finance" as an important determinant of misallocation as well "Weak Infrastructure" and "Red Tape".

Busso et al. (2013) studied the impact of resources misallocation on the productivity of Latin American countries using administrative data and survey data from World Bank's Enterprise Survey (WBES). Whereas the use of firm census data from many countries makes this work have an unparalleled coverage of firms on the region, it concurrently presents the challenge of comparability due to the different criteria used across the countries. The solution found by the authors was to present the results according to subsets in which the coverage of the data is similar and also recurring to WBES. The time span considered depend on each country, but overall data vary from mid 1990s to mid 2000s. Hsieh and Klenow (2009) framework is used without great modifications and the parameters are set as $\sigma = 3$ and α_s corresponding to the capital share in that industry in US. They considered four measures of misallocation, which are similar to those presented by Hsieh and Klenow, and use the variance of the log of those as a summary measure.

To study misallocation determinants these authors summarize information related to policy constraints from WBES by factor analysis. Then estimate an OLS regression

[2]The data are from 2005 and 2006. PICS is a previous version of the Enterprise Survey discussed on section 4.

[3]As pointed by Hsieh and Klenow (2009, p. 1414) the gains of misallocation are increasing in σ and the range for that parameter originally considered was from 3 to 5.

of the logarithm of their misallocation indicator against the policy variables. They found "the difficulty to access capital" and "the stringent labour regulations" in Latin America as major explanations for misallocation.

Inklaar et al. (2015) investigate how much of the differences in TFP among 54 countries can be attributed to resource misallocation, reproducing in large scale the original exercise by Hsieh and Klenow, but using survey data from WBES. They found that the possible TFP gains by reallocation would be of 60% on average, however they found that the severity of misallocation is uncorrelated with observed productivity level (i.e. the countries that could benefit more from a "reform" are not those with lower productivity).

Finally, the work from García-Santana and Ramos (2015) studies misallocation considering the relation between TFP and firm size distribution, following Lucas (1978). These authors claim that even though the relation between the amount of labour allocated to small plants and TFP is theoretically ambiguous, most of the models considered on resource misallocation literature predict a negative relationship between these quantities. They corroborate this point by producing measures of the average plant size by country's TFP category (high/low TFP) and country's GDP-per-worker category, using data from WBES and estimates from Caselli (2005).

Then the authors try to explain the share of labour allocated to small firms (less than 20 employees), which is their indicator of misallocation, estimating a series of OLS regressions of this share against controls and indicators of economic distortions (indexes from World Bank Doing Business). They found the general-compound index from Doing Business as an relevant explanatory variable. Considering individual indexes, "financial constraints" and "entry costs" were identified as relevant explanatory variables. Even though firm size distribution is intertwined with resource misallocation, assuming the share of small plants as an misallocation indicator is not appropriate since there are many other factors influencing firm size distribution. Furthermore, since the indicator chosen is a limited variable the studied relations would be better estimated by a logit-type regression, for instance.

The works discussed above share the concern that misallocation is a serious factor reducing TFP, moreover they also consider institutional factors as an important determinants of the misallocation extent. While Kalemli-Ozcan and Sorensen (2012) uses similar databases to those used here, it considers only an extreme case on the framework proposed by Hsieh and Klenow (2009) and focus on a small set of African economies. On the other hand, the work of Busso et al. (2013) uses high quality data and the full-fetched version of Hsieh and Klenow (2009) focusing on Latin American countries, but it only considers the information provided by the firms when considering possible determinants on misallocation. As explained on next sections, there is possible gains in combining information external to the firms, as that available in Doing Business database, and approaching the problem on firm and on country level. The work of Inklaar et al. (2015) uses data from WBES and finds that misallocation gains are not correlated with observed productivity level and that low-income countries wouldn't be the biggest winners from a reform. This certainly challenges the view that poor economic environment is related to misallocation. At last, García-Santana and Ramos (2015) studies misallocation determinants using WBES data but chose a misallocation indicator that is very different from the others.

Another possible problem with these works is that they might have used some of the formulas from Hsieh and Klenow (2009) (noticeably for the $\overline{TFPR_s}$) which were corrected later on by the authors (Hsieh and Klenow, 2013). This certainly affect their results and the comparability with the calculations provided on this paper that use the last version of the formulas.

3 Theoretical Framework and Misallocation Indicators

This section summarizes the theoretical model developed by Hsieh and Klenow (2009) and presents the misallocation measures that arise from it. The economy considered has three levels: firm level (Y_{si}), industry level (Y_s), and aggregate level (Y).

$$\overbrace{\overbrace{Y_{1,1}\ldots Y_{1,M_1}}^{Y_1} \quad \overbrace{Y_{2,1}\ldots Y_{2,M_2}}^{Y_2} \quad \ldots \quad \overbrace{Y_{S,1}\ldots Y_{S,M_S}}^{Y_S}}^{Y}$$

On the aggregate level, the final good is produced by a representative firm in a perfectly competitive market by the combination of the S intermediary goods using a technology Cobb-Douglas with constant returns to scale: $Y = \prod_{s=1}^{S} Y_s^{\theta_s}, \sum_{s=1}^{S} \theta_s = 1$.

On the industry level, there are S different industries and each produces an intermediary good by the combination of differentiated products using a Constant Elasticity of Substitution (CES) in a perfectly competitive market: $Y_s = \left(\sum_{i=1}^{M_s} Y_{si}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$, where σ is the elasticity of substitution. The number of differentiated products (or the number of establishments, M_s) can vary from industry to industry. However, the CES elasticity (σ) is assumed the same for all industries as a simplifying device.

On the firm level, the economy is populated establishments[4] that are heterogeneous with respect to their productive efficiency (A_{si}, the TFP, is specific of each firm) and produce differentiated products, i.e. this is a monopolistic competitive market. These establishments combine capital (K_{si}) and labour (L_{si}) using a Cobb-Douglas technology: $Y_{si} = A_{si} K_{si}^{\alpha_s} L_{si}^{1-\alpha_s}$, where α_s is the industry capital share. The factor shares can vary across industries but *not within* industries.

There are distortions that are specific to each establishment and create wedges between market prices and the prices observed by that establishment. Considering two production factors it is possible identify two sorts of distortions:

- **Output distortion (τ_{Ysi})**: affects capital *and* labour at the same time. This distortion increases or reduces the marginal product of capital and labour in the same proportion. It could be thought as a tax/subsidy on the final output, transportation costs, anything that affects this establishment output's price idiosyncratically.
- **Factor distortion (τ_{Ksi})**: affects the marginal products of capital *relative* to the marginal products of labour. It could be thought as a subsidy on interest rates or a

[4] It is specially important in this model distinguish between firms and establishment. While a firm may have many plants, the establishment is the individual productive unit. It is the latter that is considered on the most disaggregate level of the economy even though it is named "firm level".

payroll taxation on labour. Again the distortion is an establishment's idiosyncratic characteristic.

With these distortions the profit function takes form[5]:

$$\pi_{si} = (1 - \tau_{Ysi})P_{si}Y_{si} - wL_{si} - (1 + \tau_{Ksi})RK_{si} \tag{1}$$

Then one can see, for example, how a subsidy to an establishment that decreases the interest rate paid (i.e. a negative τ_{Ksi}) affects its objective function. It would reduce establishment's Marginal Revenue Product of Capital ($MRPK_{si}$) and consequently increase its demand for capital. So, through factor demands, distortions change the resource allocation among establishments from what it would be in an undistorted economy.

Following Foster et al. (2008, p. 400), one can distinguish two sorts of productivity that can be retrieved from data and are theoretically consistent: revenue productivity (TFPR = $P_{si}A_{si} = \frac{P_{si}Y_{si}}{K_{si}^{\alpha_s}L_{si}^{1-\alpha_s}}$) and physical productivity (TFPQ = $A_{si} = \frac{Y_{si}}{K_{si}^{\alpha_s}L_{si}^{1-\alpha_s}}$). The latter corresponds to A_{si} above and its identification depends on either having data on plant-specific prices (i.e. the quantities and prices sold by each establishment must be known), or assuming a structural model that allow the identification from other (known) quantities.

It can be shown that $TFPQ_{si}$ can be written as $TFPQ_{si} = A_{si} = \kappa_s \frac{(P_{si}Y_{si})^{\frac{\sigma}{\sigma-1}}}{K_{si}^{\alpha_s}L_{si}^{1-\alpha_s}}$. This expression is based only on values observed on data ($P_{si}Y_{si}$, K_{si} and L_{si}) and κ_s. The quantity κ_s is not observed, but since one is not studying reallocation among sectors – indeed only within sectors – this value does not affect firm's relative productive within sector (all establishments have the same κ) neither reallocation gains. Therefore there is no loss in fixing $\kappa_s = 1$ ∀ s ∈ S, what allows the estimation of A_{si} from firm-level data.

Concurrently, it can be shown that revenue productivity can be written as: $TFPR_{si} = \left(\frac{\sigma}{\sigma-1}\right)\left(\frac{(1+\tau_{Ksi})R}{\alpha_s(1-\tau_{Ysi})}\right)^{\alpha_s}\left(\frac{w}{(1-\alpha_s)(1-\tau_{Ysi})}\right)^{1-\alpha_s}$

As noted by Hsieh and Klenow (2009, p. 1410), $TFPR_{si}$ only varies among establishments of the same industry if they experience distortions (capital and/or output). Therefore, analysing the dispersion of this quantity within each industry would result in a measure of how severe are the distortions.

With the $TFPR_{si}$ formula at hand it is possible to construct a measure of industry productivity ($TFPR_s$). With that objective and following Hsieh and Klenow strategy, it is necessary to define the following quantities[6]:

[5]The interpretation of the distortions depend on how the τ values enter in the profit function (adding or subtracting). Here the form adopted by Hsieh and Klenow (2009, p. 1407) and followed by others was kept.

[6]Notice that the following three formulas follow Hsieh and Klenow (2013).

$$\overline{MRPK}_s \equiv \frac{R}{\sum_{i=1}^{M_s} \frac{(1-\tau_{Ysi})}{(1+\tau_{Ksi})} \frac{P_{si}Y_{si}}{P_s Y_s}} \tag{2}$$

$$\overline{MRPL}_s \equiv \frac{w}{\sum_{i=1}^{M_s} (1-\tau_{Ysi}) \frac{P_{si}Y_{si}}{P_s Y_s}} \tag{3}$$

$$\overline{TFPR}_s \equiv \frac{\sigma}{\sigma-1} \left(\frac{\overline{MRPK}_s}{\alpha_s} \right)^{\alpha_s} \left(\frac{\overline{MRPL}_s}{1-\alpha_s} \right)^{1-\alpha_s} \tag{4}$$

One can focus on what measures derived from the model can be obtained from data as well how misallocation can be measured. As noted before the existence of distortions (τ_{Ksi} and τ_{Ysi}) create wedges between the first order conditions and market prices, it is possible to write these FOC as:

$$(1-\tau_{Ysi}) = \left(\frac{\sigma}{\sigma-1} \right) \frac{wL_{si}}{(1-\alpha_s) \quad P_{si}Y_{si}} \tag{5}$$

$$(1+\tau_{Ksi}) = \frac{\alpha_s}{1-\alpha_s} \quad \frac{wL_{si}}{RK_{si}} \tag{6}$$

The gain from reallocation, as defined by Hsieh and Klenow (2009, p. 1414), would arise if the establishments were treated evenly instead of each one facing its idiosyncratic distortion. What would mean each establishment being subject to an average distortion within each industry ($\overline{\tau}_{Ks}$ and $\overline{\tau}_{Ys}$)[7]. Therefore a possible measure of misallocation extension would be the dispersion of actual τs relative to its hypothetical average.

A similar rationale can be used for the productivity measures $TFPR_{si}$ and $TFPQ_{si}$. However, for these quantities the reference values are not simply averages. While the $\overline{TFPR}_s$ is given by equation (4), $\overline{A}_s$ is a CES aggregation of each A_{si} in industry s: $\overline{A}_s \equiv \left(\sum_{i=1}^{M_s} A_{si}^{\sigma-1} \right)^{\frac{1}{\sigma-1}}$.

With the quantities above defined it is possible to define objective measures of misallocation. As said above, we consider the measures relative to its respective average and proceed a log transformation as usual in the literature, defining the indicators in the formulas (7) to (10)[8]. The dispersion of these indicators will be the misallocation indicators.

[7]Notice that the comparison is not with a case without distortions but indeed with a case in which the distortion is the same for all productive units.

[8]For the graphs presented below formula (9) is slightly modified. The numerator is scaled by the number of establishments in the industry (M_s) to the power of $1/(\sigma-1)$, resulting in: $(M_s^{\frac{1}{\sigma-1}} A_{si})/\overline{A}_s$.

$$log\left(\frac{1+\tau_{Ksi}}{1+\overline{\tau}_{Ks}}\right) \tag{7}$$

$$log\left(\frac{1+\tau_{Ysi}}{1+\overline{\tau}_{Ys}}\right) \tag{8}$$

$$log\left(\frac{A_{si}}{\overline{A}_{s}}\right) \tag{9}$$

$$log\left(\frac{TFPR_{si}}{\overline{TFPR_{s}}}\right) \tag{10}$$

4 Data

4.1 World Bank's Enterprise Survey[9]

The main database used in this work comes from a World Bank initiative called Enterprise Survey (WBES). With the goal of studying business environments and firms performance (productivity and job creation) in a internationally comparable way, this survey collects data on firm-level from certain manufacturing and service sectors in all world's regions using standardized instruments and sampling method.

Furthermore, WBES was designed to be a longitudinal database such that establishments from previous survey waves are reinterviewed and can be tracked. Each specific country-year wave generates their own database since not all countries are surveyed every year.

The ultimate goal of this paper is to make comparisons among countries so it uses a consolidated dataset version, built by World Bank itself, that put together and standardize those individual country-year databases. The standardization process consists in matching country data to a standard set of questions and exclude those country-specific survey questions which cannot be matched. The downside of using this consolidated database is that it is not so well documented as the individual ones and sometimes it is necessary to refer back to country-specific surveys.

The standardized dataset is continuously updated with the inclusion of amendments and new countries. The version used used in present work is referred to 26 July 2015 and has information from 2006 to 2014.

4.1.1 Sampling Characteristics and Weighting[10]

The WBES is a survey that follows a stratified random sampling strategy designed to be representative of sampled sectors. The population sampled in each country is

[9]To access this data is necessary to register with the World Bank's Enterprise Analysis Unit (DECEA) and complete the Enterprise Surveys Data Access Protocol. As demanded in the protocol, the data here is used in accordance with World Bank rules governing "strictly confidential" information.

[10]This section is based on World Bank (2009b) and World Bank (2014).

comprised by the formal companies with five or more employees of the following two-digit ISIC[11] sectors: manufacturing (15 to 37); construction (45); services (50 to 52 and 55); transport , storage, and communications (60 to 64), and computer and related activities (72).

Stratification conveys better precision of estimates and allows analyses at each level, it is done following three criteria: sector of activity (1 to 9 strata, depending on economy's size in GNP terms), firm size (3 strata), and geographical location (depend on the distribution of non-rural activities over the country). Sample size depend on population size and is chosen to ensure a minimum of precision (7.5% for 90% confidence intervals) for estimates of: *population proportions* and *log of sales*, both at industry level.

Due to this complex survey design, it is necessary to weight the observations in order to make proper estimates and inference about the respective population. Whereas the point estimates does not change with the use of full sample design or only weights in the calculations, the standard deviations and confidence intervals do change. Moreover, if one ignored the sampling structure at all, even the point estimates obtained would be wrong.

The WBES uses probability weights and the datasets usually have three options of weighting (weak, median, and strict)[12], depending on each observation's eligibility to the sample. The eligibility concept exists because WBES is made in two stages, first the establishment is contacted by phone (screening stage) to check if it really matches with the information used for sampling (which industry, for instance) and later the same establishment is surveyed (interviewing stage). So, after the screening, the eligibility criteria tries to correct distortions between the theoretical sample (i.e., firms sampled from a master list) and the real establishments actually found.

For the published Enterprise Surveys indicators, World Bank uses the weight-median option with full sample's design specification[13]. The calculations in this paper made were done in the same way using the survey commands on Stata.

4.1.2 Industry identification - ISIC Levels

As could be seen in the theoretical framework, industry definition is fundamental for the research, once misallocation effects are studied within each industry. Ideally the industries would be defined such that they encompassed similar establishments while having a sufficiently large sample in each group.

WBES collects information on each establishment main product using a four-digit ISIC, an aggregation level similar to that used by Hsieh and Klenow (2009). However, probably due to heterogeneity among datasets and the standardization process, this information has problems and is reported for many observations with only two or three digits, besides missing values. Meanwhile, the database has another variable that reports a two-digit ISIC, whose construction is not explained in any part of the documentation.

To properly evaluate the options about industry definition, a two-digit ISIC variable

[11] International Standard Industrial Classification.

[12] Whose values can vary greatly, as in World Bank (2009a), or in a smaller scale, as in World Bank (2010).

[13] This guideline was not included in any of the consulted WBES documents and was only found at the FAQ section at `www.enterprisesurveys.org`.

was constructed from the four-digit one with a twofold objective: test the consistency of the original two-digit variable and increase the number of observations with ISIC information. It was done using the hypothesis that those observations reporting an ISIC-4 with two or three digits were missing the last digits so that the information could be used (e.g. an reported ISIC-4 equal 173 was supposed to be 173? and, therefore, ISIC-2 equal 17). Even for observations with complete information (ISIC-4 and ISIC-2), dataset's original two-digit ISIC couldn't be matched once it often differs from the ISIC-4.

Considering these features, this work defines industry according to the constructed two-digit ISIC variable since it is consistent with the best information on the database (ISIC-4) and maximizes the number of observations available.

4.1.3 US Capital Shares - The Fundamental Benchmark

Other fundamental aspect for misallocation study is to obtain the benchmark capital share in a relatively undistorted economy. As done by Hsieh and Klenow (2009, p. 1413-1414), the NBER Manufacturing Industry Productivity Database[14] is used for that.

Following Hsieh and Klenow, capital shares (α_s) are defined as 1 minus labour share in the industry, whereas labour share is calculated as Total Payroll divided by Total Value Added within each industry. And labour shares are corrected, because they are substantially smaller than what is observer in data from national accounts, what happens because non-wage benefits are not considered in the former. The authors proceeded a correction of multiplying labour shares by a factor of 1.5, what was also done in this work.

Due to the option for an ISIC-2 classification stated previously, it is necessary to adapt information from NBER since industry information is presented with an ISIC four-digit. So, to obtain and ISIC two-digit equivalent information, value added and payroll were summed within each two-digit category and the labour and capital shares were calculated.

4.2 Doing Business

Even though WBES has a set of variables on entrepreneurs or managers perceptions about their business environment, using these indicators as explanatory variables can be misleading since self-reported perceptions may be subject to endogeneity or other biases. For instance, an entrepreneur lacking managerial ability may attribute its poor performance to external factors, whereas a talented manager might not highlight environment factors as relevant barriers. Moreover, how to use this kind of firm-level perception information when analysing country-level outcomes is not clear.

Doing Business, another initiative from World Bank, measures regulations that enhance business activity and those that constrain it. This survey collects data yearly in many countries: interviewing local experts – such as lawyers, accountants, and business consultants – and consolidating this information with the laws and regulations in those countries. Furthermore, information on time and cost to complete common business transactions are collected based on standardized case-studies.

[14]Data can be found in `www.nber.org/data/nberces.html`. Bartelsman and Gray (1996) explains the constructions and characteristics of this database.

In contrast to the perception questions from WBES, one could highlight that information from Doing Business uses factual information about what laws and regulations say. On the other hand, it is not a statistical survey, it only summarizes relevant laws and regulations (hard data) and expert opinions (soft data)[15]. The goal of Doing Business is to portrait what a representative firm, obeying all regulations, should do, pay, and wait. With these characteristics, the World Bank itself recognize that WBES and Doing Business are complements to each other.

Doing Business covers eleven topics: starting a business, dealing with construction permits, employing workers, registering property, getting credit, getting electricity, protecting investors, paying taxes, trading across borders, enforcing contracts, and closing a business. Each of these topics are related to more than one variable in the database (for instance, related to starting a business, one has information on how long it takes (in days), how many procedures are necessary, etc.). Over the years and across the countries there are some differences in what variables are available. For instance, concerning the topic starting a business, one could have the information about how long it takes for all countries and years, but the information about how many procedures will be available only from a specific year. Therefore, the database used is a result of the availability of misallocation indicators and data from Doing Business.

4.3 The Final Sample

Even though the WBES original sample has many observations of firms from 127 countries surveyed between 2006 and 2014, this sample is greatly reduced for the analysis performed here.

First, we concentrate on two developing regions with distinct characteristics (Latin America and Africa) and, as formerly noticed, observations with bad ISIC coding are discarded. Second, although WBES surveys not only manufacturing firms but also firms from other sectors, the study of misallocation is typically focused on manufacturing. Therefore, those observations that were not matched with the correspondent US manufacturing sector from the NBER were also discarded.

Moreover, it must be recognized that the study of misallocation is highly data-demanding and much information is necessary to estimate the quantities on formulas (22) to (26) from the previous section. Following other studies (Kalemli-Ozcan and Sorensen (2012, p. 10-11), Busso et al. (2013), Inklaar et al. (2015), and García-Santana and Ramos (2015)) observations without information enough to calculate the misallocation indicators were discarded.

Finally, because all misallocation indicators are based on the dispersion of the quantities estimated from the formulas (7) to (10) it is natural that they are somewhat sensible to extreme values. Following what was originally proposed by Hsieh and Klenow (2009, p. 1416) and repeated by others, the sample is further reduced by dropping the observations in the top and the botton 1% of $log\left(\frac{A_{si}}{A_s}\right)$ and $log\left(\frac{TFPR_{si}}{TFPR_s}\right)$ considering all industries together within each country and year. After this procedure, the quantities relative to industry or economy totals (e.g. M_s, K_s) are recalculated for the final sample.

[15] Further details can be found in `www.doingbusiness.org/methodology/methodology-note`

5 Econometric Analysis

This section presents the results for the extent and determinants of misallocation for Latin American and African countries where data were available and allowed to estimate the indicators[16].

Besides the data preparation issues defined previously, it is also necessary to set the values of some parameters to proceed the analysis. These are set to the same values originally used by Hsieh and Klenow: R=0.10, σ=3. Moreover, instead of considering separately wages (w) and total labour hired (L) by each firm, the total payroll expenditure (wL) is considered due to a database limitation.

The first subsection presents the misallocation extent results, whereas the second presents the association of misallocation indicators with economic environment factors.

5.1 Misallocation Extent

The existence of idiosyncratic firm-level distortions makes the marginal revenue products ($MRPK_{si}$ and $MRPL_{si}$) diverge among firms. This fact implies that how much a quantity (distortion, and physical or revenue productivity) faced by a specific firm differs from the average (in that firm's sector) indicates the sternness of misallocation confronted by that firm. Indeed, a measure of the dispersion of these indicators acts as a summary measure of misallocation in the overall economy.

Starting with a graphical inspection of the dispersion and considering sub-samples with countries that have data in two points in time, the TFPQ ($log(M_s^{1/(\sigma-1)} A_{si}/\overline{A}_s)$) is plotted for Latin America (in figures 1 and 2) and for Africa (in figures 3 and 4).

[16]The definition of Latin America used here is quite loose and refers to all countries in the continent except US and Canada. The lack or inadequacy of data made that some relatively large economies were not included (e.g. Venezuela whose sample is plenty of missing information) or included only once (namely Brazil that had a survey previous to 2009, which is not comparable to the current version of WBES).

Figure 1: Kernel Distribution of $log\left(\frac{M_s^{\frac{1}{\sigma-1}}A_{si}}{\overline{A}_s}\right)$ for Latin America in 2006

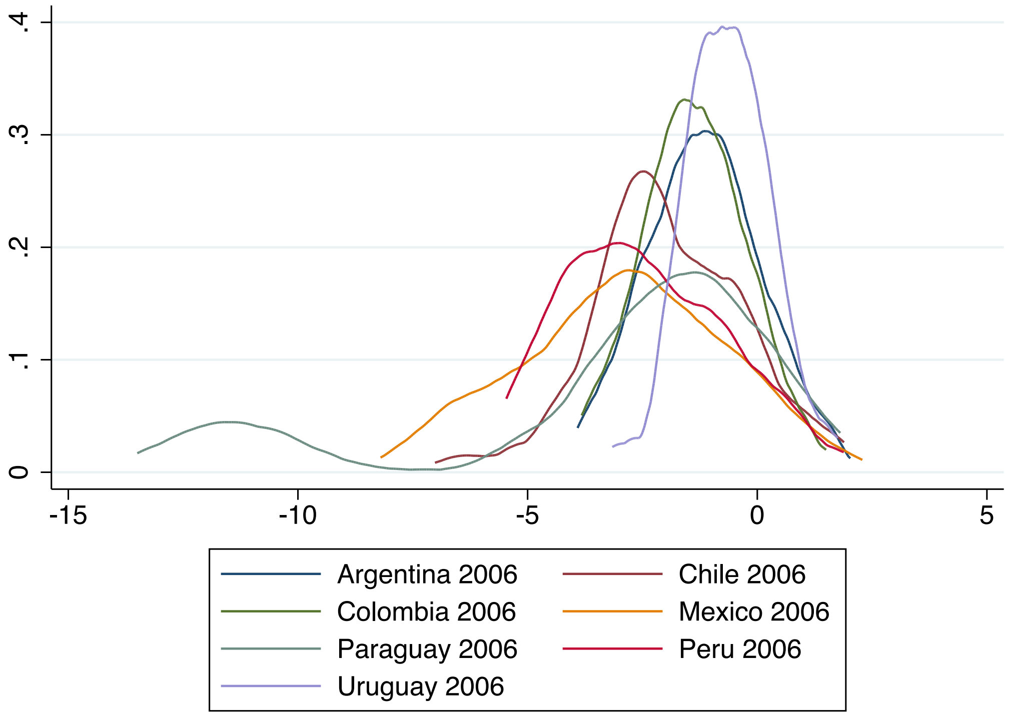


Figure 2: Kernel Distribution of $log\left(\frac{M_s^{\frac{1}{\sigma-1}}A_{si}}{\overline{A}_s}\right)$ for Latin America in 2010

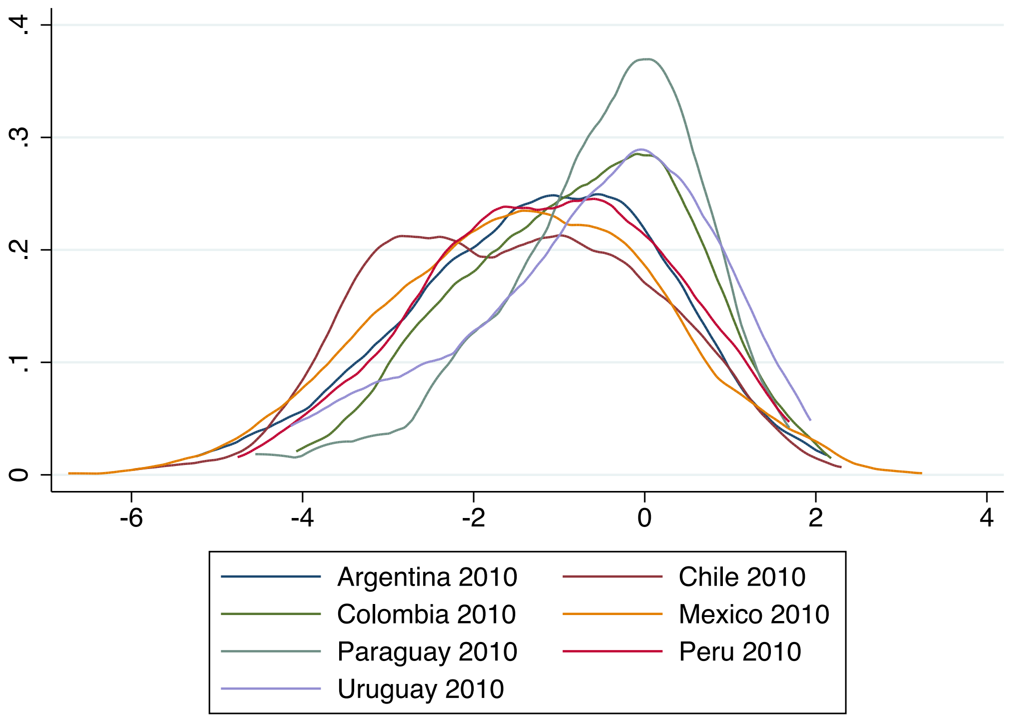


Figure 3: Kernel Distribution of $log\left(\frac{M_s^{\frac{1}{\sigma-1}}A_{si}}{\overline{A}_s}\right)$ for Africa in 2006/7

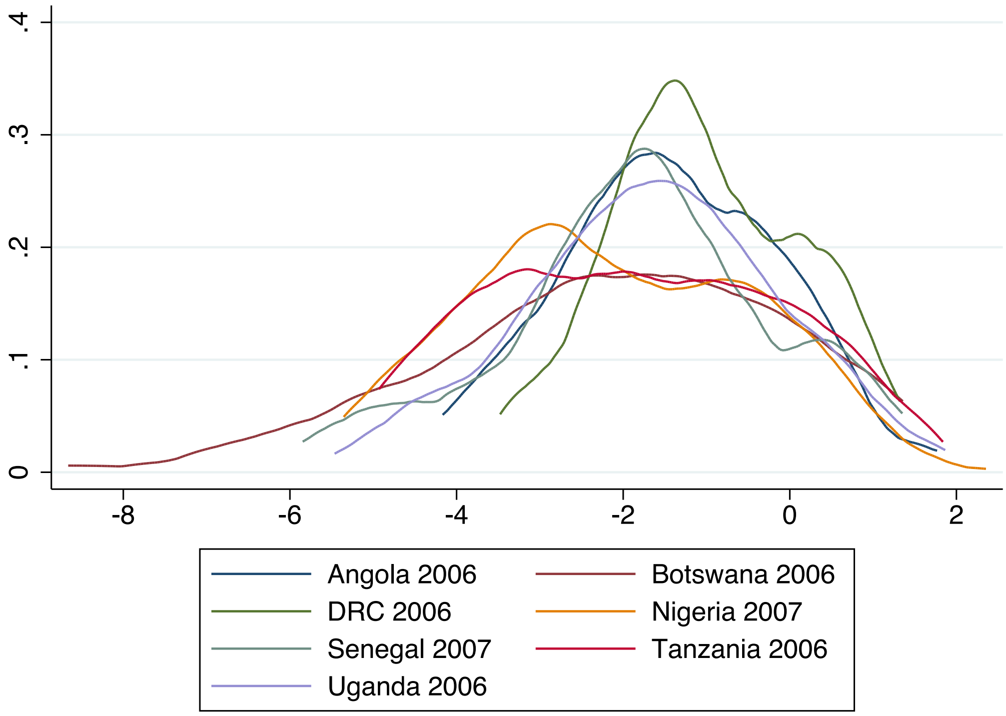
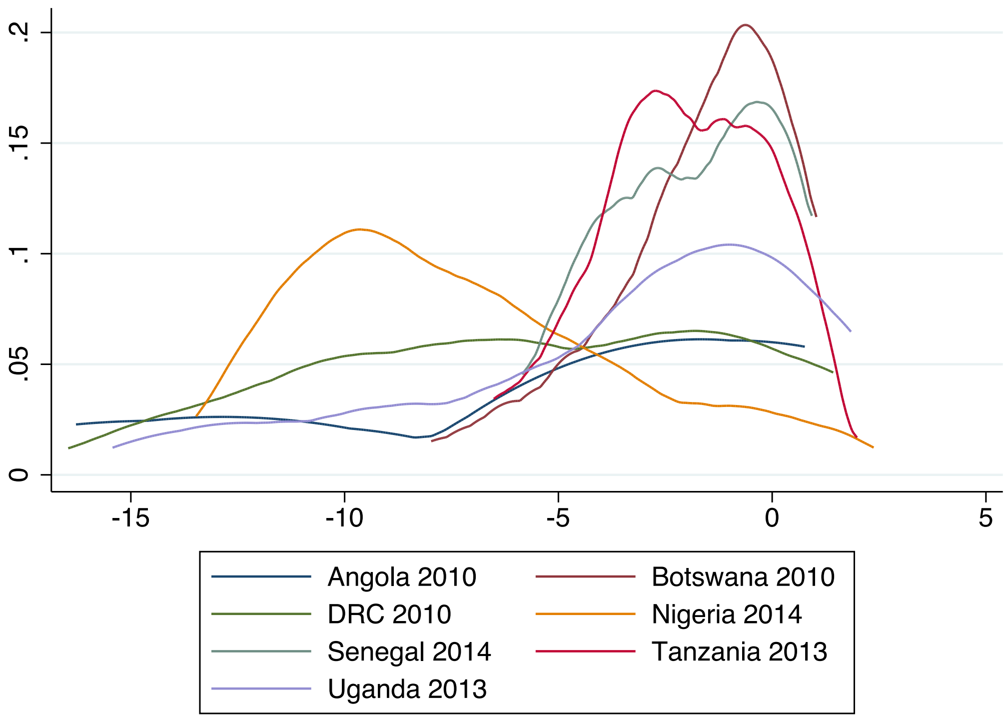


Figure 4: Kernel Distribution of $log\left(\frac{M_s^{\frac{1}{\sigma-1}}A_{si}}{\overline{A}_s}\right)$ for Africa in 2010/13/14



Looking the distribution of TFPQ for the Latin American countries in figure 1; it is discernible that in 2006 Argentina, Colombia, and Uruguay had the smaller dispersion, whereas Paraguay had a quite thick left tail indicating a high dispersion and, therefore, stern misallocation. This pattern changed in 2010 when the dispersion in all countries considered become more similar, except for Colombia where TFPQ is quite less disperse.

On the other hand, the opposite behaviour is observed for the African countries considered. While Botswana that had a flattened distribution in 2006 reduced notably the thickness of its distribution tails in 2010, other countries worsened a lot increasing the dispersion and, specially, Nigeria in 2014 whose distribution not only became more disperse but also moved leftwards (i.e. to lower levels of TFPQ).

The general results from graph inspection are confirmed by the dispersions measures of $log(A_{si}/\overline{A}_s)$. The estimated standard deviation (sd), interquartile range, and 90th-10th percentile range for the distribution on TFPQ are presented on table 1 for Latin American countries and in table 2 for African countries.

The three considered measures are generally smaller in Latin American countries than in African countries. Considering the time dimension it is noticeable that while the dispersion decreased or increased (relatively) little in Latin America, it increased for African countries.

A pattern of moderate increase for the dispersion of TFPQ in Latin America was found by Busso et al. (2013, p. 911). Even though the results reported in that study correspond to different databases (administrative data), the levels and the trend found by them are corroborated by the data presented on table 1.

Moving the analysis to the dispersion of revenue productivity, the $log(\frac{TFPR_{si}}{TFPR_s})$ distribution is plotted for Latin American countries (in figures 3 and 4) and for African countries (in figures 7 and 8).

Again, one can see the pattern of the Latin American countries becoming more similar; each country with a distribution of TFPR less dispersed over time. Notably, Paraguay that had quite long tails in 2006 converged to a fairly concentrated distribution in 2010. Meanwhile the African countries diverged. While in the starting years (2006/2007) Botswana and Tanzania had the most dispersed distributions, later Botswana (in 2010) and Tanzania (in 2013) got a distribution of TFPR very little dispersed and the other countries increased their distribution dispersion on different degrees.

This graphical depiction is mirrored by the dispersion measures on table 3 for Latin America and table 4 for Africa. The three considered dispersion measures are generally higher in the African economies than in the Latin American ones. Some economies in Africa have quite high dispersion measures, as Angola in 2010[17].

The higher heterogeneity of productivity suggest that misallocation in Africa is more severe than in Latin America. Moreover while firms in Latin American economies became more similar in terms of TFPR over time the opposite happened in Africa.

Considering the distortions themselves (τ_{Ysi} and τ_{Ksi}) the pattern of Africa with a higher dispersion than Latin America is repeated, however it is worth noting that for both regions the dispersion of τ_{Ksi} is higher than that of τ_{Ysi}. This means that in both regions the distortions in input markets are more strong than those in output

[17]Even though the sample considered by Kalemli-Ozcan and Sorensen (2012) is encompassed by the one used here that authors do not present statistics for individual countries, instead they consider all economies together as "African countries". Additionally, the misallocation indicators used by that authors are not comparable with the ones presented here.

Table 1: Dispersion Measures of $log\left(\frac{A_{si}}{A_s}\right)$ for Latin American countries

	year	sd	90th-10th	75th-25th
Antigua and Barbuda	2010	0.78	1.66	1.36
Argentina	2006	1.30	3.47	1.97
Argentina	2010	1.53	4.31	1.95
Bahamas	2010	1.86	5.04	2.68
Barbados	2010	1.08	2.86	1.76
Belize	2010	1.30	3.31	1.84
Bolivia	2006	1.55	3.87	2.04
Brazil	2009	1.56	4.62	2.61
Chile	2006	1.67	3.91	2.60
Chile	2010	1.58	3.84	2.60
Colombia	2010	1.47	4.33	1.34
Costa Rica	2010	1.83	5.98	2.02
Dominican Republic	2010	1.66	3.40	1.58
Ecuador	2006	1.63	4.00	2.80
Ecuador	2010	1.85	5.52	3.27
El Salvador	2006	3.29	9.46	3.71
El Salvador	2010	1.29	3.29	1.92
Grenada	2010	0.58	1.74	0.49
Guatemala	2006	2.29	5.63	3.92
Guatemala	2010	1.51	3.86	2.38
Guyana	2010	1.72	3.98	3.28
Honduras	2006	1.72	4.67	2.12
Honduras	2010	3.52	8.09	5.91
Jamaica	2010	0.95	2.60	1.42
Mexico	2006	2.30	6.23	3.06
Mexico	2010	1.84	4.63	3.34
Nicaragua	2006	3.71	10.58	4.65
Nicaragua	2010	1.11	2.97	1.42
Panama	2006	2.26	5.95	3.35
Panama	2010	2.03	4.95	2.20
Paraguay	2006	4.19	12.33	3.60
Paraguay	2010	1.49	3.86	2.02
Peru	2006	1.99	5.23	2.63
Peru	2010	1.43	3.80	1.91
Suriname	2010	1.29	3.73	1.63
Uruguay	2006	1.12	2.93	1.58
Uruguay	2010	1.49	4.36	1.95

Table 2: Dispersion Measures of $log\left(\frac{A_{si}}{A_s}\right)$ for African countries

	year	sd	90th-10th	75th-25th
Angola	2006	1.47	3.76	2.05
Angola	2010	7.25	17.07	9.13
Botswana	2006	2.26	5.97	3.02
Botswana	2010	2.09	4.86	2.72
Burkina Faso	2009	1.62	3.45	3.08
Burundi	2006	1.49	3.99	2.47
Burundi	2014	1.12	2.48	1.06
Cameroon	2009	3.35	9.84	3.62
Ivoire	2009	2.10	5.89	3.13
DRC	2006	1.24	3.12	2.02
DRC	2010	4.96	13.09	9.10
DRC	2013	2.72	7.89	3.59
Egypt	2013	1.72	4.87	2.17
Ethiopia	2011	2.51	6.37	3.85
Gambia	2006	1.47	3.35	2.65
Ghana	2007	1.83	4.83	2.14
Ghana	2013	2.93	7.48	4.23
Guinea	2006	1.49	4.10	2.02
Guinea Bissau	2006	1.51	4.09	2.18
Kenya	2007	1.37	3.80	1.90
Kenya	2013	2.62	7.31	4.79
Morocco	2013	2.22	5.96	3.86
Mozambique	2007	2.93	6.04	3.50
Namibia	2006	1.95	4.96	3.16
Namibia	2014	3.66	9.41	5.50
Nigeria	2007	1.95	5.49	2.94
Nigeria	2014	4.20	12.43	5.88
Rwanda	2006	1.93	4.53	3.12
Senegal	2007	1.63	4.14	2.14
Senegal	2014	2.12	5.29	3.76
South Africa	2007	1.56	4.09	2.21
South Sudan	2014	1.69	4.59	3.70
Sudan	2014	0.62	1.82	0.10
Swaziland	2006	1.27	3.33	2.46
Tanzania	2006	1.98	5.34	3.40
Tanzania	2013	2.20	5.67	3.66
Tunisia	2013	2.27	6.38	2.98
Uganda	2006	1.62	4.35	2.39
Zambia	2007	1.29	3.23	1.69
Zambia	2013	1.96	5.46	3.19
Zimbabwe	2011	1.90	5.43	2.38

markets. The dispersion measures for $log\left(1+\tau_{Ysi}/1+\overline{\tau}_{Ysi}\right)$ (tables 11 and 12) and $log\left(1+\tau_{Ksi}/1+\overline{\tau}_{Ksi}\right)$ (tables 13 and 14) are presented on the Annex of this work.

Figure 5: Kernel Distribution of $log(TFPR_{si}/\overline{TFPR_s})$ for Latin America in 2006

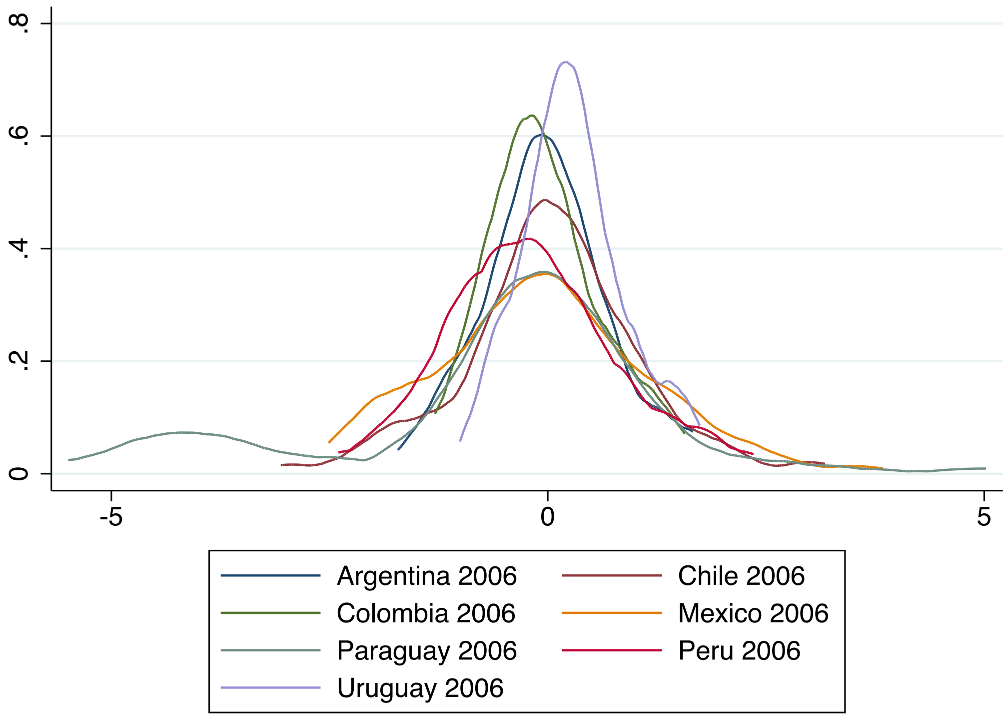


Figure 6: Kernel Distribution of $log(TFPR_{si}/\overline{TFPR_s})$ for Latin America in 2010

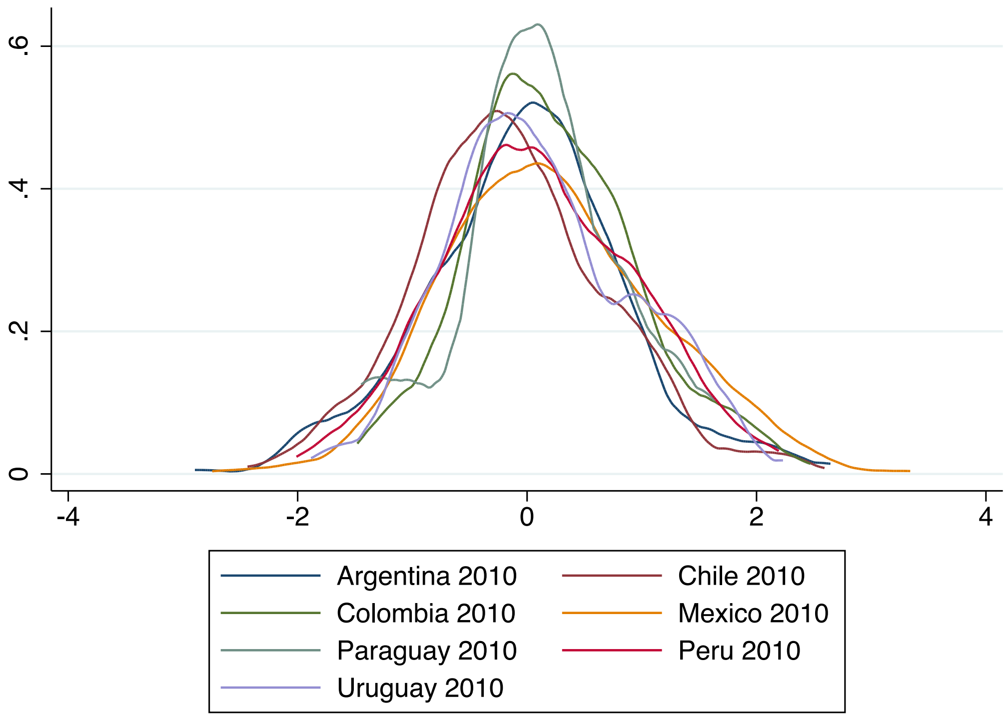


Figure 7: Kernel Distribution of $log(TFPR_{si}/\overline{TFPR_s})$ for Africa in 2006/7

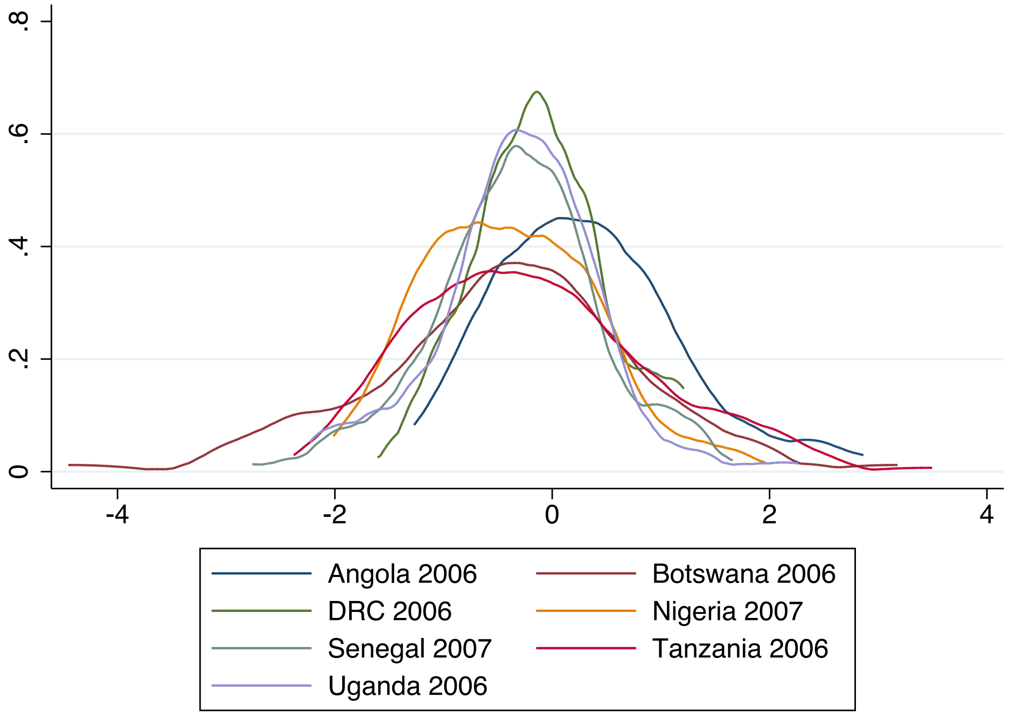


Figure 8: Kernel Distribution of $log(TFPR_{si}/\overline{TFPR_s})$ for Africa in 2010/13/14

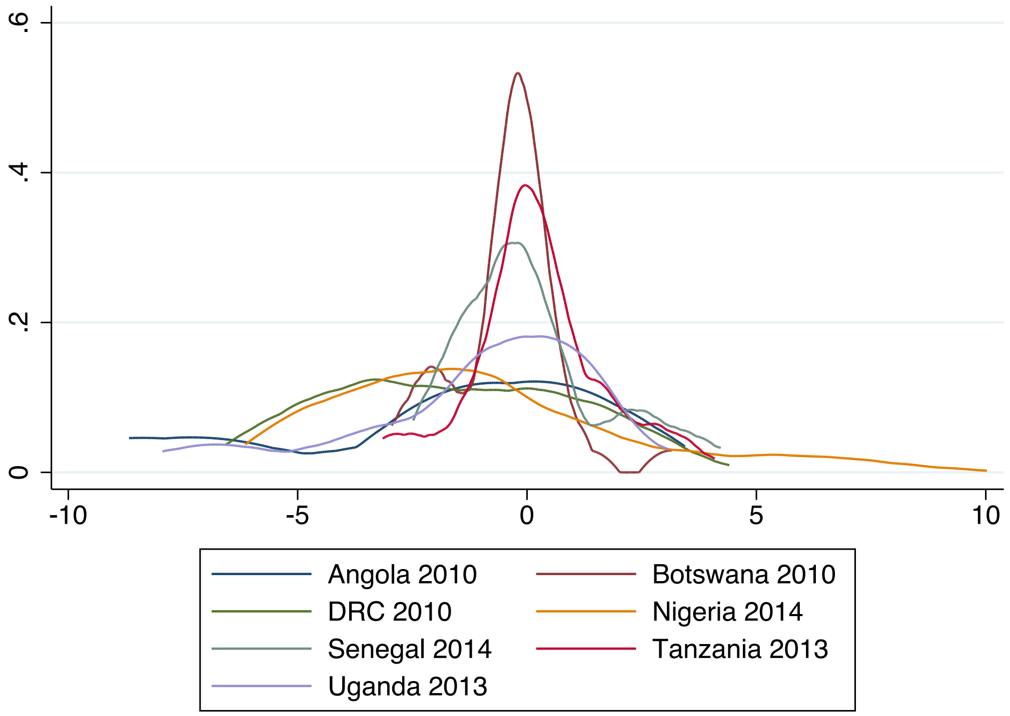


Table 3: Dispersion Measures of $log(TFPR_{si}/\overline{TFPR_s})$ for Latin American countries

	year	sd	90th-10th	75th-25th
Antigua and Barbuda	2010	0.30	0.77	0.10
Argentina	2006	0.76	2.07	1.00
Argentina	2010	0.83	2.05	0.90
Bahamas	2010	0.68	2.03	0.51
Barbados	2010	0.53	1.55	0.46
Belize	2010	0.69	1.71	0.80
Bolivia	2006	0.83	1.81	0.57
Brazil	2009	1.18	3.35	1.63
Chile	2006	0.97	2.17	0.97
Chile	2010	0.80	1.89	0.97
Colombia	2010	0.65	2.00	0.65
Costa Rica	2010	0.96	2.42	1.40
Dominican Republic	2010	1.02	2.30	1.21
Ecuador	2006	0.85	2.08	1.16
Ecuador	2010	1.02	2.82	1.04
El Salvador	2006	1.72	4.36	1.86
El Salvador	2010	0.65	1.31	0.63
Grenada	2010	0.26	0.84	0.03
Guatemala	2006	1.89	5.48	2.07
Guatemala	2010	0.75	1.70	0.86
Guyana	2010	0.88	2.00	0.80
Honduras	2006	0.99	2.80	1.01
Honduras	2010	1.69	4.13	2.33
Jamaica	2010	0.59	1.58	0.60
Mexico	2006	1.22	3.19	1.54
Mexico	2010	0.89	2.42	1.13
Nicaragua	2006	2.04	5.20	2.81
Nicaragua	2010	0.67	1.67	0.58
Panama	2006	1.24	2.49	1.28
Panama	2010	0.83	2.09	0.44
Paraguay	2006	1.92	5.21	1.49
Paraguay	2010	0.74	2.20	0.73
Peru	2006	1.00	2.78	1.22
Peru	2010	0.85	2.21	1.16
Suriname	2010	0.57	1.50	0.51
Uruguay	2006	0.62	1.79	0.82
Uruguay	2010	0.80	2.10	1.07

Table 4: Dispersion Measures of $log(TFPR_{si}/\overline{TFPR_s})$ for African countries

	year	sd	90th-10th	75th-25th
Angola	2006	0.85	2.13	1.15
Angola	2010	3.92	9.22	4.17
Botswana	2006	1.24	2.82	1.40
Botswana	2010	1.02	2.32	0.86
Burkina Faso	2009	0.52	1.09	0.64
Burundi	2006	0.91	2.12	1.31
Burundi	2014	0.81	1.42	0.33
Cameroon	2009	1.84	5.74	1.52
Ivoire	2009	1.14	2.17	1.49
DRC	2006	0.63	1.85	0.69
DRC	2010	2.79	7.12	5.28
DRC	2013	1.70	4.39	2.25
Egypt	2013	0.98	2.37	1.19
Ethiopia	2011	1.62	3.56	1.66
Gambia	2006	1.02	2.73	0.81
Ghana	2007	0.97	2.53	1.32
Ghana	2013	1.52	3.74	1.89
Guinea	2006	0.79	1.69	0.89
Guinea Bissau	2006	0.95	2.60	0.94
Kenya	2007	0.64	1.55	0.81
Kenya	2013	1.45	3.86	2.10
Morocco	2013	1.60	4.97	2.06
Mozambique	2007	1.66	3.55	2.02
Namibia	2006	1.02	2.16	1.10
Namibia	2014	2.19	3.85	1.90
Nigeria	2007	0.80	1.98	1.21
Nigeria	2014	3.54	9.13	4.21
Rwanda	2006	0.99	2.43	0.96
Senegal	2007	0.77	1.82	0.82
Senegal	2014	1.80	4.98	2.92
South Africa	2007	0.75	1.83	0.82
South Sudan	2014	1.04	2.60	1.15
Sudan	2014	0.45	1.28	0.00
Swaziland	2006	0.72	1.73	1.35
Tanzania	2006	1.08	2.89	1.38
Tanzania	2013	1.76	4.89	2.75
Tunisia	2013	2.20	5.70	3.31
Uganda	2006	0.75	1.93	0.82
Zambia	2007	0.86	1.74	0.95
Zambia	2013	1.29	3.62	1.79
Zimbabwe	2011	1.11	2.66	1.09

5.1.1 Reallocation Gains

Knowing the extent of misallocation helps to acknowledge how severe are the idiosyncratic distortions in each economy. However, the interest on misallocation arises due to its negative impact on the aggregate TFP. Hsieh and Klenow (2009, p. 1420-1421) present a way to estimate by how much the TFP could be increased (%) by equalizing the distortions on a given economy [18] with the caveat that this kind of estimate is highly sensitive to measurement errors or model misspecification.

The estimates of possible TFP gains resulting from a reallocation are presented on table 5 for a subsample of countries. Nonetheless, the reader should be informed that the possible problems highlighted by Hsieh and Klenow might be more serious in the present exercise due to the use of survey data. Moreover, the estimates here used the version for $\overline{TFPR}_s$ presented in Hsieh and Klenow (2013) what prevents the direct comparison with similar attempts on the literature, e.g. Busso et al. (2013).

While some estimates seems in line with the original results for China and India from Hsieh and Klenow others are much higher (e.g. Chile in 2006, Cameroon in 2009). The latter should reflect problems of sample size that were not controlled[19].

[18]This is done using the formula $\frac{Y}{Y_{\text{efficient}}} = \frac{TFP}{TFP_{\text{efficient}}} = \prod_{s=1}^{S} \left[\sum_{i=1}^{M_s} \left(\frac{A_{si}}{\overline{A_s}} \frac{\overline{TFPR}_s}{TFPR_{si}} \right)^{\sigma-1} \right]^{\theta_s/(\sigma-1)}$ and calculating $((TFP_{\text{eff}}/TFP_{\text{act}}) - 1) \times 100\%$.

[19]Some countries were previously excluded from the sample because, after all calculations, they got so few observation that the firm distribution degenerated (Colombia in 2006, Bolivia and Dominica in 2010, and Djibouti and Uganda in 2013). This does not seem to be the case of this odd estimates of TFP gains.

Table 5: Reallocation Gains

	year	TFP Gain (%)		year	TFP Gain (%)
Antigua and Barbuda	2010	7.15	Angola	2006	264.65
Argentina	2006	89.84	Botswana	2006	219.04
Argentina	2010	77.31	Botswana	2010	241.11
Bahamas	2010	31.37	Burkina Faso	2009	76.93
Barbados	2010	28.48	Burundi	2006	80.11
Belize	2010	10.19	Burundi	2014	239.97
Bolivia	2006	71.79	Cameroon	2009	362.73
Brazil	2009	267.01	Ivoire	2009	179.09
Chile	2006	313.23	DRC	2006	60.64
Chile	2010	77.72	DRC	2013	313.15
Colombia	2010	38.59	Ethiopia	2011	230.84
Costa Rica	2010	180.13	Gambia	2006	121.93
Dominican Republic	2010	128.17	Ghana	2007	90.93
Ecuador	2006	46.11	Guinea	2006	87.49
Ecuador	2010	69.29	Guinea Bissau	2006	103.74
El Salvador	2010	39.63	Kenya	2007	87.60
Grenada	2010	4.79	Kenya	2013	72.17
Guatemala	2006	300.43	Morocco	2013	175.49
Guatemala	2010	25.49	Namibia	2006	191.13
Guyana	2010	19.45	Namibia	2014	163.59
Honduras	2006	173.85	Nigeria	2007	106.04
Honduras	2010	29.11	Rwanda	2006	97.38
Jamaica	2010	72.68	Senegal	2007	85.46
Mexico	2010	48.54	South Africa	2007	125.83
Nicaragua	2010	94.89	South Sudan	2014	44.80
Panama	2010	14.03	Sudan	2014	3.37
Paraguay	2010	47.80	Swaziland	2006	57.62
Peru	2006	398.51	Tanzania	2006	152.05
Peru	2010	81.56	Uganda	2006	58.06
Suriname	2010	48.97	Zambia	2007	58.21
Uruguay	2006	106.62	Zambia	2013	99.82
Uruguay	2010	110.24	Zimbabwe	2011	274.13

5.2 Misallocation Determinants

While analysing the extent of misallocation on the selected economies, two sorts of indicators were calculated: on firm level, quantities relative to the average in a firm's sector indicate the size of misallocation confronted by that firm; on country level, dispersion measures of that quantities estimated for each firm summarize the size of misallocation within the country.

To study the determinants of misallocation, two approaches will be followed. On firm-level quantities ($ln(A_{si}/\overline{A}_s)$, $ln(TFPR_{si}/\overline{TFPR}_s)$, $ln(1+\tau_{Ysi}/1+\overline{\tau}_{Ys})$, $ln(1+\tau_{Ksi}/1+\overline{\tau}_{Ks})$) are regressed against variables that represent entrepreneurs perceptions about barriers to business development (from WBES database). On country level, misallocation summary measures (standard deviation for each of the four quantities) are regressed against variables that represent experts perceptions and factual information about the business environment (from Doing Business database).

Table 6 presents the results for the firm level regressions. For each misallocation indicator two regressions are presented, one with all the possible covariates and another only with the covariates that are statistically significant. Because these regressions are using relative quantities as dependent variables, the sign of the coefficients is not a matter for interpretation; what is important is that a significant coefficient, positive or negative, is moving a variable from its average and, therefore, indicating misallocation.

Regressions (1) and (2) show that physical productivity $(ln(A_{si}/\overline{A}_s)$ can be related to "Customs and Trade Regulations", "Crime theft and disorder", and "Business Licensing and Operation Permits".

Regressions (3) and (4) present that revenue productivity $(ln(TFPR_{si}/\overline{TFPR}_s)$ can also be related to "Customs and Trade Regulations" but additionally with "Anti-competitive or informal practices" and "Access to Financing". For both productivities the coefficient for the dummy variable differentiating African economies from the Latin American ones is not significant.

Regressions (5), (6), and (7) are concerned with the output distortion $(ln(\tau_{Ysi}/\overline{\tau}_{Ys})$, again "Customs and Trade Regulations" is a significant variable but also are "Tax Rates" and the dummy variable for African economies. The variable "Legal system/conflict resolution" is significant in one of the specifications (7), however when this variable is included "Tax Rates" become not significant.

Finally, for the input distortion $(ln(\tau_{Ksi}/\overline{\tau}_{Ks})$ one has the regressions (8) and (9). In this case the significant variables are "Anti-competitive or informal practices" and "Access to Financing" paralleling the specification for physical productivity.

Even though regression analysis was able to identify some significant variables, for all cases considered the explicative power of those variables was considerably low, as shown by the R^2 levels on table 6.

Moving to the country-level analysis, the standard deviation of misallocation indicators are used as dependent variable and regressed against the Doing Business variables. This imposes a challenge since a much smaller sample than that used on firm-level analysis is available (80 countries) with far more numerous possible covariates, with the additional problem that variables related to "Getting Electricity" from Doing Business are available only to a fraction of the sample (44 countries).

Since the variables related to "Getting Electricity" are significant most of the times, the regressions were estimated considering a restricted sample with these variables and in a full sample without these variables. Moreover, given the quantity of possible covariates and the fact that many of them are not statistically significant, presenting a first regression with all covariates as done before is unmanageable in the present case. So the choice made was to show regressions for each independent variable in a separate table considering only covariates with a reasonable p-value.

Table 7 shows the regressions of TFPQ. In all cases considered variables related to international trade ("time to export" and "cost to import") are significant. A bigger "time to export" increases TFPQ dispersion, what follows arguments that favours international trade and economic openness as drivers of productivity. Nevertheless, the negative - despite small - coefficient of "import price" means that a higher import price would reduce TFPQ dispersion, what goes in the opposite direction.

As mentioned before, variables related to "Getting Electricity" are important for the restricted sample for which they are available. In the case of TFPQ, the number of procedures necessary to get electricity increases the dispersion therefore reducing aggregate productivity.

Table 8 shows the regressions of TFPR. The pattern seen before for the variables related to "trading across borders" is repeated in all regressions presented. For the full

Table 6: Regressions of Firm-level Misallocation Indicators

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	ln_Asi_s	ln_Asi_s	ln_TFPRsi_s	ln_TFPRsi_s	ln_Tau_Ysi_s	ln_Tau_Ysi_s	ln_Tau_Ysi_s	ln_Tau_Ksi_s	ln_Tau_Ksi_s
Electricity	-0.00849		-0.0714		0.0325			-0.0472	
	(-0.12)		(-1.32)		(0.83)			(-0.66)	
Customs and Trade Regulations	0.517***	0.531***	0.245***	0.237***	-0.188***	-0.138***	-0.148***	0.146**	
	(3.96)	(4.07)	(2.64)	(3.44)	(-2.65)	(-2.57)	(-2.22)	(1.73)	
Anti-competitive or informal practices	0.0809		-0.156***	-0.167***	0.0146			-0.193***	-0.179***
	(1.17)		(-2.18)	(-2.76)	(0.32)			(-2.58)	(-2.36)
Access to Land	0.0865		-0.00919		-0.00687			-0.0552	
	(0.95)		(-0.15)		(-0.15)			(-0.63)	
Crime, theft and disorder	-0.254***	-0.177***	-0.0506		0.0220			-0.140**	
	(-3.33)	(-2.68)	(-0.88)		(0.63)			(-1.81)	
Tax rates	-0.105		-0.0245		0.0474	0.0955***		0.0660	
	(-1.15)		(-0.47)		(1.32)	(2.37)		(0.78)	
Tax administration	-0.0462		0.105		-0.0155			0.0958	
	(-0.38)		(1.39)		(-0.30)			(0.79)	
Business Licensing and Operating Permits	-0.290***	-0.255***	0.00962		0.00403			0.0183	
	(-2.94)	(-2.00)	(0.20)		(0.10)			(0.27)	
Corruption	0.0700		-0.00520		0.00895			-0.0312	
	(0.84)		(-0.09)		(0.20)			(-0.50)	
Access to Financing	-0.0477		0.0911***	0.120***	-0.00515			0.182***	0.284***
	(-0.73)		(2.05)	(2.58)	(-0.14)			(2.64)	(3.29)
Labor Regulations	-0.00546		-0.180**		0.0779			-0.183***	
	(-0.07)		(-1.70)		(1.06)			(-2.11)	
Skills and Education of Available Workers	0.0306		0.118***		-0.0658			0.0670	
	(0.41)		(2.03)		(-1.31)			(1.04)	
Legal system/conflict resolution	0.0557		-0.0331		0.0813*		0.127***	0.173	
	(0.54)		(-0.38)		(1.44)		(2.45)	(1.37)	
Dummy for African countries	-0.242*		-0.191*		0.281***	0.293***	0.326***	0.190	
	(-1.55)		(-1.46)		(2.91)	(7.93)	(9.56)	(1.36)	
Intercept	-5.296***	-5.508***	0.271**		-0.0415			0.555***	0.478***
	(-27.25)	(-41.12)	(1.70)		(-0.32)			(2.44)	(3.37)
N	7771	8717	7771	8214	7771	8818	8211	7771	8782
R^2	0.139	0.110	0.147	0.141	0.0736	0.0904	0.0911	0.101	0.0551
F	4.328	7.210	2.490	5.583	1.962	42.49	37.78	3.868	7.101

t statistics in parentheses
* $p < 0.15$, ** $p < 0.10$, *** $p < 0.05$

Table 7: Regressions of Country-level Misallocation Indicators - TFPQ - $sd\left(ln\left(\frac{A_{si}}{\overline{A}_s}\right)\right)$

	(1) sd(Asi/As_bar)	(2) sd(Asi/As_bar)	(3) sd(Asi/As_bar)	(4) sd(Asi/As_bar)	(5) sd(Asi/As_bar)
Starting a Business - Cost (% of income per capita)	0.00221	0.00324**	0.00249***	-0.00000768	
	(1.08)	(1.96)	(2.09)	(-0.01)	
Getting Electricity - Procedures (number)	0.265***	0.207***	0.189***		
	(2.42)	(3.78)	(3.98)		
Getting Electricity - Cost (% of income per capita)	0.0000157				
	(0.46)				
Trading Across Borders - Time to export (days)	0.0702***	0.0643***	0.0631***	0.0436***	0.0472***
	(3.63)	(3.52)	(3.50)	(3.61)	(4.10)
Trading Across Borders - Cost to export (deflated US$ per container)	0.000359			-0.0000399	
	(0.87)			(-0.21)	
Trading Across Borders - Cost to import (deflated US$ per container)	-0.000561*	-0.000252***	-0.000260***	-0.000166	-0.000181***
	(-1.53)	(-2.19)	(-2.29)	(-1.13)	(-2.73)
Enforcing Contracts - Cost (% of claim)	-0.00535	-0.00519		0.00554	
	(-0.56)	(-0.66)		(1.20)	
Intercept	-0.422			1.214***	1.299***
	(-0.65)			(4.65)	(5.54)
N	44	44	44	80	80
R^2	0.512	0.865	0.864	0.199	0.180
F	5.406	50.05	63.34	3.665	8.453

t statistics in parentheses
* $p < 0.15$, ** $p < 0.10$, *** $p < 0.05$

sample, only these variables are significant (regressions (6) to (8)).

Considering the restricted sample, the time spent to get electricity is significant and imply that longer times are associated with higher dispersion of TFPR. The coefficient related to the cost of "enforcing contracts" is presented on the table with a negative value in the regressions 1 and 2, but with a p-value higher than 15%.

Table 9 shows the regressions of sd($ln\,(1+\tau_{Ysi}/1+\overline{\tau}_{Ys})$). Again, the pattern seen before for the variables related to "trading across borders" is repeated in all regressions. However, in this case the "time to export" is a relevant covariate, with longer times increasing the dispersion and thus reducing productivity.

Table 8: Regressions of Country-level Misallocation Indicators - $sd\left(ln\left(\frac{TFPR_{si}}{TFPR_s}\right)\right)$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Getting Electricity - Procedures (number)	0.153***	0.0833**	0.0674*					
	(2.03)	(1.77)	(1.59)					
Getting Electricity - Cost (% of income per capita)	0.0000377**	0.0000350**	0.0000228**	0.0000201*				
	(1.89)	(1.77)	(1.83)	(1.59)				
Registering Property - Cost (% of property value)	0.0513***					0.0176	0.0211*	
	(2.20)					(1.13)	(1.46)	
Trading Across Borders - Time to export (days)	0.0420***	0.0427***	0.0405***	0.0508***	0.0525***	0.0218***	0.0229***	0.0241***
	(3.33)	(3.38)	(3.30)	(4.80)	(4.89)	(2.77)	(3.00)	(3.15)
Trading Across Borders - Cost to import (deflated US$ per container)	-0.000170***	-0.000231***	-0.000229***	-0.000262***	-0.000239***	-0.000110***	-0.000107***	-0.0000998***
	(-2.11)	(-2.78)	(-2.76)	(-3.20)	(-2.91)	(-2.47)	(-2.42)	(-2.25)
Enforcing Contracts - Cost (% of claim)	-0.00881	-0.00501				0.00185		
	(-1.32)	(-0.80)				(0.64)		
Getting Electricity - Time (days)		0.00411**	0.00400**	0.00602***	0.00586***			
		(2.01)	(1.97)	(3.72)	(3.56)			
Intercept	-0.354					0.684***	0.697***	0.793***
	(-0.81)					(4.01)	(4.14)	(5.07)
N	44	44	44	44	44	80	80	80
R^2	0.472	0.826	0.823	0.812	0.800	0.143	0.138	0.114
F	5.508	30.11	36.33	43.13	54.63	3.122	4.060	4.950

t statistics in parentheses
* $p < 0.15$, ** $p < 0.10$, *** $p < 0.05$

Table 9: Regressions of Country-level Misallocation Indicators - τ_Y - $sd\left(ln\left(\frac{1+\tau_{Ysi}}{1+\overline{\tau}_{Ys}}\right)\right)$

	(1)	(2)	(3)	(4)	(5)
Getting Electricity - Procedures (number)	0.139***	0.133**	0.153***		0.153***
	(2.05)	(1.99)	(2.31)		(2.31)
Getting Electricity - Cost (% of income per capita)	0.0000339**	0.0000248***	0.0000291***		0.0000291***
	(1.90)	(2.16)	(2.58)		(2.58)
Registering Property - Procedures (number)	0.136***	0.139***	0.149***	0.0545*	0.149***
	(2.86)	(2.96)	(3.17)	(1.56)	(3.17)
Registering Property - Cost (% of property value)	0.0344*	0.0293*		0.0119	
	(1.61)	(1.48)		(0.78)	
Trading Across Borders - Time to export (days)	0.0374***	0.0360***	0.0378***	0.0229***	0.0378***
	(3.55)	(3.52)	(3.67)	(3.04)	(3.67)
Trading Across Borders - Cost to export (deflated US$ per container)	-0.000187***	-0.000188***	-0.000210***	-0.000150***	-0.000210***
	(-2.28)	(-2.31)	(-2.57)	(-2.67)	(-2.57)
Enforcing Contracts - Cost (% of claim)	-0.00400			0.00360	
	(-0.67)			(1.23)	
Intercept	-1.303***	-1.345***	-1.361***	0.193	-1.361***
	(-2.78)	(-2.92)	(-2.91)	(0.67)	(-2.91)
N	44	44	44	80	44
R^2	0.596	0.591	0.567	0.201	0.567
F	7.603	8.928	9.963	3.727	9.963

t statistics in parentheses
* $p < 0.15$, ** $p < 0.10$, *** $p < 0.05$

For the restricted sample (regressions 1 to 3 and 5), the cost of getting electricity is again a relevant covariate. In its most parsimonious version (regression 3), dispersion of sd($ln\,(1 + \tau_{Ysi}/1 + \overline{\tau}_{Ys})$) is explained by "trading across borders" variables, "cost of getting electricity", and number of procedures to register property.

Table 10 shows the final regressions relative to sd($ln\,(1 + \tau_{Ksi}/1 + \overline{\tau}_{Ks})$). Few of the possible covariates are statistically significant so that the regressions presented considers only the restricted sample. While the variables related to "getting electricity" are significant, two variables related to resolving insolvency show up as relevant covariates. Considering that this is the input distortion, it is somewhat unexpected since covariates relative to labour market and access to finance were available but, once tested, showed themselves not statistically significant.

Table 10: Regressions of Country-level Misallocation Indicators - τ_K - $sd\left(ln\left(\frac{1+\tau_{Ksi}}{1+\overline{\tau}_{Ks}}\right)\right)$

	(1)	(2)	(3)
Getting Electricity - Time (days)	0.00429**	0.00503***	0.00376*
	(1.77)	(2.13)	(1.68)
Getting Electricity - Cost (% of income per capita)	0.0000416		
	(1.22)		
Registering Property - Cost (% of property value)	0.0617***	0.0552**	0.0458*
	(2.17)	(1.97)	(1.65)
Enforcing Contracts - Cost (% of claim)	-0.0149*	-0.0132	
	(-1.63)	(-1.45)	
Resolving Insolvency - Time (years)	-0.528***	-0.510***	-0.448***
	(-2.94)	(-2.83)	(-2.51)
Resolving Insolvency - Recovery rate (cents on the dollar)	-0.0307***	-0.0322***	-0.0243***
	(-2.48)	(-2.60)	(-2.15)
Intercept	3.580***	3.568***	2.877***
	(3.84)	(3.80)	(3.49)
N	39	39	39
R^2	0.353	0.323	0.279
F	2.908	3.146	3.296

t statistics in parentheses
* $p < 0.15$, ** $p < 0.10$, *** $p < 0.05$

6 Conclusion

This paper studied the extent and determinants of misallocation for Latin American and African economies. The main results indicate that, in contrast with previous evidence documented in the literature, distortions affecting international trade and related to business licensing and operation permits as well as to informality are key factors damaging TFP in the set of developing countries we have considered.

The extent of resource misallocation is more severe in Africa than in Latin America. Moreover, while most of the Latin American economies improved or maintained their situation in terms of the dispersion of productivity (a misallocation indicator) the opposite happened in the African economies. When looking at the determinants of misallocation, barriers to international and some factors related to financing ("Access to Financing") and informality ("Anti-competitive or informal practices") were identified as key relevant issues.

References

Banerjee, A. V. and Duflo, E. (2005). Chapter 7 - Growth Theory through the Lens of Development Economics. In Aghion, P. and Durlauf, S. N., editors, *Handbook of Economic Growth*, volume 1, Part A, pages 473–552. Elsevier.

Bartelsman, E. J. and Gray, W. (1996). The NBER Manufacturing Productivity Database. Working Paper 205, National Bureau of Economic Research.

Busso, M., Madrigal, L., and Pagés, C. (2013). Productivity and resource misallocation in Latin America. *The B.E. Journal of Macroeconomics*, 13(1).

Caselli, F. (2005). Chapter 9 Accounting for Cross-Country Income Differences. In Durlauf, P. A. a. S. N., editor, *Handbook of Economic Growth*, volume 1, Part A, pages 679–741. Elsevier.

Caselli, F. and Coleman, II, W. J. (2006). The World Technology Frontier. *American Economic Review*, 96(3):499–522.

Foster, L., Haltiwanger, J., and Syverson, C. (2008). Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability? *American Economic Review*, 98(1):394–425.

García-Santana, M. and Ramos, R. (2015). Distortions and the size distribution of plants: evidence from cross-country data. *SERIEs*, 6(3):279–312.

Hall, R. E. and Jones, C. I. (1999). Why do Some Countries Produce So Much More Output Per Worker than Others? *The Quarterly Journal of Economics*, 114(1):83–116.

Hopenhayn, H. and Rogerson, R. (1993). Job Turnover and Policy Evaluation: A General Equilibrium Analysis. *Journal of Political Economy*, 101(5):915–938.

Hopenhayn, H. A. (2014). Firms, Misallocation, and Aggregate Productivity: A Review. *Annual Review of Economics*, 6(1):735–770.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and Manufacturing TFP in China and India. *The Quarterly Journal of Economics*, 124(4):1403–1448.

Hsieh, C.-T. and Klenow, P. J. (2013). Misallocation and Manufacturing TFP in China and India - Correction Appendix. Technical report.

Inklaar, R., Lashitew, A. A., and Timmer, M. P. (2015). The Role of Resource Misallocation in Cross-country Differences in Manufacturing Productivity - Research database - University of Groningen. GGDC Working Papers GD-143, Groningen Growth and Development Centre, University of Groningen, Groningen.

Kalemli-Ozcan, S. and Sorensen, B. E. (2012). Misallocation, Property Rights, and Access to Finance: Evidence from Within and Across Africa. NBER Working Paper 18030, National Bureau of Economic Research, Inc.

Lucas, R. E. (1978). On the Size Distribution of Business Firms. *Bell Journal of Economics*, 9(2):508–523.

Mankiw, N. G., Romer, D., and Weil, D. N. (1992). A Contribution to the Empirics of Economic Growth. *The Quarterly Journal of Economics*, 107(2):407–437.

Oberfield, E. (2013). Productivity and misallocation during a crisis: Evidence from the Chilean crisis of 1982. *Review of Economic Dynamics*, 16(1):100–119.

Parente, S. L. and Prescott, E. C. (1994). Barriers to Technology Adoption and Development. *Journal of Political Economy*, 102(2):298–321.

Parente, S. L. and Prescott, E. C. (2002). *Barriers to Riches*. The MIT Press, Cambridge, Mass.

Restuccia, D. and Rogerson, R. (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics*, 11(4):707–720.

Restuccia, D. and Rogerson, R. (2013). Misallocation and productivity. *Review of Economic Dynamics*, 16(1):1–10.

World Bank, . (2009a). Brazil Enterprise Surveys Data Set. Technical report, World Bank, Washington.

World Bank, . (2009b). Enterprise Survey and Indicator Surveys - Sampling Methodology. Technical report, World Bank, Washington.

World Bank, . (2010). The Bolivia 2010 Enterprise Surveys Data Set. Technical report, World Bank, Washington.

World Bank, . (2014). Implementing Enterprise Surveys in Latin America and the Caribbean. Latin America and Caribbean Notes 10, World Bank, Washington.

7 Annex: Tables with Dispersion Measures of τ_{Ysi} and τ_{Ksi} and correlation between misallocation indicators

Table 11: Dispersion Measures of $log\left(\frac{1+\tau_{Ysi}}{1+\bar{\tau}_{Ysi}}\right)$ for Latin American countries

	year	sd	90th-10th	75th-25th
Antigua and Barbuda	2010	0.31	0.57	0.19
Argentina	2006	0.69	1.59	0.79
Argentina	2010	0.71	1.49	0.68
Bahamas	2010	0.82	1.98	0.78
Barbados	2010	0.46	1.00	0.51
Belize	2010	0.55	1.40	0.94
Bolivia	2006	0.75	1.89	0.31
Brazil	2009	0.77	1.97	0.94
Chile	2006	0.83	1.91	0.81
Chile	2010	0.76	1.85	0.75
Colombia	2010	0.60	1.56	0.67
Costa Rica	2010	0.92	2.71	1.04
Dominican Republic	2010	0.79	1.54	1.09
Ecuador	2006	1.00	2.46	0.92
Ecuador	2010	0.87	2.13	1.45
El Salvador	2006	1.79	4.18	1.78
El Salvador	2010	0.69	1.53	0.95
Grenada	2010	0.23	0.70	0.31
Guatemala	2006	1.20	2.61	0.77
Guatemala	2010	0.59	1.12	0.65
Guyana	2010	0.83	1.90	0.79
Honduras	2006	0.92	2.32	1.02
Honduras	2010	1.10	2.58	1.43
Jamaica	2010	0.41	0.65	0.35
Mexico	2006	1.27	3.31	1.54
Mexico	2010	0.78	1.76	1.02
Nicaragua	2006	1.99	4.88	2.51
Nicaragua	2010	0.79	2.30	0.65
Panama	2006	1.06	2.18	1.18
Panama	2010	1.01	2.29	0.80
Paraguay	2006	2.07	5.33	1.48
Paraguay	2010	0.73	2.12	0.81
Peru	2006	1.17	3.23	1.24
Peru	2010	0.84	2.11	0.99
Suriname	2010	0.37	1.06	0.34
Uruguay	2006	0.77	1.79	0.73
Uruguay	2010	0.72	1.82	0.74

Table 12: Dispersion Measures of $log\left(\frac{1+\tau_{Ysi}}{1+\bar{\tau}_{Ysi}}\right)$ for African countries

	year	sd	90th-10th	75th-25th
Angola	2006	0.59	1.37	0.59
Angola	2010	3.65	8.30	2.70
Botswana	2006	1.24	3.00	1.53
Botswana	2010	0.95	2.17	0.97
Burkina Faso	2009	0.70	1.65	0.08
Burundi	2006	0.65	1.45	0.85
Burundi	2014	0.74	1.23	0.35
Cameroon	2009	2.20	4.89	1.64
Ivoire	2009	0.91	1.89	0.92
DRC	2006	0.50	1.20	0.51
DRC	2010	3.13	7.13	4.45
DRC	2013	1.62	4.72	1.73
Egypt	2013	1.08	2.43	1.28
Ethiopia	2011	1.41	3.70	1.59
Gambia	2006	0.56	1.18	0.63
Ghana	2007	0.97	2.72	0.78
Ghana	2013	1.32	2.93	1.62
Guinea	2006	0.76	1.98	0.99
Guinea Bissau	2006	0.84	1.77	0.81
Kenya	2007	0.74	1.93	0.97
Kenya	2013	1.43	3.51	1.67
Morocco	2013	1.11	2.56	2.00
Mozambique	2007	1.22	2.13	1.08
Namibia	2006	1.15	2.25	0.81
Namibia	2014	2.44	5.64	1.31
Nigeria	2007	0.91	2.46	1.39
Nigeria	2014	3.62	8.89	3.89
Rwanda	2006	1.00	2.39	0.89
Senegal	2007	0.56	1.49	0.70
Senegal	2014	1.19	3.22	1.16
South Africa	2007	0.58	1.39	0.79
South Sudan	2014	0.99	2.24	0.76
Sudan	2014	0.10	0.33	0.00
Swaziland	2006	0.62	1.54	0.97
Tanzania	2006	0.98	2.66	1.48
Tanzania	2013	1.30	3.61	1.29
Tunisia	2013	1.46	2.82	1.21
Uganda	2006	0.76	1.96	0.82
Zambia	2007	0.74	1.84	0.98
Zambia	2013	1.22	3.19	1.84
Zimbabwe	2011	1.26	2.63	0.97

Table 13: Dispersion Measures of $log\left(\frac{1+\tau_{Ksi}}{1+\overline{\tau}_{Ksi}}\right)$ for Latin American countries

	year	sd	90th-10th	75th-25th
Antigua and Barbuda	2010	0.22	0.50	0.26
Argentina	2006	1.18	2.63	1.31
Argentina	2010	1.36	2.74	1.49
Bahamas	2010	0.78	2.33	1.07
Barbados	2010	0.71	1.32	0.78
Belize	2010	0.68	1.33	0.66
Bolivia	2006	1.02	2.39	0.56
Brazil	2009	1.82	4.83	2.63
Chile	2006	1.50	3.60	1.58
Chile	2010	1.38	3.11	1.56
Colombia	2010	1.01	2.36	0.89
Costa Rica	2010	1.39	3.18	1.56
Dominican Republic	2010	1.89	2.64	1.40
Ecuador	2006	1.12	2.87	1.39
Ecuador	2010	1.44	2.42	1.04
El Salvador	2006	1.53	4.13	1.88
El Salvador	2010	1.07	3.28	1.37
Grenada	2010	0.54	1.63	0.42
Guatemala	2006	2.75	7.45	2.28
Guatemala	2010	1.30	3.07	2.06
Guyana	2010	0.80	1.89	0.63
Honduras	2006	1.41	3.26	1.77
Honduras	2010	1.26	2.65	1.59
Jamaica	2010	1.08	2.56	1.32
Mexico	2006	1.59	3.96	1.86
Mexico	2010	1.19	2.54	1.02
Nicaragua	2006	1.63	3.93	1.92
Nicaragua	2010	1.62	4.84	1.57
Panama	2006	1.45	3.52	1.80
Panama	2010	0.61	0.75	0.69
Paraguay	2006	1.42	3.35	1.79
Paraguay	2010	1.66	4.55	2.11
Peru	2006	1.66	4.46	1.72
Peru	2010	1.39	3.66	1.62
Suriname	2010	0.64	1.10	0.73
Uruguay	2006	1.40	3.11	1.12
Uruguay	2010	0.66	1.46	0.88

Table 14: Dispersion Measures of $log\left(\frac{1+\tau_{Ksi}}{1+\bar{\tau}_{Ksi}}\right)$ for African countries

	year	sd	90th - 10th perc.	75th - 25th perc.
Angola	2006	1.77	4.35	2.32
Angola	2010	2.47	5.32	0.81
Botswana	2006	0.93	2.34	0.99
Botswana	2010	0.93	1.97	1.11
Burkina Faso	2009	1.68	3.61	2.30
Burundi	2006	1.51	3.66	2.25
Burundi	2014	1.21	2.76	0.62
Cameroon	2009	1.93	5.11	1.74
Ivoire	2009	1.88	5.17	2.23
DRC	2006	1.10	2.97	1.19
DRC	2010	2.32	6.12	2.32
DRC	2013	2.98	6.31	4.07
Egypt	2013	1.24	2.89	1.44
Ethiopia	2011	1.18	2.38	1.08
Gambia	2006	1.52	3.68	1.25
Ghana	2007	1.22	3.56	1.74
Ghana	2013	2.61	5.19	2.86
Guinea	2006	1.63	3.90	2.00
Guinea Bissau	2006	1.61	3.80	1.71
Kenya	2007	1.30	3.19	1.39
Kenya	2013	1.46	3.36	1.84
Morocco	2013	1.70	4.33	2.21
Mozambique	2007	2.03	4.81	2.49
Namibia	2006	0.99	2.69	1.02
Namibia	2014	2.16	5.49	1.96
Nigeria	2007	1.32	3.34	1.89
Nigeria	2014	3.85	7.65	4.65
Rwanda	2006	0.97	2.40	0.81
Senegal	2007	1.22	3.39	1.45
Senegal	2014	4.00	8.64	1.33
South Africa	2007	1.28	3.13	1.43
South Sudan	2014	1.20	2.74	0.97
Sudan	2014	1.80	5.63	0.00
Swaziland	2006	1.00	2.48	0.96
Tanzania	2006	1.77	4.01	1.81
Tanzania	2013	2.36	6.22	2.53
Tunisia	2013	3.03	8.05	5.16
Uganda	2006	1.23	3.06	1.35
Zambia	2007	1.36	2.78	1.50
Zambia	2013	2.55	6.29	3.53
Zimbabwe	2011	1.20	2.63	1.17

Table 15: Correlation Between Dispersion Measures of Misallocation Indicators

	(1)				
	$sd\left(ln\left(\frac{TFPR_{si}}{\overline{TFPR}_s}\right)\right)$	$sd\left(ln\left(\frac{A_{si}}{\overline{A}_s}\right)\right)$	$sd\left(ln\left(\frac{1+\tau_{Ysi}}{1+\overline{\tau}_{Ys}}\right)\right)$	$sd\left(ln\left(\frac{1+\tau_{Ksi}}{1+\overline{\tau}_{Ks}}\right)\right)$	Labour Share$_{<20}$
$sd\left(ln\left(\frac{TFPR_{si}}{\overline{TFPR}_s}\right)\right)$	1				
$sd\left(ln\left(\frac{A_{si}}{\overline{A}_s}\right)\right)$	0.905***	1			
$sd\left(ln\left(\frac{1+\tau_{Ysi}}{1+\overline{\tau}_{Ys}}\right)\right)$	0.901***	0.880***	1		
$sd\left(ln\left(\frac{1+\tau_{Ksi}}{1+\overline{\tau}_{Ks}}\right)\right)$	0.594***	0.412***	0.392***	1	
Labour Share$_{<20}$	-0.0706	-0.136	-0.0494	-0.0687	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$