Ph.D. Dissertation

ESSAYS ON THE INVESTMENT CLIMATE AND ITS EFFECTS ON FIRMS' EFFICIENCY

BY

JORGE PENA IZQUIERDO

Universidad Nacional de Educación a Distancia, Department of Economic Analysis II; Madrid, Spain, 2009.

Ph.D. Dissertation

ESSAYS ON THE INVESTMENT CLIMATE AND ITS EFFECTS ON FIRMS' EFFICIENCY

BY

JORGE PENA IZQUIERDO
BA in economics, Universidad de Oviedo.

Universidad Nacional de Educación a Distancia, Department of Economic Analysis II; Madrid, Spain, 2009.

Universidad Nacional de Educación a Distancia, Department of Economic Analysis II; Madrid, Spain, 2009.

PhD Dissertation prepared by: Jorge Pena Izquierdo, BA in economics, Universidad de Oviedo.

Supervisor: Álvaro Escribano Sáez, Department of Economics, Universidad Carlos III de Madrid.

Tutor of the dissertation: Ana Martín Marcos, Department of Economic Analysis II, Universidad Nacional de Educación a Distancia.

Dedication

To my parents, Mª Teresa and Julian, for having always stayed behind me and pushing me ahead with the best of gifts ever: education.

Acknowledgments

My first thoughts in this section necessarily go to Alvaro Escribano. I will never find the words to thank him for his time, and I much regret the fact that I will have little chance to pay him back for all his personal dedication.

This dissertation has its roots planted in 2005, when I first took contact with the investment climate surveys and all those professionals involved in the study of this fascinating field of knowledge at the World Bank, Washington DC. I am indebted with Jose Luis Guasch, Vivien Foster, Paulo Correa, Eric Manes, Maria Luisa Motta, Stefka Slavova, Gallina Andronova and many others for providing my with datasets and further insight on this interesting topic. My gratitude to all of them.

I have benefited from useful comments received at different stages of the project in seminars at the World Bank, Universidad Carlos III, UNED and other institutions. Thanks to all those that participated in them. I have found especially interesting comments received from Arial Pakes, James Levinsohn, Eric Verhoegen, Paulo Correa, Klaus Desmet, Eduardo Ley, Marialisa Motta, Jose Guillherme Reis, Isabel Sanchez and Rodolfo Stucchi.

I would like to acknowledge all the support, financial and intellectual, received from the Universidad Carlos III de Madrid, and especially from the department of economics. Further financial support was received from the Chair Telefonica-UC3M of economics of telecommunications and from the Government of the Community of Madrid, Spain.

Thanks to Ana Martin Marcos for her excellent guidance and for making the process of writing a dissertation infinitely easier.

My last words of gratitude go to Manuel de Orte who joined me in the early stages of a project that finally has made us both professionals. His useful and invaluable intellectual support is only comparable with his unusual and elevated sense of humor.

Thanks to the people that pushed, that pulled and that accompanied me on the way:

TABLE OF CONTENTS

INTRODUCTION

INTRODUCTION

The pursuit of greater competitiveness and a better investment climate is leading countries to undertake their own studies, to set priorities for intervention and reform. The most common instrument used has been firm-level surveys, know as Investment Climate Surveys (ICSs) from which both subjective evaluations of obstacles to economic performance and objective hard-data numbers on the quality of social and physical infrastructures available for transforming inputs into output with direct links to costs and productivity are elicited and imputed. A unique effort of data compilation done by the World Bank in close partnership with national statistical agencies has yielded a set of comparable Investment Climate Surveys covering close to 70,000 enterprises from more than 100 countries. This invaluable set of information has been used by an increasing number of applied researchers and scholars to try to place additional empirical underpinning on the role that the quality of the investment climate—broadly understood as physical infrastructures, access to finance, security, regulatory framework, competition and property rights—plays on economic success and growth from a microeconomic perspective.

The literature on investment climate has highlighted the importance of analyzing the different ways that the business environment in which firms operate may affect economic activity, particularly through incentives to invest. Successive improvements of the investment climate conditions increase returns of economic activity, and so create new investment opportunities and change the perceptions of the entrepreneurs on whether these will pay-off. Likewise, a better investment climate puts competitive pressure on sectors or firms that have received governmental protection and boosts processes of Schumpeterian *creative destruction*. In the other side, a bad investment climate, besides discouraging investment, may lead businesses to undertake inefficient and costly alternative investments such as security systems, own generators or inventories (see Dethier et al., 2008).

Notwithstanding all the research effort done so far with ICSs, taking into consideration that these data have been collected at considerable cost and given the variety of questions that might be highlighted with them, there may be the perception that so far the IC surveys have not been used to the full extent possible. All the emerging body of literature using investment climate data, including this dissertation, puts the way forward for a greater and better use of it and as a consequence for the analysis of economic success, growth, development and the traditional dichotomy between rich and poor regions from a microeconomic perspective. The contribution of this dissertation goes in this line of thought, by exploring the role of the investment climate on economic success, understood as the ability of the economies to transform inputs into outputs at the maximum level of efficiency possible (or total factor productivity), and trying to gather empirical evidence and patterns in the data that allow us to extend and derive causal inference on the determinants of economic success in successive stages of the research.

Most of the methodological aspects of this dissertation are based on the framework of the modern empirical industrial organization, and in particular on TFP literature. Concretely, I apply the recent econometric methodology developed for the World Bank by Escribano and Guasch (2005, 2008) and Escribano et al. (2008a and b) that departs from the work of Olley and Pakes

(1996). defining and using country by country homogeneous quantitative techniques to first identify and second assess the contributions to within industry aggregate productivities (or total factor productivities) of having restrictive business practices coming from bad investment climate conditions.

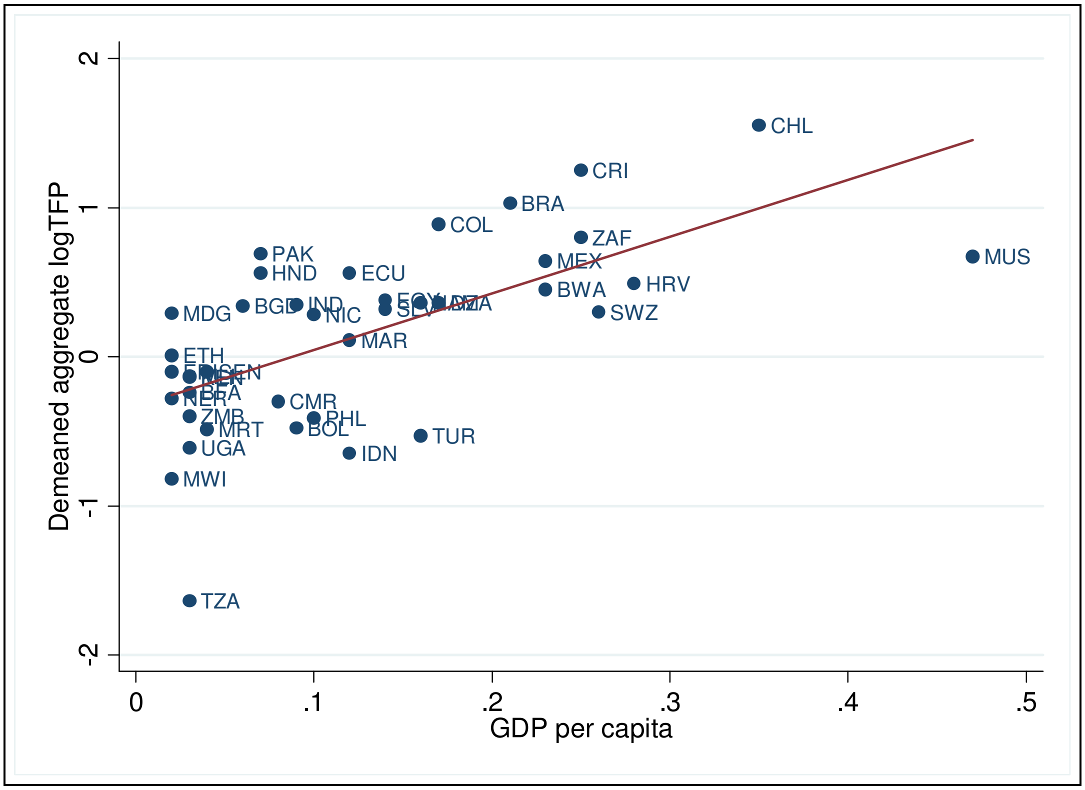
The underlying philosophy of this thesis is therefore essentially empirical. I understand that TFP as a residual, or that part of firms' output not explained by the inputs, is a black box. Moreover, in the literature there is no plenty consensus on what that box contains, or in the extreme if it can even be used as a measure of the true technical efficiency of a given firm or economy. The elusive quest for a robust theory of TFP is yet an unsettled issue in economics (Prescott, 1998), under the TFP concept might be a function of a variety of factors some of them as disparate as technology, macroeconomic uncertainty, crime, traffic, diseases, weather or even computer viruses if these make people working more inefficiently. Under a strictly empirical point of view, I attempt to bring some light on what TFP, understood as a black box, contains. The intuition behind this formulation is clear, we do not know what TFP is or what it contains, however the ICSs provide a practical underpinning on how is the business environment in which firms operate. Under the assumption that the IC is a key part of the technical efficiency of any country, industry, province or firm the objective is to use those hard data numbers to bring the abstruse TFP concept to the empirical world.

Accepting this definition, the "composition of TFP" in terms of the different investment climate variables can be estimated—following Escribano and Guasch (2005 and 2008)—and the relative weighs of each variable or group of variables on the aggregate productivity of the population analyzed empirically defined through standard econometric techniques. Once we have identified those IC effects with a statistically significant effect on explaining productivity we are able to aggregate the effects at the firm level, obtaining the IC effects on aggregate log-productivity. Aggregate log-productivity, according to the Olley and Pakes decomposition, can be in turn decomposed into average log-productivity and allocative efficiency; symmetrically, we can also decompose the aggregate effect of any IC variable on the un-weighted average effect and the allocation of resources effect.

In addition, once accepted the limitations imposed by the data and considering that we can apply the same "composition of TFP" proposed above to any country for which IC data is available, we can still isolate that part of TFP that is country by country homogeneous. We call this share of TFP "*demeaned productivity*", or that part of TFP strictly associated with the investment climate conditions.

The cross-plot between these *demeaned aggregate TFPs* and GDP per capita in more than 60 emerging and transition economies is shown in Figure 1. The results included in Figure 1 are the outcome of more than 5 years of research on the investment climate and partially they come from Escribano and Guasch (2005 and 2008), Escribano et al. (2008a and b) and other background papers done for several World Bank's Investment Climate Assessments. The larger and positive the demeaned aggregate productivity is the better investment climate conditions for doing business. The intuitive positive relation found between the demeaned aggregate TFP highlights and introduces the dramatic importance that creating and stimulating proper investment climate conditions have for economic growth and living standards.

Figure 1: Cross-plot between demeaned aggregate TFP and GDP per capita in 60 emerging and transition economies



Source: Author's elaboration based on the works of Escribano and Guasch (2005 and 2008) and Escribano et al (2008a and b) with ICSs.

Contributions of the dissertation

The main contributions of this dissertation are therefore of two types: empirical and theoretical. In one hand and from a strictly empirical point of view, the objective is to gather empirical evidence on what are the main determinants of TFP within the business environment in which firms operates. I come up to this objective from the analysis of the investment climate in more than 30 economies, with different levels of development among them. This way of approaching the analysis allowed us to compile rich and useful conclusions on where the main bottlenecks for economic success are (or could be) and therefore where we should put further efforts for interventionism and reform. More importantly, I consider the kind of analyses included in this dissertation as a first stage of a wider research path. In this sense, I do believe that the conclusions we reach should be considered as a preliminary and crude set of IC effects on firms' efficiency.

More empirical effort is needed in order to be able to discharge spurious effects and to place real causal relations between the IC and economic success in further stages of the research path.

The second types of contributions are therefore theoretical. More precisely, I consider important the way we (my coauthors and I) conduct the econometric strategy to identify the IC effects on TFP. What I think what is especially remarkable is the way we approach three different econometric issues: robustness, omitted variables and data quality issues. The underlying idea of getting robust results to different econometric specifications is always present throughout the entire dissertation. I am totally convinced that this is a key first step in order to discharge most spurious relations we may get and, in addition, what makes really rigorous an otherwise weaker analysis.

The omitted variables problem is a well-known issue in econometrics. In order to approach this question we take advantage of the incredibly large set of information included in the ICSs, which in most cases include information for almost 200 IC variables. We use this large set of information to approach the omitted variables problem, or more precisely to describe the idiosyncratic differences among the subjects of our study: the firms or establishments. Even in those cases in which we were only interested in estimating the effect of a single IC variable on TFP, we find of key importance to control for all the information available on the IC. This control approach is in turn related with the robustness issue. Besides of eliminating most spurious relations–as long as we estimate the conditional expectation of TFP on the widest set of information possible—, we believe that is what allowed us to get so robust results among so different specifications for the TFP equation.

Lastly, our concern has been also to give a proper treatment to data quality problems. Unfortunately, missing data, outliers and measurement errors are too frequently found in ICSs. The objective has been to use the control and robustness approaches to check the sensitivity of the results to different data quality and to know to what extent the results may be derived by these issues. In addition we have also derived an easy to implement mechanism to deal with missing data issues, which is presented in the third chapter of the dissertation.

Structure of the dissertation

The dissertation is structured in three independent chapters. The first one focuses in the evaluation of the investment climate in Spain, putting the exercise in international perspective by including other peer countries to the analysis—Germany, Ireland, Korea, Portugal and Greece.[1] In the comparison of the results we also include other sources of information like the EU-Klems database (2007), the Global Competitiveness Report (2009) and the Doing Business Report (2009). The interest of this analysis, besides studying the Spanish case in detail, is to extend the analysis of the investment climate, so far constrained to emerging and transition economies, to developed countries.

[1] The first chapter was written jointly with A. Escribano.

In addition, the Spanish case is of special interest given the international crisis context and the difference performance of Spain as compared with other European countries. The variety of questions that arises is enormous. For instance, what is the role that the quality of the investment climate plays in the different performance of Spain as compared with Germany or Ireland. The conclusions of the chapter are clear, the quality of the investment climate matters in explaining Spanish performance in presence of the crisis context. We found that the empirical results reaffirm most of those obtained from the GCR and DBR, and at the same time they suggest further economic bottlenecks of the Spanish economy in terms of productivity. Thus, the excessive bureaucracy, the poorly understood state paternalism and the regulatory burden creates unjustified barriers to private entrepreneurship and reduces the efficiency of the economy as a whole. We observe that these effects are more preeminent in Spain than in other countries, playing a clear negative role on aggregate productivity.

The second chapter (joint with A. Escribano and L. Guasch) changes the scenario of the analysis to developing countries. In this case our aim is to analyze and assess the role of the poor quality of the provision of physical infrastructures on the productivity of African firms. For that purpose, we apply the methodological procedure aforementioned to investment climate surveys of 26 African countries carried out in different years during the period 2002–6, making country-specific evaluations of the impact of investment climate (IC) quality on aggregate TFP, average TFP, and allocative efficiency. We divided countries into two blocks: high-income-growth and low-income-growth. The empirical results show that infrastructure quality has a low impact on TFP in countries of the first block and a high (negative) impact in countries of the second. We found heterogeneity in the individual infrastructure elements affecting countries from both blocks. Poor-quality electricity provision affects mainly poor countries, whereas problems dealing with customs while importing or exporting affects mainly faster-growing countries. Losses from transport interruptions affect mainly slower-growing countries. Water outages affect mainly slower-growing countries. There is also some heterogeneity among countries in the infrastructure determinants of the allocative efficiency of African firms.

The kind of questions discussed in the third chapter (joint with A. Escribano) of the dissertation has to do with methodological aspects of data quality of investment climate surveys. Concretely, we study the sensitivity of the results we get on the investment climate conditions, when we are forced to work with datasets with high proportions of missing data. Low data quality is a ubiquitous problem in econometrics and especially when we work with data from developing countries. In some cases the problem is so serious that it reduces the number of observations available to conduct a proper regression analysis to 0% of the original sampling frame.

Under the same methodology used in chapters I and II we propose a simple and easy to use method to deal with the problem of missing observations, which we call ICA method and which departs from the class of EM algorithms. We analyzed the mechanism generating missing values in order to know to what extent this may considered as missing are random or as opposite it was non-ignorable. We evaluate the performance of the ICA method under the two cases and we compare the performance of the ICA method with other more sophisticated imputation mechanisms, such as EM algorithms, or multiple imputation based on MCMC methods. We also include a comparison of the ICA method with a class of estimators *a la* Heckman.

The conclusions of the study are rather satisfactory. The performance of the ICA method is reasonable as long as we follow a robustness approach and we control for the IC variables related with the mechanism generating missing data (control for omitted variables). We found that the more parsimonious ICA method leads to similar and even in some cases to more consistent results than more sophisticated imputation mechanisms.

References

Dethier, J.J, M. Hirn and S. Straub (2008): “Explaining Enterprise Performance in Developing Countries with Business Climate Survey Data,” The*World Bank Policy Research Working Paper* #4792.

Escribano A. and J. L. Guasch (2005): “Assessing the Impact of the Investment Climate on Productivity using Firm Level Data: Methodology and the Cases of Guatemala, Honduras and Nicaragua,” *The World Bank Policy Research Working Paper # 3621*.

Escribano A. and J. L. Guasch (2008): “Robust Methodology for Investment Climate Assessment on Productivity: Application to Investment Climate Surveys from Central America,” *Universidad Carlos III de Madrid*, Working Paper # 08-19 (11).

Escribano, A., J. L. Guasch, M. de Orte and J. Pena. (2008a): “Investment Climate Assessment Based on Demean Olley and Pakes Decompositions: Methodology and Applications to Turkey’s Investment Climate Survey,” *Universidad Carlos III de Madrid*, Working Paper # 08-20 (12).

Escribano, A., J. L. Guasch, M. de Orte and J. Pena. (2008b): “Investment Climate and Firm’s Performance: Econometric and Applications to Turkey’s Investment Climate Survey,” *Universidad Carlos III de Madrid*, Working Paper # 08-20 (12).

Olley G. S. and A. Pakes (1996): “The Dynamics of Productivity in the Telecommunications Equipment Industry,” *Econometrica*, Vol. 64, 6, 1263-1297.

Prescott, E. C. (1998): "Needed: A Theory of Total Factor Productivity," *International Economic Review*, Department of Economics, University of Pennsylvania and Osaka University Institute of Social and Economic Research Association, vol. 39(3), pages 525-51, August

CHAPTER I

EVALUATION OF SPANISH INVESTMENT CLIMATE: EFFECTS ON FIRMS' PRODUCTIVITY

Evaluation of Spanish investment climate: Effects on firms' productivity*

An international comparison based on firm level data from 2005

by

Álvaro Escribano† and Jorge Pena‡

July, 2009

Abstract

In this paper we analyze how a number of economic variables such as physical infrastructures, finance, regulations, institutions, innovation or labor skills that are understood to form the investment climate available for doing business in Spain are associated with firms' productivity. For that purpose we apply robust microeconometric techniques from Escribano and Guasch (2005 and 2008) to data obtained from the Business Environment and Enterprise Survey (BEEPS) of the World Bank for Spain to first, identify those investment climate variables associated with sample variability of firm level productivities and, second, assess the relative importance of each investment climate factor in explaining Spanish aggregate productivity, following Escribano et al. (2008a). For comparison purposes, we apply the same investment climate assessment to data for Germany, Ireland, Korea, Portugal and Greece. In addition, we put the results obtained in international perspective by using other reports analyzing the competitiveness conditions in Spain and comparator countries, such as the Global Competitiveness Report (2008) (GCR), the Doing Business Report (2009) (DBR) and the EU-KLEMS Report (2008). We found that the empirical results from the econometric analysis reaffirm most of those obtained from the GCR and DBR and at the same time suggesting further economic bottlenecks for productivity of the Spanish economy.

Key words: Spain, Total Factor Productivity, Investment Climate, Competitiveness, Rankings of Global Competitiveness, Rankings of the Ease of Doing Business.

* We gratefully acknowledge the financial support received from the Government of the Community of Madrid for the realization of this work, as well as the cooperation of the World Bank granting proper access to the BEEPS database. All the errors are the sole responsibility of the authors.

† Telefonica-UC3M Chair of Economics of Telecommunications. Department of Economics, Universidad Carlos III de Madrid; alvaroe@eco.uc3m.es.

‡ Department of Economics, Universidad Carlos III de Madrid; jpizquie@eco.uc3m.es.

1. Introduction

By September-October 2008, when Spanish industrial production experienced negative rates of growth and especially when international financial liquidity problems become more acute, economic authorities began to see clear signals of a lasting economic recession. Also by that date, the European Union as well as most of more preeminent multilateral organizations (WB, IMF, OECD, etc) began to lower down their forecasts about global economic growth rates. In this context Spain suddenly abandoned its process of convergence with respect to EU income levels. The rapid economic convergence between Spain and EU-15 countries during the last 30 years had been direct consequence of the lower Spanish initial level of income, accompanied by rates of growth of GDP per capita over the average of the EU. Nonetheless, this positive growth differential vanished during 2009 as a consequence of the crisis context, turning Spain into a depressed economy with the handicap of being also the country with the highest unemployment rates within the EU. The important question that arises at this point is: what are the factors making Spain the most vulnerable economy of the EU in the presence of the global crisis context? Trying to explore this question by looking for the determinants of Spanish firms' competitiveness within the investment climate conditions is the scope of this paper.

As described and illustrated in Van Ark et al. (2007), "*the success story of Spanish economic growth has been shadowed by an extremely poor productivity performance*". In spite of Spain has been one of the EU members with the highest rates of growth of per capita income, it has also been the country with the lowest rates of growth of labor productivity. Figure 1 shows how Spanish GDP per capita was around 81% of the average of Euro area countries in 1997, and how it reached the 96% ten years later.[1] If we analyze the components of Figure 1 it becomes clear that the process of convergence observed in per capita income was driving exclusively by employment creation (dot line in the graph), and more importantly despite the stagnation of labor productivity or output per worker (dash-dot line in the graph).

[FIGURE 1 ABOUT HERE]

Figure 1 illustrates how Spain's strengths in the short term are its own debilities in the long, and how these have to be looked for in labor markets and in the lack of labor productivity growth. Besides the pure accumulation of production factors, the most important variable explaining output per worker is the total factor productivity (TFP). The TFP is a measure of the ability of an economy to keep increasing its output with a fixed level of inputs.[2] This point constitutes, along with the structure of labor markets, one of the weakest points of Spanish economy and a key structural barrier for firms' competitiveness. The EU-KLEMS database is

[1] Figure 1 illustrates the process of convergence between Spain and the rest of EU-15 members by plotting Spanish GDP per capita as a percentage of that of the average of EU-15 members. Likewise, the figure also decomposes GDP per capita into its two components: labor productivity and workforce participation. These two terms are expressed also in relative terms.

[2] Large differences in output per worker between rich and poor countries have been attributed to differences in Total Factor Productivity (TFP), see Caselli (2005), Hall and Jones (1999), and Klenow and Rodríguez-Clare (1997) among others.

very useful to illustrate the problem of low productivity of the Spanish economy, even in periods of low unemployment rate. Figure 2 plots the evolution of TFP in Spain along with those of several comparator countries. We observe that Spain, Italy and Portugal were the only countries that suffered negative inter-annual TFP growth rates between 2001 and 2005. More importantly, TFP in Spain decreased during all the period considered 1990-2005, with negative growth rates of -2% in 1996-2000 and -3.2% in 2001-2005, far away of the outstanding behavior of TFP in other countries such as Korea, Germany, France, UK or USA.

Further insight into this problem is provided by Figure 3, where TFP growth rates by activity sectors are shown. Only construction has experienced positive growth rates of TFP during the 2001-2005 period. In the other hand, TFP has had a especially bad performance in manufacturing and in hotels and restaurants, for which the estimated fall of TFP has been of -10.2 and -7.5% respectively.

[FIGURES 2 & 3 ABOUT HERE]

Once identified the debilities of the Spanish economy in terms of productivity, the study of the microeconomic determinants of this problem has deserved special attention in the economic literature during the last years. This body of literature has tried to identify the causes of the problem within a variety of determinants. For example, Delgado et al. (2002) analyze the role of exports on TFP. The impact of innovation on Spanish manufacturing firms' productivity has been studied by Huergo and Jaumandreu (2004) and the diffusion effect of technology by Ornaghi (2006). Likewise, productivity dynamics, firms' heterogeneity in productivity distributions and cyclical patterns in productivity have been analyzed in Fariñas and Ruano (2005), Lopez-García et al. (2007) and Escribano and Stucchi (2008). Other recurrent references in the literature are the works of Martin-Marcos and Jaumandreu (2004), Huergo and Jaumandreu (2004), Segura (2006), Jaumandreu and Doralzesky (2007), Dolado and Stucchi (2008) or Stucchi (2008).[3]

The enormous task of stimulating the productivity growth in Spain has therefore received great part of the attention of both academic and political debates in the last years. The aim of this work is to go in depth into this debate. Concretely, our aspiration is to empirically identify, assess and evaluate the main economic bottlenecks for productivity that the quality of the investment climate (IC) available for doing business has in Spain.

Nowadays, it is conceptually accepted and empirically demonstrated that the quality of the business environment in which firms operate has an enormous influence in the way they develop and evolve throughout its life cycle [4] The World Bank (2005) defines the investment

[3] The literature on productivity and its determinants is one of the most active fileds of research in economics. Two excellent and comprehensive surveys on this literature are the articles of Bartelsman et al. (2000) and Ahn (2001).

[4] The literature discussing the impact of the business climate on productivity and growth in developing countries has considerably grown during the last years thanks to an unprecedented data collection effort that has yielded a set of comparable enterprise surveys covering close to 70,000 firms from over 100 countries in all continents. As a result, a number of studies have started to analyze the impact of the business climate variables contained in these surveys on different dimensions of firm performance. The main results of this body of literature emphasize the key role of the quality of the investment climate over the productivity of firms, especially those located in developing countries.

climate as: i) the set of location-specific factors shaping the opportunities and incentives for firms to invest productively, create jobs and expand and ii) the institutional, policy and regulatory environment in which firms operate. The Investment Climate series of surveys (ICSs) are the instruments where all this set of information is compiled.[5] The surveys capture firms' experience in a range of areas related with the economic performance: financing, governance, corruption, crime, regulation, tax policy, labor relations, conflict resolution, infrastructures, supplies and marketing, quality, technology, and training among others.

Most of the empirical and theoretical effort done so far regarding investment climate has been almost exclusively focused in emerging and transition economies, given the ICSs have been initially designed to boost economic growth and development in these countries. However, the database used in the empirical exercise proposed in this paper comes from the Business Environment and Enterprise Survey (BEEPS in what follows), which, as opposite to the usual ICSs, compiles data on the quality of the investment climate of several developed countries — Spain, Germany, Ireland, Korea, Portugal and Greece—for 2005.[6] Exploiting this invaluable set of information allows us to first, approach the analysis of the investment climate from a different perspective, say that of developed countries; and second and more important, to contrast hypotheses over stylized facts and empirical regularities about TFP determinants in Spain unknown so far.

The methodological aspects of the work comprise the identification by means of robust microeconometric techniques of those factors of the investment climate with a statistically significant association with the TFP of the sample of firms included in the BEEPS.[7] For that purpose we follow Escribano and Guasch (2005 and 2008) and we base our analysis in the estimation of the IC effects on TFP under a simultaneous system of equations involving production and the investment climate. We work under the basic hypothesis that the IC available for doing business affects firms' optimum decisions of production through TFP, determining, at least in part, the heterogeneity in productivity distributions observed in the BEEPS. This simple hypothesis allows us to identify significant investment climate bottlenecks in Spain and compare them with those of other countries (Germany, Korea, Ireland, Portugal and Greece).[8]

By definition, investment climate is a broad concept comprising a wide range of economic processes and interactions. As a consequence, IC is understood in this work to include

---

Dethier et al. (2008) encompass the main results of the last years of investigation on this topic in a comprehensive survey. Other popular references within this body of literature are the works of Reinikka and Svensson (1999), Bastos and Nasir (2003), Veeramani and Goldar (2004), Eifert et al. (2005), Haltiwanger and Schweiger (2005), Frazer (2005), Glodberg et al (2005), Escribano and Guasch (2005 and 2008) Goedhuys et al. (2006), Fernandes and Pakes (2008), Escribano et al. (2008a and b) or Escribano, et al. (2009), among others.

[5] Concretely, the Investment Climate surveys were created within a new development strategy which aim is to put more emphasis in measuring the impact of intangible assets such as institutions or culture on growth and economic activity.

[6] To our knowledge, this is the first paper that tries to analyze the quality of the investment climate, understood as a broad concept, in a group of developed countries.

[7] It must be pointed out that inferring causal relations between TFP and IC is beyond of the scope of this work. Implementing those techniques that allow doing causal inference are unfeasible with data for a single cross-section of firms. For that reason we satisfy ourselves with getting useful statistical associations between TFP and the IC.

[8] As already pointed out, the BEEPS is intended to compile information for firms from Spain, Germany, Korea, Ireland, Portugal and Greece. Within each country the BEEPS includes information for several sectors and cities (regions); in the case of Spain it puts special emphasis in those firms from Madrid and Barcelona.

a wide range of variables too. Initially, we use more than 150 variables in the analysis, although after a selection process that goes from the general to the specific to avoid omitted variables biases we end up with a number of significant variables in between 20 and 30, depending on the country. Furthermore, we apply a robust econometric procedure that allow us to estimate IC effects on productivity that are robust to six different productivity measures, avoiding the problem of having results conditional on a single TFP measure.[9]

Once we identify those IC factors associated with firm level TFP, our objective is to analyze the contribution of them to each country aggregate TFP. For that purpose, we follow Escribano et al. (2008a) and we use the Olley and Pakes (1996) decomposition, which allows us to analyze the relative weight of each IC variable on aggregate productivity, which in turn can be decomposed into the relative weights on average productivity and on the processes of allocative efficiency. This second step of the analysis is important as it is what effectively allows us to rank IC bottlenecks and to make a proper country by country comparison. Moreover, by means of this kind of analysis we are able to test whether the effect of a given IC variable is transmitted to aggregate productivity through the level of efficiency of the representative firm (average TFP) or as opposite it is how the IC effect is distributed among firms with different share of sales (allocation effect) what dominates the transmission mechanism to aggregate TFP.

Likewise, in the second step of the analysis we are able to compute a synthetic index of investment climate conditions, introduced in Escribano et al. (2008a) which they call *demeaned aggregate TFP*. This index gives us a single and simple estimate of the effect of the investment climate on aggregate TFP and its components—average TFP and allocative efficiency—making possible to do international comparisons of investment climate conditions as a whole.

The structure of the paper is as follows. Section 2 tries to motivate why we think it is important to analyze the investment climate in Spain and to prepare the way forward for the econometric analysis of the next sections. For that, we describe from an international perspective what the current literature and especially other sources of information such as the Global Competitiveness Report (Porter and Schwab, 2008) and the Doing Business Report (World Bank, 2009) say about the deficiencies and strengths of the quality of physical and social infrastructures to run competitive business in Spain.

Section 3 provides details on the BEEPS and introduces a proper definition of the investment climate. Section 4 describes the econometric methodology used, stressing all the drawbacks we found in the estimation process. Section 4 present the empirical results obtained. Finally, section 5 concludes. The definition of the variables used in the analysis as well as the tables and figures with the results are included in two appendices at the end of the paper.

[9] Another interesting property of the Escribano and Guasch (2005 and 2008) approach is the way they approximate the fixed effects with information from the ICs, allowing to discharge most spurious correlations between the variables.

2 The quality of physical and social infrastructures in Spain at first glance

Is the investment climate available for doing business in Spain enhancing firms' competitiveness?[10] To give a proper answer to this question and to put the econometric analysis of the following sections in perspective, it is useful to start with a brief summary of what some recent reports have said about the strengths and debilities of the Spanish economy in terms of competitiveness. Basically, we rely in two sources of information: the *Global Competitiveness Report* from 2009 (GCR in what follows) and the *Doing Business Report* 2009 (DBR) prepared by the World Bank in 2008.

The recent GCR evaluate and rank 138 economies based on 12 basic pillars for competitiveness. The report summarizes its conclusions by elaborating an aggregate index of competitiveness, say the *Global Competitiveness Index*, based on the behavior of each economy in the 12 pillars. Figure 4 intuitively illustrates the importance of this index by showing the strong positive correlation that it has with GDP per capita in the sample of 138 countries included in the GCR. According to the index Spain (SPN in the chart) is still far away of other economies in terms of competitiveness, such as USA, Japan, UK, Netherland or Germany. In terms of rankings Spain ends up in the 29th position. The overall competitiveness index estimated for Spain is 4.72, equal for instance to those of China and Chile, over the level of Portugal, Greece and Italy but considerably lower than most of developed economies—U.S, Japan, U.K, Germany, Netherlands, etc.

[FIGURE 4 ABOUT HERE]

The GCR classifies the pillars in three main groups. The factors of the first group, say basic requirements, for Spain and other peer countries are in Table 1a. The overall performance of Spain in this group is reasonably good thanks in part to the first-class quality of its physical infrastructures. Spain ends up 27th out of 138 economies, with an overall score of 5.34, being 7 the maximum. Nonetheless, the performance of Spain in what refers to the quality of the institutional environment is considerably worse than in the rest of the factors of this group. Within the countries considered in Table 1a only the quality of the institutions of Italy and Greece received a score lower than that of Spain.

[TABLES 1a, 1b & 1c ABOUT HERE]

Continuing with the GCR, the next group of factors is presented in Table 1b. The GCR refers to them as "*Efficiency enhancers*", and are intended to be factors associated with competitiveness in early stages of economic development. Overall, Spain performs slightly better than Italy, Portugal and Greece. Within the positive factors, it is worth mentioning the outstanding role of Spain in what refers to the market size available for national companies, and

[10] A larger descriptive analysis of the Spanish IC from an international perspective is in Escribano et al. (2009b).

the reasonably good score received in the financial market sophistication factor. Nonetheless, there are also negative aspects. The efficiency of labor markets is especially problematic in Spain, finishing 96th with a score of 4.11, only higher than that of Italy, Greece and France.

Table 1c presents the last group of factors evaluated in the GCR, say the *innovation and sophistication factors*, those associated with competitiveness in the last stages of economic development. Spanish performance within these factors is similar to that of Portugal, Greece and Italy. Overall, Spain ranks 24th with a score of 4.89, one point lower than the first country of the list: Germany.

In summary, according to the GCR Spain is still far away of other economies such as US, UK or Germany in some key pillars for competitiveness, being Spanish performance more similar to that of Italy, Portugal and Greece. Within the positive aspects of Spanish competitiveness are the market size, the sophistication of the financial market and the first class infrastructures. In the negative side we find the inefficient institutional environment, the lack of innovative capacity, and especially the scarce flexibility of labor markets.[11]

The DBR elaborated annually by the World Bank in 181 economies is also very illustrative for the purposes of this section. The economies in this report are classified by the ease of doing business (absence of barriers to private entrepreneurship): e.g. how many permits or licenses are needed to open a business, how easy is to export or enforce contracts, etc. As expected, the conclusions of this report are fully consistent with those from the GCR. For example, opening a business in Spain takes in average 47 days and involves 10 administrative procedures, while if we want to open a business in India we will spend 30 days, 32 in Argentina and 40 in China. In the other side, it only takes 2 days to start an enterprise in Singapore and Canada. Other examples can also be very instructive, getting a construction permit in Spain requires a waiting lapse of 233 days, enforce a contract 515 and to register a property 18.

The importance of having a good institutional environment over efficiency and competitiveness becomes even clearer if we think in the effects this kind of barriers have over firms' incentive schemes and processes of *Schumpeterian* creative destruction.[12] Furthermore, the empirical association between the ease of creating, running and closing businesses efficiently and per capita income is patent in Figure 5. The cross-plot between the ease of doing business and GDP per capita suggests a clear positive relation between these two variables, suggesting ways to achieve the economic success. In global terms, Spain ends up in 49th place in the ease of doing business, just behind Portugal and only slightly better than Italy and Greece.

[FIGURE 5 ABOUT HERE]

[11] Conclusion supported by other studies, see for instance Stucchi and Dolado (2008) or Sanchez and Toharia (2001).

[12] Some references addressing the importance of reducing frictions (barriers) in the markets in order to boost entry/exit processes with considerable improvements of productivity levels are the works of Olley and Pakes (1996), Bartelsman, Haltiwanger and Scarpetta (2004), Foster, Haltiwanger and Krizan (2006), Bartelsman, Haltiwanger and Scarpetta (2006), Hsieh and Klenow (2006), Restuccia and Rogerson (2007) or Alfaro et al. (2007).

More information on the different determinants of the ease of doing business in Spain is provided by Figures 2a and 2b. Spain shows an especially bad performance in starting a business and employing workers, ranking respectively 140th and 160th out of 181 economies. The mechanisms to grant proper protection to investors and to pay taxes are neither satisfactory in Spain, ranking 88th and 84th respectively. Likewise, trading across borders, enforcing contracts or getting credit is easier in almost all the countries considered in Table 2 than in Spain. In general, Spain does not show an outstanding behavior in none of the aspects of the ranking of ease of doing business, but in closing a business, aspect in which Spain ranks 19th.

[TABLES 2a & 2b ABOUT HERE]

The empirical evidence is clear on the debilities of the Spanish economy and puts the way forward for further efforts and reforms (either structural or not). The inefficient institutional environment, the obtrusive regulation—especially in labor markets—, and the lack of innovative capacity currently constitutes serious bottlenecks for economic growth in Spain. The literature has documented the importance of these factors and the way they might affect competitiveness and productivity. At an empirical level, Lewis (2005) compiles a number of examples in which diverse institutional aspects such as excessive taxes, bad regulation, misunderstood protectionism or a too slow and rambling administration are translated to barriers to efficiency and competitiveness as well as to distortions to the markets, preventing economies to achieve their production possibilities frontier. At a more theoretical level but with the same message, the *institutions hypothesis* of the macro literature advocates that the quality of institutions is one of the main determinants of a country's level of development (Acemoglu, Johnson and Robinson, 2001; Rodrik, Subramanian and Trebbi, 2004; Easterly and Levine, 2003; Basu and Das, 2008; and Basu, 2008).[13]

Have these institutional factors a statistical association with firm level TFP in Spain? Or in other words, to what extent and how are they related with the poor performance of Spanish TFP? This is the question we will try to give answer in the following sections.

[13] As other examples of the role of intuitions on economic activity, Kasper (2002) shows that poorly understood "state paternalism" has usually created unjustified barriers to entrepreneurial activity, resulting in poor growth and a stifling environment. Kerr (2002) shows that a quagmire of regulation is a massive deterrent to investment and economic growth. McMillan (1988) argues that obtrusive government regulation before 1984 was the key issue in New Zealand's slide in the world per-capita income rankings. Hernando de Soto (2002) describes one key adverse effect of significant business regulation and weak property rights: with costly firm regulations, fewer firms choose to register and more become informal. Also, if there are high transaction costs involved in registering property, assets are less likely to be officially recorded, and therefore cannot be used as collateral to obtain loans, thereby becoming "dead" capital. Schimtz (2005) points to competitiveness and restrictive work practices as two key drivers of firms' productivity. Likewise, Erosa and Cabrillana (2007) point out that the ability to enforce contracts affects resource allocation across entrepreneurs of different productivities, and across industries with different needs for external financing.

3. Data and definition of the investment climate

3.1. Definition of data, sampling methodology and cleaning process

The information we have for the analysis comes from the *Business Environment and Enterprise Performance Survey* (BEEPS). Initially, this database was carried out by the World Bank and the European Bank for Reconstruction and Development (EBRD) in a number of low and middle income countries from east Europe and central Asia. For comparison purposes several high income countries were also included in the project during 2005-2006, Spain, Germany, Ireland, Portugal, Korea and Greece among them. It will be this last bunch of countries in which we will base our analysis.[14]

To ensure a proper representation of the universe of firms in each country a stratified random sampling was applied by sector, region and size.[15] In each country a sample frame as representative as possible were created from information compiled from national statistical institutes, chambers of commerce and industry, published information in industry registers and commercial directories. Minimum quotas of firms by industry, size, location, as well as exporters and foreign owned firms were established in the design of the survey. As a result the final sample of target firms is as representative as possible given these minimum quotas.[16]

The survey focuses in the next productive sectors: i) mining and quarrying; ii) manufacturing; iii) construction; iv) transport, storage and communication; v) wholesale trade; vi) real state; vii) hotels and restaurants; viii) other services. Table 3 summarizes the number of establishments surveyed by country and industry. Those observations included for the mining and quarrying sector represents less than 1% in each country. We exclude this sector from the analysis due to the low number of observations available for it.

[TABLE 3 ABOUT HERE]

Table 3 also includes the number of observations available for the analysis after excluding missing values and outliers. Data quality is a ubiquitous problem in the context of ICSs and BEEPS datasets, both outliers and missing values are too frequent in this kind of data. Table 3 shows that the number of observations available is considerably reduced in all the countries and industries when we exclude those observations with data quality problems in any variable. As a result, efficiency and representativeness in the estimation process is visibly affected and a proper treatment of this problem becomes a requisite. In the cleaning process we

[14] The survey was first undertaken on behalf of the EBRD and the World Bank in 1999 – 2000, when it was administered to approximately 4,000 enterprises in 26 countries of Central and Eastern Europe (including Turkey) and the Commonwealth of Independent States to assess the environment for private enterprise and business development. In the second round of the BEEPS, the survey instrument was administered to approximately 6,500 enterprises in 27 counties. In the third round ("BEEPS III") the BEEPS instrument was administered to approximately 9,500 enterprises in the 27 countries covered by the second round of the BEEPS.

[15] Size classification is as follows: small firms (less than 20 employees); medium firms (in between 20 and 100 employees); large firms more than 100 employees).

[16] This sampling methodology ensures that there is enough weight in the tales of the distribution of firms by industry, location and size. Further details on the sampling methodology can be obtained in European Bank for Reconstruction and Development (2005).

follow Escribano and Pena (2009) and we replace missing values and outliers by using a robust imputation methodology that departs from the EM type algorithms.[17] The imputation mechanism, referred to as *ICA Method* in Escribano and Pena (2009), basically estimates missing values as the expectation of each variable conditional on the information we have from industry, region and size each firm belongs to; in other words $E(k_{ij} \mid D_{ij}) = \delta' D_{ij}$, where k may be any variable with missing values in it and D is a vector of industry, region and size variables.[18] This cleaning process allows us to retrieve for the analysis a considerable number of observations, avoiding losing efficiency and representativeness in the estimation (see tables 3 and 4 for the number of observation available after the cleaning process). Further details on the possible biases in the estimation results caused by data quality are commented and evaluated in Escribano and Pena (2009).

Table 4 shows the number of observations in Spain by region, industry and size. The cities included in Spain are Madrid and Barcelona—in both cases including metropolitan area and surroundings—, large cities (in between 250,000 and 1,000,000 inhabitants), medium cities (in between 50,000 and 250,000 inhabitants) and small cities (less than 50,000 inhabitants). Likewise, the numbers of observations available before and after the cleaning process are detailed in the table, with equal implications than those described in the previous paragraph.

[TABLE 4 ABOUT HERE]

3.2. Definition of the investment climate

In order to properly describe the business environment available for doing business in each country a set of more than 300 variables were collected directly from managers' experience.[19] Out of this set of information 153 observations were susceptible to be used in the regression analysis. We include in Appendix I a detailed description of each IC variable together with the units of measurement used.

By simplicity we classify the IC variables in five broad groups. In the first group, says infrastructures, we include all the variables related with customs clearance, power and water supply, telecommunications (including phone connection and information technologies) and transportation. In the second group, bureaucracy and others, are included all the IC factors regarding tax rates, conflicts resolution, crime, bureaucracy, informalities, corruption and regulations. The next group is finance which contains factors related with governance, investments, informalities in payments of sales and purchases, access and cost of finance and

[17] Basically, there are two alternatives to deal with the problem of missing values. We can work with the observations available or complete case what in general is only acceptable if it implies losing less than 5% of the original number of observations and it does not introduce a omitted variables bias. The other alternative is to impute the missing values; this is the way we follow in this work.

[18] Escribano and Pena (2009) suggest testing the sensitivity of the results to a number of imputation mechanisms, multiple imputation, different specifications of the EM algorithm, amount of information embodied in the imputation mechanism as well as endogenous sampling schemes such as the Heckman model. The results for Spain are robust to all these mechanisms. Results are available upon from request.

[19] Although in the BEEPS the unit of observation is the establishment we will use both establishment and firm to refer to it throughout all the paper.

accountability (or auditing). The next set of IC variables is quality, innovation and labor skills; this group includes the quality certifications, technology usage, product and process innovation, research and development, quality or skills of the workforce, training, managers' experience and education. The last group –other control variables– is not properly a group of investment climate factors but a group of other firms' control characteristics, we classify into this group all those factors that we consider may have an important impact on the economic performance but not considered as a proper IC factor: exports and imports, age, FDI, number of competitors, etc.

Lastly, the BEEPS also includes all the information needed to construct firm level productivity (TFP) indices: sales (as measure of gross output), purchases of intermediate materials, net book value of capital stock, employment (total hours worked by year) and labor costs. A detailed description of them can be found in Appendix I. As no price indices are available at the firm level, the database only provides the nominal values of the variables. To try to solve this problem to the extent possible we use disaggregated price indices at the sector level specific for sales and intermediate materials from the EU-KLEMS database

4. Econometric strategy

In this section we introduce the econometric models used. Further details can be found in Escribano and Guasch (2005 and 2008), Escribano et al. (2008a) and Escribano and Pena (2009). Basically the methodology consists of two steps; the first one is the identification of those IC variables with a statistically significant association with firm level TFP, while the second is the evaluation of IC contributions in each country's aggregate TFP.

4.1 Identification of IC effects on productivity

In Escribano and Guasch (2005 and 2008) TFP of firm *i* in country *j* is determined as part of a structural system of equations as the following;

$$\log Y_{ij} = \alpha_L \log L_{ij} + \alpha_M \log M_{ij} + \alpha_K \log K_{ij} + \log TFP_{ij} \tag{1}$$

$$\log TFP_{ij} = \alpha_P + \alpha_W \log W_{ij} + \alpha_D' D_{ij} + a_{ij} + v_{ij} \tag{2}$$

$$a_{ij} = \alpha_{IC}' IC_{ij} + \varepsilon_{ij} \ . \tag{3}$$

Where in the production function of equation (1) TFP is total factor productivity, *Y* is sales, *M* is materials, *L* is employment as measured by total hours worked per year, and *K* denotes capital stock. Likewise, in equation (2) *W* is real wages per employee and *D* is a vector of sector/size/region dummies. In equation (3) *IC* is the vector of investment climate and other control variables. We are interested in getting a consistent estimator of the vector α_{IC}, which maps the relation between TFP and the IC. Likewise, the vector α_{IC} varies by country as the estimation is done for each country separately. The posterior evaluation and comparison of the quality of the investment climate conditions for doing business in each country will be conditional on a consistent estimation of α_{IC}.

The underlying philosophy of the econometric strategy is essentially empirical. We understand that TFP as a residual, or that part of firms' output not explained by the inputs, is a black box. Moreover, in the literature there is no plenty consensus on what that box contains, or in the extreme if it can even be used as a measure of the true technical efficiency of a given firm or economy. The elusive quest for a robust theory of TFP is yet an unsettled issue in economics (Prescott, 1998), under the TFP concept might be a function of a variety of factors some of them as disparate as technology, macroeconomic uncertainty, crime, traffic, diseases, weather or even computer viruses if these make people working more inefficiently. Under a strictly empirical point of view, we attempt to bring some light on what Spanish TFP, understood as a black box, contains.

Endogeneity issues

According to the system (1)-(3), the logarithm of firms' TFP can be expressed as the sum of real wages (as proxy for human capital), the average constant technical efficiency of each country (α_P), sector/region/size fixed affects and two random error terms, a_i and v_i. According to Escribano and Guasch (2005 and 2008) v_i is a usual *i.i.d* random term determined independently of the inputs of the production function, whereas a_i is that part of firm level TFP usually unobserved by the econometrician but known by firms' managers and therefore correlated with the inputs of the production function. This well-known endogeneity problem renders standard parametric estimators of input-output elasticities**,** such as OLS, inconsistent (Marschak and Andews, 1944; Griliches and Mairesse, 1995). Under the control approach proposed in the system (1) to (3), we assume that the usually unobserved part of TFP, the main source of endogeneity of the inputs, is now observed and approximated by the information on the IC from the BEEPS.

In other words, we attempt to model the expected value of productivity conditional on all the information we have from the BEEPS, what means using more than 150 IC variables, to give form to the *black box*. The intuition behind this formulation is clear: we do not know what TFP is or what it contains, however the BEEPS provide a practical underpinning on how is the business environment in which firms operate. Under the assumption that the IC is a key part of the technical efficiency of any country, industry, province or firm the objective is to use those hard data numbers to bring the abstruse TFP concept to the empirical world.

In addition, as long as we model the conditional expectation of TFP on this large set of information we can get—under certain regularity conditions—a consistent estimator of $\boldsymbol{\alpha_{IC}}$ for each country, eliminating most spurious correlations between TFP and IC caused by the bias introduced by the omitted variables contained in a_i. Nonetheless, this exogeneity condition does not hold for all IC variables. Many IC variables are likely to be determined simultaneously along with any TFP measure. The traditional instrumental variable (IV) approach is difficult to implement, given that we only have information for one year. Therefore, we cannot use the natural instruments for the variables, such as those provided by their own lags, and in addition it is difficult to find good instruments from the list of IC variables. As a simple alternative correction for the endogeneity of the IC variables, we use the region-industry averages[20] of plant-level IC variables, which is a common solution in panel data studies at firm level.[21]

Productivity measurement

Particularly important in the methodology implemented in the identification step is the way productivity measurement is approached. Escribano and Guasch (2005 and 2008) argue that there is not a single salient measure of productivity and therefore any empirical evaluation of the IC effects on productivity will intimately depend on the particular way productivity is measured. By looking for robust results for several productivity measures we reduce the degree of uncertainty of IC estimates that would arise otherwise when we rely in only a single estimator.[22]

Therefore, we want the IC elasticities and semi-elasticities to be robust to: i) different technologies in equation (1) (Cobb-Douglas and Translog); ii) different assumptions on market conditions—constant returns to scale, perfectly competitive input markets—leading to parametric (single step estimation of the system (1)-(3) by OLS) and non-parametric (Solow, 1957 and Hall, 1990) estimation of productivity; and iii) different input marginal productivities by sector (restricted input-output elasticities for each country, unrestricted by industry input-output elasticities within countries).[23]

From the idea of robustness presented above we get two basic projections of productivity onto the IC variables. First, from the single step estimation approach with a Cobb-Douglas production function we have

$$\log Y_{ij} = \alpha_L \log L_{ij} + \alpha_M \log M_{ij} + \alpha_K \log K_{ij} + \alpha_P + \alpha_W \log W_{ij} + \alpha'_D D_{ij} + \alpha'_{IC} IC_{ij} + \varepsilon_{ij} + v_{ij} \qquad (4)$$

that is, we plug (2) and (3) into (1).[24] Second, from the non-parametric estimation of the two steps estimation approach we have the next productivity projection

$$\log TFP_{ij} = \alpha_P + \alpha_W \log W_{ij} + \alpha'_D D_{ij} + \alpha'_{IC} IC_{ij} + \varepsilon_{ij} + v_{ij} \qquad (5)$$

where productivity in this case comes from the traditional decomposition of growth proposed in Solow (1957) and revised by Hall (1990) and therefore the input-output elasticities are assumed to be equal to each input's cost shares and therefore inputs markets are assumed to be perfectly

[20] Because of the low number of available regions in most of the countries, we had to use the industry-region-size variables instead of the region-industry averages. For the creation of cells a minimum number of firms are imposed—there must be at least 15 to 20 firms in each industry-region-size cell to create the average, otherwise we apply the region-industry averages. If the problem persists, we apply the industry-size or the region-size average.

[21] This two-step estimation approach is a simplified version of an instrumental variable estimator. Using industry-region-size averages also mitigates the effect of having certain missing individual IC observations at the plant level, which—as mentioned in section 3—represents one of the most important difficulties using ICS or BEEPS data.

[22] The justification of this procedure can be found in the sensitivity analysis approach of econometrics, see for instance Magnus and Vasnev (2004). For a longer discussion on productivity measurement see for instance Solow, 1957; Jorgenson, Gollop and Fraumeni, 1987; Diewert and Nakamura, 2002; Jorgenson, 2003; or van Biesebroek, 2007.

[23] In addition, Escribano and Guasch (2008) show that results are robust to further approaches such as Levinsohn and Petrin (2003) and Ackerberg et al. (2009) procedures. These methods have been also applied for the case of Spain. The results are robust to all these methods and are all available upon from request.

[24] Equation (4) corresponds to a Cobb-Douglas production function; the case of a Translog is symmetrical.

competitive and there are constant returns to scale in production processes. From either (4) or (5) and under the different assumptions proposed, we can estimate by OLS 6 sets of IC elasticities and semi-elasticities with respect to productivity, which we want to have the same sign and vary within a reasonable range of values across specifications.

Production function versus sales generating function

The role of prices in the system (1)-(3) deserves special attention. As our dependent variable is sales, rather than units of physical output, it reflects prices. In fact, according to the current literature, the term *sales generating function* seems more appropriate for equation (1) rather than production function, as in the work of Olley and Pakes (1996). If prices are not identical across firms, what seems to be a high productive plant may be just an establishment that is charging high prices, what in turn may be consequence of either market power (non zero mark-ups) or differences in quality of final goods. While with homogeneous products high productivity could be a reflection of high prices, or in other words a reflection of market power (Melitz, 2000: Bernard, et al., 2003), under heterogeneous or differentiated products high prices could be consequence of higher quality, what could be translated to overmeasured productivities as some plants would be able to produce higher quality—and price— products with the same amount of output (Levinsohn and Melitz, 2002; de Loecker, 2007; Katayama, et al., 2006; Gorodnichenko, 2007). These points are especially important in presence of market power.

Addressing these issues is not a straightforward task with the data at hand though. A more comprehensive analysis would need information on plant level input prices to incorporate the demand side of the model. As long as this data is not available a plausible solution is to test the assumptions of market power or differentiated products with a careful reduced form analysis of the number of firms or the different trade policy implemented in a given industry. For instance a low number of industries and/or restrictive trade policies may be a cause of concern,[25] this is the approach followed for instance in Escribano, de Orte and Pena (2009).

Selection of the final set of significant IC variables

The econometric methodology applied for the selection of the final set of significant IC variables goes from the general to the specific. The otherwise omitted variables problem that we encounter— starting from a too simple model—generates biased and inconsistent parameter estimates. We start the selection of variables with a wide set compounded by up to 153 IC variables (see the appendix on definition of variables). We avoid using simultaneously variables that provide the same information and are likely to be correlated, mitigating the problem of multicollinearity that could arise otherwise. We then start removing from the regressions the less significant variables one by one, until we obtain the final set of variables, all significant in at least one of the regressions and with parameters varying within a reasonable range of values. Once we have selected a preliminary model we test for omitted IC variables.

Eventually, we end up with 23 significant IC variables in Spain, distributed by groups as follows: 4 in infrastructures, 4 in the group of bureaucracy and others, 8 in finance, 3 in quality,

[25] We are indebted to James Levinsohn and Ariel Pakes for useful comments on this point.

innovation and labor skills and 4 in the group of other control variables. In Germany we estimate 14 significant IC variables, 18 in Greece, 23 in Portugal and 15 in Ireland and Korea.[26] We estimate the IC coefficients by OLS, controlling for heteroskedasticity with the usual White robust standard errors and for correlation within clusters introduced by the sampling structure with cluster standard errors by industry and region. The IC variables can enter the regression in two forms, as industry-region average or with the missing values imputed according to the method explained in section 3.

The robust coefficients of the IC variables with respect to TFP, along with their level of significance, are detailed in tables 5 to 10 of the appendix on tables and figures included at the end of the paper. Indications on the form the variables enter the regression—industry-region average or missing values imputed—are also included. The results are in all the cases robust across the six specifications proposed for TFP equation. The IC elasticities and semi-elasticities estimated never change the signs and the numerical values obtained vary within a very reasonable range. The main reason for getting these robust results under so different assumptions on the production process is to control for IC variables from all the blocks (infrastructures; bureaucracy and others; finance; quality, innovation and labor skills; and other control variables) to avoid having omitted variables biases (Escribano and Guasch, 2005 and 2008). Finally, with the exception of the variable 'Days to clear customs in exports' in Spain and Greece, we do not use the industry-region average of any variable given that Housman tests do not support the hypotheses of endogeneity for any variable.

[TABLES 5 TO 10 ABOUT HERE]

4.2 Evaluation of IC effects on aggregate productivity

Once we have identified those IC effects with a statistically significant effect on explaining variability of productivity we are able to aggregate the effects at the firm level, obtaining the IC effects on aggregate log-productivity of each country. Aggregate log-productivity, according to the Olley and Pakes decomposition, can be in turn decomposed into average log-productivity and allocative efficiency. Symmetrically, we can also decompose the aggregate effect of any IC variable into the un-weighted average effect and the allocation of resources effect.

The Olley and Pakes (1996) decomposition of aggregate log-TFP of country j is thus given by

$$\log TFP_j = \overline{\log TFP}_j + N_j \operatorname{cov}(s_{ij}, \log TFP_{ij}) \qquad (6)$$

where $\log TFP_j$ is aggregate TFP of country j; or in other words the weighted average of firm level productivities, with weights given by each firms' shares of sales. $\overline{\log TFP}_j$ is the un-weighted average and the last term of the decomposition is the allocative efficiency. This last term measures whether high productive firms are also those with the largest market shares,

[26] By means of simplicity we do not include the input-output elasticities in tables 5 to 10. The results are available upon from request.

therefore using most of the resources of the economy, in which case the term is positive and large. If the allocative efficiency is negative it implies that TFP and market shares are negatively correlated, and as a consequence aggregate TFP is reduced as the resources of the economy are being used by low productive firms.

The useful additive property of equation (2) in logarithms, allow us to obtain an exact closed form solution of the decomposition of aggregate log productivity. Following Escribano et al. (2009a) we can express aggregate log productivity as a weighted sum of the average values of the IC, dummy D variables, the intercept, the productivity residuals from (2) and (3); and, the sum of the covariances between the share of sales and investment climate variables IC, dummies D and productivity residuals

$$\begin{aligned} \log TFP_j &= \hat{\alpha}_P + \hat{\alpha}_W \overline{\log W}_j + \hat{\alpha}_D \overline{D}_j + \hat{\alpha}_{IC} \overline{IC}_j + \overline{\hat{\varepsilon}}_j + \overline{\hat{v}}_j + N_j \hat{\alpha}_W \,\hat{\text{cov}}(s_{ij}, \log W_{ij}) \\ &+ N_j \hat{\alpha}_D \,\hat{\text{cov}}(s_{ij}, D_j) + N_j \hat{\alpha}_{IC} \,\hat{\text{cov}}(s_{ij}, IC_{ij}) + N_j \,\hat{\text{cov}}(s_{ij}, \hat{\varepsilon}_{ij}) + N_j \,\hat{\text{cov}}(s_{ij}, \hat{v}_{ij}) \end{aligned} \qquad (7)$$

where by simplicity IC and D represent now scalars rather than vectors and the set of estimated parameters used comes from the two-step TFP estimation, having the restricted Solow's residual as dependent variable in (4).[27]

From equation (7) each IC variable may affect aggregate log TFP through both the un-weighted average and the covariance with respect to the share of sales. This complements the information provided by the marginal effects (IC elasticities). Suppose that an IC variable with a low impact, in terms of marginal effects (elasticities), affects most of the firms in a given country. In that case the impact of such an IC variable in terms of average productivity could increase significantly because it would be suffered by most of the firms in the population.

Similarly, a variable with a negative marginal effect on average productivity may have either a positive or a negative effect on allocative efficiency. If the covariance of an anti-productive IC variable and the market share is positive, then we say that a large proportion of output is in hands of establishments with high levels of a variable that harms TFP, and consequently aggregate TFP decreases. In contrast, a negative covariance means that those establishments with the highest levels of the IC variable are those with the lowest market shares, and, therefore, the negative effect of the IC variable on average productivity is somehow compensated through the effect on the reallocation of resources among firms.

The importance of using the share of sales as weighting variable in order to denote the relative importance of each firm in the sample is simply that high share of sales firms are those using the largest proportions of resources of the economy. As a consequence, negative or positive IC effects can be either amplified or mitigated depending on which firms are affected by them. When the respective contributions to aggregate and average TFP are almost equal we say that there are no significant differences in the impact of the IC by types of firms.

At this point, one problem remains unresolved: how can we compare the investment climate conditions country by country? Traditionally, in the literature international TFP comparisons are restricted to only some kind of macro data with the optimal properties to do this

[27] Note that the set of IC elasticities and semi-elasticities used is not so important proved that we have been able to get robust results to different assumptions and estimation processes.

type of exercises (see Caves, Christensen and Diewert, 1982). The low quality of the micro-level data collected in developing countries along with differences in the sampling methodology across IC surveys make it unfeasible to do this kind of comparisons. Apart from these data quality problems, there still remains the problem of comparing apples and oranges (Bartelsman, Haltiwanger and Scarpetta, 2006), as we do not know what it is contained in the TFP measure in different countries.

Once accepted the limitations imposed by the data and considering that we can apply the same *composition of TFP* proposed in (7) to any country for which IC data is available, we can still isolate that part of TFP that is country by country homogeneous. We call this share of TFP *demeaned productivity* (or TFP). According to our formulation of TFP at the firm level, *demeaned productivity* is simply firm level TFP minus the constant effect, residual and size/region/sector fixed effects, or in other words that part of TFP strictly associated with the investment climate conditions. This concept of *demeaned productivity* allow us to avoid the problem of comparing apples and oranges in international productivity comparisons, as all the measurement errors are contained in the constant and residual and size/region/sector fixed effects, and therefore what remains is the pure effect of the investment climate. We argue that demeaned productivities are cross-country comparable because the productivity equation used comprises the same set of IC variables for all the countries considered and complementarily we use the same methodology for the identification and assessment of IC effects.

Therefore, demeaned aggregate TFP of country *j* obtained as the result of aggregating firm level demeaned TFP estimates is given by

$$Demeaned \log TFP_j = \hat{\alpha}_{IC}\overline{IC}_j + N_j \hat{\alpha}_{IC}\,\hat{\text{c}}\text{ov}(s_{ij}, IC_{ij}) \qquad (8)$$

The interpretation of the demeaned aggregate TFP is straightforward. We interpret it as a synthetic index of the investment climate conditions available for doing business in a given economy (country, region, industry, and so). It allows us to compare IC conditions among countries as well as to compare the contribution or relative importance of each IC variable (or groups of IC variables) in the demeaned aggregate TFP country by country. Furthermore, we are able to know whether the contribution of a given IC variable comes from the effect on average TFP or as opposite through the effect on the allocative efficiency, what may have important consequences in terms of economic policy and investment climate reforms.

5. Empirical results

Before we start introducing the main results of the econometric analysis it is convenient to clarify some points on the estimation process. Among them especially important is the interpretation of the results, we cannot be sure that the estimated IC effects on TFP can be interpreted as causal relations. Therefore, we have to satisfy ourselves with inferring associations and empirical regularities between IC and TFP. In either case we consider the results very useful in the task of identifying environmental restrictions to competitiveness and efficiency.

As already pointed out, the results obtained are robust to different specifications of the production process. All of them are significant in at least one of the equations, never change the direction of the corresponding effect and more important they vary within a very reasonable range of values. Given the robustness of the results, we rely in only one set of IC coefficients for the posterior IC evaluation in aggregate TFP, say those from the two step procedure with the Solow residual as TFP measure computed with input-output elasticities restricted to be equal within each country.

Likewise, the R^2 is reasonably high in all the countries. Concretely, R^2 in Spain varies in between 0.46 and 0.61, depending on the specification used. In other countries it is even higher.

Furthermore, the average wage per worker (proxy for human capital) has a significantly high association with TFP in all the countries. Nonetheless, in the remaining of this section we will rely only in the results obtained for the IC, as analyzing the role of real wages on TFP is beyond of the scope of this paper and we only include it in the regressions as a control variable

In what follows we present the results of the evaluation of IC on demeaned aggregate TFP. In particular, we concentrate in the IC contributions to demeaned aggregate and average TFPs. The difference between them is the contribution to the allocative efficiency, so although by means of simplicity we do not include it in any of the tables and figures, it will be implicitly present in all the results.

The remaining of this section is structured as follows. First we introduce the results obtained by groups of IC variables, preparing the way forward for the analysis of the effects of individual IC variables. We finish by introducing the estimated demeaned aggregate TFP. We always focus in the Spanish case in comparison with Germany, Greece, Portugal, Ireland and Korea.

5.1 IC absolute contributions to demeaned aggregate TFP by blocks of variables

In this section we introduce the results of equation (8). We add up the effects by groups of IC variables in absolute terms to avoid compensating positive and negative effects. All the results are in Figure 6.

[FIGURE 6 ABOUT HERE]

From Figure 6 we observe that the contribution of the IC group of infrastructures is noticeably larger than in the rest of the countries. The relative importance of this group in explaining the demeaned aggregate TFP in Spain is 30%, whereas in the rest of the countries it is around 0%-2%, but in Portugal where it reaches 18%. The group of bureaucracy contributes with less than 10% of demeaned aggregate TFP, almost equal than the contribution of this group in Germany, Korea and Portugal and lower than in Ireland and Greece.

The block of finance variables contributes with almost 20% of Spanish demeaned aggregate TFP, the third in order of relative importance. Finance is also important in the rest of the countries analyzed with contributions ranging from 13% in Germany to 34% in Greece.

Quality, innovation and labor skills group also has a significantly high relative importance. The contribution of this group in Spain is more than 12%, almost equal than in Germany, Korea and Ireland and slightly higher than in Greece and Portugal. Finally, the group of other control variables has the higher relative importance in all the countries, contributing with more than 30% to demeaned aggregate TFP in Spain and more than 66% in Germany.

In the other hand, it is worth mentioning the differences between aggregate and average TFP contributions. As we have also mentioned, the differences between them are due to the correlation between firm level demeaned TFP and the share of sales. For instance in Spain there is no significant difference between the contribution of the block of infrastructure to aggregate and average TFP, so the effect of the infrastructures is evenly distributed among the sample of firms. As opposite, we do observe ostensible differences in the impact of finance variables, for which the contribution to the aggregate TFP is lower than that to the average term. This can imply either that positive IC effects are reduced as they are concentrated in low share of sales firms or that the negative effects are concentrated in firms with high share of sales. This effect is also observed in Portugal, although it is less intense than in Spain.

5.2 Individual IC contributions to demeaned aggregate and average TFP

We now introduce the results of equation (8) by individual IC variables. The entire set of results for Spain and peer countries is in Table 11.

[TABLE 11 ABOUT HERE]

As already pointed out, the contribution of the infrastructures group is relatively more important in Spain than in the remaining countries. The main results by individual IC variables are as follows:

- In spite of the results show a considerably high effect of infrastructures on aggregate TFP, the main contributor within this group is a pro-productive variable, say ICT usage.[28] That is, those firms using e-mail with clients and suppliers are associated with higher productivity levels. Moreover, the contribution of this variable to (demeaned) aggregate TFP is higher than that to the average, so this positive effect is proportionally associated with firms with high share of sales. In the other hand this variable is significant only in Greece.
- In the negative side, the average number of days spent with red tape when exporting contributes in Spain with -16.2%. This variable only affects TFP in Greece, although with a lower contribution. In addition, a higher contribution to aggregate TFP indicates that the effect is concentrated in high share of sales firms.

[28] The literature related with the ICT problem or the so called "Solow's Paradox" in Spain has frequently focused the lack of positive externalities and productivity gains from the use of these technologies. Nonetheless, during the last years the trends in the literature are others. The benefits of the use of ICT can be observed in the statistics as new ways of organization of work in order to get the maximum profit of the use of these technologies appear (see Mas, 2004 and Mas and Quesada, 2004).

- Lastly, the variable 'average duration of phone outages' also affects aggregate TFP, although with a lower relative contribution than other variables of this group (-2.5%).

Within the block of bureaucracy and others, whose relative absolute contribution is 8.6%, the results we want to address are the following:

- The variable '*Dummy firm would reduce its current level of employees*' (firms that would reduce staff if labor regulation were more flexible) contributes in Spain with -8% of aggregate TFP. This variable is only significant in Spain.
- National regulation, concretely the fact that firms' decisions are affected by it, contributes negatively in Spain with -5% of (demeaned) aggregate TFP. We only observe this in Spain.
- The number of cases of commercial or civil arbitrage as plaintiff contributes negatively in Spain, Greece, Ireland and Korea. Among them the lowest contribution is observed in Spain with -4%. It is worth mentioning that the contribution of this variable is almost equal in all the countries, around -3%, and what is making more negative the contribution to the aggregate TFP of Greece, Ireland and Korea is the fact that the negative effect is more concentrated in firms with high share of sales in these countries. In Spain we do not observe significant difference of this variable by the type of firms affected.
- Finally, the percentage of unreported sales to tax authorities contributes with -1.5% in Spain, similar to the effect estimated for Ireland. In Greece and Korea what is negatively associated with aggregate TFP is the percentage of workforce unreported, another measure of informality. It is important to note that the effect of the informalities tends to be concentrated in firms with low share of sales, what mitigates the negative effect of this variable—in Spain the effect is reduced from -2.3% to -1.6%, and in Ireland from -6.3% to -1.7%.

Finance is the group of variables with the highest relative contribution in Spain, concretely 19.7%, slightly lower than in other countries. Furthermore, in all the countries but Germany, the contribution to average TFP of finance is reduced when we move from average to aggregate TFP:

- Belonging to a trade chamber contributes positively in Spain and Ireland. While in Ireland there are no differences between the contributions to aggregate and average TFP, in Spain it can be observed a small bias in the effect toward firms with the largest market shares.
- The percentage of working capital financed by internal funds has a negative contribution in Spain, Greece and Korea. The contribution of this variable in Spain is -9.2%. We do not observe substantial differences in the average and aggregate TFP effects.
- Having a loan with collateral has a positive association with TFP and more importantly it affects mainly firms with small market share. The contribution to average TFP is reduced from 12.6% to 7.4% in the case of the aggregate. This variable is also significant in explaining sample TFP differences in Portugal, but with a contribution several times lower.

- The interest rate of the loan contributes negatively with -8.9% of aggregate TFP in Spain. This variable also contributes negatively in Portugal with -24%. In both countries it seems that the effect tends to be brought together in firms with low market shares, so the negative effect on the average is somewhat mitigated.
- Another variable with a significant and positive contribution (4.4%) is the percentage of purchases paid for before delivery. This variable is also significant in Ireland with a contribution rather similar, 3.6%.

Regarding the group of quality, innovation and labor skills, whose absolute contribution to aggregate TFP is 12.3%—and 6% to average TFP—the main results are the next:

- The effect of R&D investment is very important, as its positive contribution of 22.7% to aggregate TFP indicates. The effect is concentrated among firms with large share of sales. This variable is also significant in Germany and Korea, but with a relative contribution somewhat lower than in Spain.
- The percentage of staff with university education is the other important contributor within this group. This variable also contributes positively in Germany, but while the effect in aggregate terms in Spain is 3.3%, in Germany it is 1.7%.

In what refers to the last group, other control variables, we address the next results:

- It can be observed a positive relative contribution of the percentage of sales exported to aggregate TFP of 34%. Several observations are worth mentioning with respect to this variable. It is not significant in any other country, although the percentage of imports and the FDI contributes positively in Korea and Portugal. The effect is clearly biased towards firms with high share of sales—the contribution to average TFP is 11.4%, while the effect to the aggregate is amplified to 34%.
- Finally there are two variables within this group making reference to the organizational structure of the firm with a statistically positive association with TFP. These variables are the firms constituted as incorporated companies and also those firms that have acquired other companies. Their contributions are 15.4% and 17.5%.

5.3 Summary of overall investment climate conditions

Figure 7 shows the Olley and Pakes decomposition of demeaned aggregate TFP in Spain and peer countries. In general terms, what this figure shows is the overall effect of the investment climate conditions on TFP when we aggregate pro-productive and anti-productive IC effects. In other words, the higher the demeaned aggregate TFP is the better investment climate conditions faced by the firms. As a consequence the capacity of the firms to expand the production possibilities frontier beyond the pure accumulation of inputs is amplified. In turn, a positive IC effect can be due either to the effect on the average firms or to the allocative efficiency. [29]

[FIGURE 7 ABOUT HERE]

[29] With this we do not refer to how market shares are reassigned among firms as we regress TFP on IC and not market shares on IC.

The effect of the investment climate is positive in all the countries. In Germany the effect is the largest, followed by Ireland, Spain, Korea, Greece and Portugal. Nonetheless, the results show that demeaned aggregate TFPs are rather similar among countries, in between 0.1 and 0.36.

In Spain, out of the 0.33 points of the demeaned aggregate TFP, almost one half is due to the allocation effect. That is, in Spain the positive effects of the investment climate tend to be concentrated in firms with high share of sales and the negative in firms with small market shares. This is a positive characteristic of the investment climate conditions in Spain, as in spite of the effect in the average firm is not too high, when we take into account distributive aspects, the pro-productive aspects of the IC are amplified, while the anti-productive are mitigated.

In Greece and Ireland the effects are not different in what respect to the effect by market shares, however, the effects over the average or representative firm are somewhat better than in Spain. In Germany the effect is positively high in both components. Lastly, the allocation effect in Portugal is negative, constraining the positive effect of the IC on average TFP.

6. Conclusions

Spanish convergence to UE income level has been the result of the incorporation of low skilled workers to the productive structure, and no thank to but in spite of TFP evolution. Both determinants of growth are pro-cyclical, making Spain more vulnerable to the international crisis.

I order to boost economic growth in Spain stimulating firms' competitiveness becomes a requisite. According to the *Global Competitveness Report* (2009) the main factors constraining firms' competitiveness in Spain are the institutional environment, the lack of innovative potential and the scarce flexibility of labor markets. The Doing Business Report (2009) is also very instructive in identifying barriers to competitiveness. This report signals as key economic bottlenecks to do business in Spain efficiently the current mechanisms available to start a business, employ workers, grant proper protection to investors, pay taxes, trade across borders and enforce contracts.

In order to go in depth into the investigation of the causes of the poor evolution of Spainsh competitiveness, we have explored by means of robust microeconometric techniques the relationship between firm level TFP and the investment climate available for doing business in Spain. The analysis is complemented with an international comparison of IC conditions in Spain with those available in Germany, Ireland, Korea, Greece and Portugal. The empirical results obtained in this work with the BEEPS database are reasonable because i) all the effects and signs of the IC on firm level TFP are as expected; ii) they are also robust to six different productivity measures; iii) the results are able to explain up to 80% of the variance of TFP with cross-sectional data from 2005, and iv) the human capital (measured as wage per worker) is always significant in explaining TFP in all the countries analyzed. The international comparison of IC conditions is done through the concept of *demeaned TFP* to avoid the conceptual problem of comparing '*apples and oranges*'.

We found that the empirical results reaffirm most of those obtained from the GCR and DBR and at the same time suggesting further economic bottlenecks of the Spanish economy in terms of productivity. Thus, an excessive bureaucracy, poorly understood state paternalism or the regulatory burden create unjustified barriers to private entrepreneurship and reduces the efficiency of the economy as a whole. We observe that these effects are more preeminent in Spain than in other countries, playing a clear negative role on aggregate productivity. The main findings of the report can be summarized as follows:

i) *Infrastructures*. According to the econometric analysis the quality of the physical infrastructures in Spain is satisfactory. They contribute with approximately 30% of demeaned aggregate TFP. The largest effect on TFP comes from the use of information technologies. Nonetheless, the waiting lapse wasted in customs to export constitutes a relatively serious bottleneck as it contributes negatively with -16% of aggregate TFP, and this effect is only a problem in Greece (in line with what the DBR says). Finally, it seems that the average duration of the phone outages is also constraining average TFP to some extent, the contribution of this variable is negative (-2.5%).

ii) *Bureaucracy and others*. The excessive bureaucracy and regulation constitutes a massive deterrent to private entrepreneurship in Spain. This is reflected in the econometric analysis by three IC variables (besides the time dealing with customs in exports which also reflects bureaucracy): the proportion of firms that would reduce their staff if labor regulation would be less restrictive (-8%) contribution to demeaned aggregate TFP), the restrictive national regulation (-5%) and the number of cases of commercial or civil arbitrage as plaintiff (-4%). It is important to note that the first two are only significant in Spain and they are fully consistent with the conclusions of the GCR and DBR. In the other side, in what refers to the informal economy, the percentage of unreported sales to taxes contributes with -1.5% to demeaned aggregate TFP, possibly related with difficulties to pay taxes, as signaled in the DBR.

iii) *Finance*. Spanish financial sector receives a good evaluation in the GCR, as well as in other international forums. It is the second group of variables contributing to the demeaned aggregate TFP just after infrastructures. In general terms, financing tend to affect more firms with low market share, what makes the contribution of this group to the demeaned average TFP larger than the contribution to the aggregate TFP. Three variables affects positively and two negatively. The positive effects come from belonging to a trade chamber (12%), having a loan with collateral (7.4%) and the percentage of purchases paid for before delivery (4.4%). The negative contributions are the percentage of working capital financed with internal funds (-9.2%) and the interest rate of the loan (-8.9%).

iv) *Quality, innovation and labor skills*. This group contributes with 12% of aggregate TFP and its effects tend to be concentrated in the largest firms in terms of share of sales. The aggregate effect of the entire group is only higher in Korea and Germany. The two key variables of this group are R&D investment which contributes with 23% of demeaned aggregate TFP affecting mainly largest firms; and the percentage of staff with university education constituting 3.3% of aggregate TFP. R&D intensity is significant also in Germany and Korea, the largest contribution estimated for Spain is directly consequence of the fact that R&D is more concentrated in largest firms in Spain than in Germany and Korea.

v) *Other control variables*. Within the most important variables of this heterogeneous group we find the percentage of sales directly exported which contributes with 34% of aggregate TFP with effects concentrated in high share of sales firms. The other variables with a positive effect are *dummy for incorporated company* (15%) and the *dummy for acquired another firm* (17%).

References

Acemoglu, D., S. Johnson, and J. A. Robinson (2001): "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review* 91 (5) 1369-1401.

Ahn, S. (2001): "Firm Dynamics and Productivity Growth: A Review of Micro Evidence from OECD Countries," *OECD Economics Department*, Working Paper No. 297, OECD Publishing.

Alfaro, L., A. Charlton and F. Kanczuk (2007): "Firm-size Distribution and Cross-Country Income Differences," unpublished manuscript.

Bartelsman E. and M. Doms (2000): "Understanding Productivity: Lessons from Longitudinal Microdata," *Journal of Economic Literature*, American Economic Association, vol 38(3), 569-594, September.

Bartelsman, E, J. Haltiwanger and S. Scarpetta (2006): "Cross-Country Differences in Productivity: The Role of Allocative Efficiency," unpublished manuscript.

Bastos, F. and J. Nasir (2004): "Productivity and the Investment Climate: What Matters Most?" *The World Bank Policy Research Working Paper Series:* #3335.

Basu, S.R. (2008): "A new way to link development to institutions, policies and geography," *HEI* Working Paper Number: 04-2008, International Economics Department, Graduate Institute of International Studies, Geneva. March 2008.

Basu, S.R. and M. Das, (2008): "Institution and Development Revisited: A Nonparametric Approach," *HEI* Working Paper No: 05/2008.

Bernard, A. B., J. Eaton, J. B. Jensen, and S. Kortum (2003): "Plants and Productivity in International Trade," *American Economic Review*, 93(4), 1268-1290.

Caselli, F., (2005): "Accounting for Income Differences Across Countries," chapter 9 in the *Handbook of Economic Growth* Vol. 1A, P. Aghion and S. Durlauf, eds., North Holland.

Caves, D., L. Christensen, and E. Diewert (1982): "Multilateral comparisons of output, input, and productivity using superlative index numbers," *The Economic Journal*, 92(365), 73-86.

de Loecker, J. (2007): "Product Differentiation, Multi-Product Firms, and Estimating the Impact of Trade Liberalization on Productivity," *NBER Working Paper* #13155.

De Soto, Hernando (2002): "The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else," *New York: Basic Books Press*.

Delgado, M., J. Fariñas, and S. Ruano (2002): "Firm Productivity and Export Markets: a Non-Parametric Approach," *Journal of International Economics*, 57(2), 397–422.

Dethier, J.J, M. Hirn and S. Straub (2008): "Explaining Enterprise Performance in Developing Countries with Business Climate Survey Data," The*World Bank Policy Research Working Paper* #4792.

Doraszelski, U. and J. Jaumandreu (2007): "R&D and productivity: Estimating production functions when productivity is endogenous," unpublished manuscript.

Easterly, W. and R. Levine (2003): "Tropics, Germs, and Crops: How Endowments Influence Economic Development," *Journal of Monetary Economic*, 50 (1) 3 - 39.

Eifert, B., A. Gelb and V. Ramachandran (2005): "Business Environment and Comparative Advantage in Africa: Evidence from the Investment Climate Data," *Center for Global Development Working Paper* No. 56. Available at SSRN: http://ssrn.com/abstract=997383

Erosa, A and A. Cabrillana (2007): "On Finance as a Theory of TFP, Cross-Industry Productivity Differences, and Economic Rents," *University of Toronto working paper*.

Escribano A. and J. L. Guasch (2005): "Assessing the Impact of the Investment Climate on Productivity using Firm Level Data: Methodology and the Cases of Guatemala, Honduras and Nicaragua," *The World Bank Policy Research Working Paper # 3621*.

Escribano A. and J. L. Guasch (2008): "Robust Methodology for Investment Climate Assessment on Productivity: Application to Investment Climate Surveys from Central America," *Universidad Carlos III de Madrid*, Working Paper # 08-19 (11).

Escribano, A. and J. Pena (2009): "Empirical Econometric Evaluation of Alternative Methods of Dealing with Missing Values in Investment Climate Surveys," *Universidad Carlos III de Madrid*, unpublished manuscript.

Escribano, A., and R. Stucchi (2008): "Catching Up in Total Factor Productivity through the Business Cycle: Evidence from Spainish Manufacturing Surveys," *Universidad Carlos III de Madrid*, Working Paper #08-51.

Escribano, A., J. L. Guasch and J. Pena. (2008): "Assessing the Impact of Infrastructure Quality on Firm Productivity in Africa," *World Bank Policy Research Working Paper* (forthcoming).

Escribano, A., J. L. Guasch, M. de Orte and J. Pena. (2009): "Investment Climate Assessment in Indonesia, Malaysia, Philippines and Thailand: Results from Pooling Firm Level Data," *Singapore Economic Review*, Special Issue on the Econometric Analysis of Panel Data, Vol. 54, #3, August 2009.

Escribano, A., J. L. Guasch, M. de Orte and J. Pena. (2008a): "Investment Climate Assessment Based on Demean Olley and Pakes Decompositions: Methodology and Applications to Turkey's Investment Climate Survey," *Universidad Carlos III de Madrid*, Working Paper # 08-20 (12).

Escribano, A., J. L. Guasch, M. de Orte and J. Pena. (2008b): "Investment Climate and Firm's Performance: Econometric and Applications to Turkey's Investment Climate Survey," *Universidad Carlos III de Madrid*, Working Paper # 08-20 (12).

Escribano, A., M. de Orte and J. Pena. (2009): "Econometric Methods for Investment Climate Assessment on Economic Performance in Pakistan.," Background paper prepared for the Investment Climate Assessment of Pakistan, the World Bank, Washington DC, 2009.

European Bank for Reconstruction and Development (2005): "The Business Environment and Enterprise Performance Survey (BEEPS) 2005: A brief report on observations, experiences and methodology from the surveys in Spain and the Irish Republic," mimeo the European Bank for Reconstruction and Development and Synovate, November 2005.

Fariñas, J., and S. Ruano (2004): "The Dynamics of Productivity: A Decomposition Approach Using Distribution Functions," *Small Business Economics*, 22(3-4), 237–251.

Fernandes A.M. and A. Pakes (2008): "Factor Utilization in Indian Manufacturing: A Look at the World Bank Investment Climate Surveys Data," *NBER Working Papers* #14178, National Bureau of Economic Research, Inc.

Foster, L., J. C. Haltiwanger, C. J. Krizan (2001):"Aggregate Productivity Growth: Lessons from Microeconomic Evidence," *Chicago: University of Chicago Press*.

Foster, L., J. Haltiwanger and C. Syverson (2008): "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitablity?" *American Economic Review*, March.

Frazer, G., (2005): "Which Firms Die? A Look at Manufacturing Firm Exit in Ghana," *Economic Development and Cultural Change*, University of Chicago Press, vol. 53(3), pages 585-617, April.

Glodberg, I, B. L.J. Radulovic, M.E Schaffer (2005): "Productivity, ownership, and the investment climate: International lessons for priorities in Serbia," *World Bank Policy Research Working Paper* #3681.

Goedhuys, M., N. Janz and P. Mohnen (2006): "What drives productivity in Tanzanian manufacturing firms: technology or institutions?" *UNU-MERIT working paper*, Netherlands.

Gorodnichenko, Y. (2007): "Using Firm Optimization to Evaluate and Estimate Returns to Scale," unpublished manuscript, UC Berkeley, Sept.

Griliches, Z. and J. Mairesse (1995): "Production Functions: The Search for Identification," *NBER Working Paper* #5067, National Bureau of Economic Research, Inc.

Haltiwanger, J. y H. Schweiger (2005): "Allocative Efficiency and the Business Climate", unpublished manuscript.

Hall, R. E. and C. I. Jones (1999): "Why Do Some Countries Produce So Much More Output Per Worker Than Others?" *Quarterly Journal of Economics* #114: 83-116.

Hsieh, C-T and P. J. Klenow (2007): "Misallocation and Manufacturing TFP in China and India," Unpublished manuscript.

Huergo, E., and J. Jaumandreu (2004): "Firms' Age, Process Innovation and Productivity Growth," *International Journal of Industrial Organization*, 22, 541-559.

Jorgenson D.W. (2001): "Information Technology and the U.S. Economy," *American Economic Review*, Vol. 91, 1, 1-32.

Jorgenson D.W., F. Gollop and B. Fraumeni (1987): "Productivity and U.S. Economic Growth," Cambridge: Harvard University Press.

Kasper, W. (2002): "Losing Sight of the Lodestar of Economic Freedom: A Report Card on New Zeland's Economic Reform," *NZ Business Roundtable*.

Katayama, H., S. Lu, and J. R. Tybout (2006): "Firm-Level Productivity Studies: Illusions and a Solution," unpublished manuscript, The University of Sydney, Charles River Associates and Pennsylvania State University.

Kerr, R. (2002): "The Quagmire of Regulation," *NZ Business Roundtable*. http://www.nzbr.org.nz/documents/speeches/speeches200/the_quagmire_of_regulation.pdf

Klenow, P. J. and A. Rodríguez-Clare (1997): "The Neoclassical Revival in Growth Economics: Has It Gone Too Far?" in B. Bernanke and J. Rotemberg, eds., NBER Macroeconomics Annual (*MIT Press, Cambridge*) 73-103.

Levinsohn, J and M. Melitz (2002): "Productivity in a Differentiated Products Market Equilibrium". unpublished manuscript, NBER

Lewis, W.W. (2005): "The Power of Productivity: Wealth, Poverty, and the Threat to Global Stability." *The University of Chicago Press.*

Lopez-García, P., S. Puente, and A. L. Gómez (2007): "Firm Productivity Dynamics in Spain," *Bank of Spain,* Documentos de Trabajo 0739.

Marschak, J. and W.H. Andews (1944): ""Random Simultaneous Equations and the Theory of Production," *Econometrica*, 1944-5.

Martin-Marcos, A. and J. Jaumandreu (2004): "Entry, exit and productivity growth: Spanish manufacturing during the eighties," *Spanish Economic Review*, 6, 211-226

Mas, M. y J. Quesada, (2005). "ICT and Economic Growth in Spain 1985-2002". *EU-KLEMS Working Paper series* #1.

Mas, M. (2006). "Infrastructures and ICT: Measurement Issues and Impact on Economic Growth". *EU-KLEMS Working Paper series* #12.

McMillan, J. (1998): "Managing Economic Change: Lessons from New Zealand," *The World Economy*, 21, 827-43.

Melitz, M. (2000): "Estimating Firm-level Productivity in Differentiated Product Industries," unpublished manuscript, *Harvard University.*

Olley G. S. and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, Vol. 64, 6, 1263-1297.

Ornaghi, C. (2006): "Spillovers in Product and Process Innovation: Evidence from Manufacturing Firms," *International Journal of Industrial Organization*, 24, 349-380.

Porter, M.E. and K. Schwab (2008). "Global Competitiveness Report 2008-2009". *The World Economic Forum*, Geneva Switzerland, 2009.

Prescott, E. C. (1998): "Needed: A Theory of Total Factor Productivity," *International Economic Review*, Department of Economics, University of Pennsylvania and Osaka University Institute of Social and Economic Research Association, vol. 39(3), pages 525-51, August.

Reinikka, R., and J. Svensson (1999): "How Inadequate Provision of Public Infrastructure and Services Affect Private Investment?" *the World Bank Policy Research Working Paper* #2262.

Restuccia, D. and R. Rogerson (2007): "Policy Distortions and Aggregate Productivity with Heterogeneous Plants," *NBER Working Paper* No. W13018.

Rodrik, D., A. Subramanian and F. Trebbi (2004): "Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development," *Journal of Economic Growth*, 9 (2) 131 - 165.

Sanchez, R., and L. Toharia (2000): "Temporary workers and productivity: the case of Spain," *Applied Economics*, 32, 583–591.

Schimtz (2005): "What Determines Productivity? Lessons From the Dramatic Recovery of the U.S. and Canadian Iron Ore Industries Following Their Early 1980s Crisis," *Federal Reserve Bank of Minneapolis Research Department Staff Report 286.*

Segura, J. (2006): "La Productividad de la Economía Española," *Fundación Ramón Areces*.

Solow R. M. (1957): "Technical Change and the Aggregate Production Function," *The Review of Economics and Statistics*, 39 (3), 312-320.

Stucchi, R (2008): "What determines productivity dynamics at the firm level? Evidence from Spain," *MPRA Paper* No. 6564.

Stucchi, R., y J.J Dolado (2008): "Do Temporary Contracts Affect TFP? Evidence from Spanish Manufacturing Firms," *IZA Discussion Paper #3823*.

Van Ark, B., M. O´Mahoney and G. Ypma (2007): "EU-KLEMS Productivity Report: An Overview of Results from the EU KLEMS Growth and Productivity Accounts for the European Union, EU Member States and Major Other Countries in the World," *European Commission, Research Directorate General.*

Van Biesebroek, J. (2007): "Robustness of Productivity Estimates," *Journal of Industrial Economics*, vol. 55, n°3.

Veeramani, C., and G. Goldar (2004): "Investment Climate and Total Factor Productivity in Manufacturing: Analysis of Indian States," *Indian Council for Research on International Economic Relations*, Working paper #127, New Delhi, India.

World Bank (2005): "World Development Report 2005: A Better Investment Climate for Everyone," *The World Bank and Oxford University Press*, Washington, D.C.

World Bank (2009): "Doing Business Report 2009: Comparing Regulation in 181 Economies," *The World Bank Group*, Washington DC, USA.

Appendix I: Definition of variables

I. Production function (PF) variables and general information at plant level

1. Sales: Used as the measure of output for the production function estimation. Sales are defined as total annual sales. The series are deflated by using the Wholesale Price Index (WPI), base 2003. Translated to US dollars (IMF database)

2. Employment: Total number of permanent and temporal workers.

3. Total hours worked per year: Total number of employees multiplied by the average hours worked per year.

4. Materials: Total costs of intermediate and raw materials used in production (excluding fuel). The series are deflated by using the Producer Price Indexes (PPI), base 2000.

5. Capital stock: Net book value of machinery and equipment. The series are deflated by using the Producer Price Indexes (PPI), base 2000.

6. User cost of capital: The user cost of capital is defined in terms of the opportunity cost of using capital; it is defined as a 15% of the net book value of machinery and equipment.

7. Labor cost: Total expenditures on personnel. The series are deflated by using the Producer Price Indexes (PPI), base 2000.

8. Industrial classification: i) food and beverages; ii) textiles and wearing apparel; iii) chemical products, petroleum, coal, rubber and plastics; iv) non-metallic metal products; v) fabricated metal products, excluding machinery and equipment; vi) machinery and equipment, excluding electrical; vii) electrical machinery apparatus, appliances and supplies; viii) transport equipment.

9. Regional classification: i) Marmara; ii) Ege; iii) Ic Anadolu; iv) Akdeniz; v) Karadeniz (Dogu Anadolu).

10. Size classification:

II. Investment climate (IC) variables

Infrastructures

1. Days to clear customs for exports: Average number of days to clear customs to export (log).

2. Days to clear customs for imports: Average number of days to clear customs to imports (log).

3. Number of power outages: Number of power outages suffered by the plant in fiscal year 2003 (FY03) (log).

4. Average duration of power outages: Average duration of power outages suffered by the plant in hours (log) in FY03, conditional on the plant reports having power outages.

5. Losses due to power outages: Value of the losses due to the power outages as a percentage of sales in FY03 (conditional on the plant reports having power outages).

6. Number of water outages: Number of water outages suffered by the plant in 2003 (log).

7. Average duration of water outages: Average duration of water outages suffered by the plant in hours (log).

8. Losses due to water outages: Value of the losses due to the water outages as a percentage of sales (conditional on the plant reporting water outages).

9. Shipment losses, domestic: Fraction of the value of the plant's average cargo consignment that was lost in transit due to breakage, theft, spoilage or other deficiencies of the transport means used.

10. Shipment losses, international: Fraction of the value of the plant's average cargo consignment that was lost in transit due to breakage, theft, spoilage or other deficiencies of the transport means used.

11. Dummy for email: Dummy variable taking value 1 if the plant uses email.

12. Dummy for internet page: Dummy variable taking value 1 if the plant has a website.

Bureaucracy and others

1. Dummy for crime: Dummy variable taking value 1 if the plant suffered any criminal attempt during 2003.

2. Losses due to crime: Value of losses due to criminal activity (log).

3. Security expenses: Cost in security (equipment, staff, etc) (log).

4. Illegal payments for protection: Cost due to protection payments e. g. to organized crime to prevent violence (bribery) (log).

5. Dummy for consulting: Dummy variable taking value 1if the firm uses consultants or employments to help deal with bureaucratic issues.

6. Dummy for payments to deal with bureaucratic issues: variable taking value 1 if firms in the main sector occasionally need to give gifts or make informal payments to public officers in order to "get things done" with regard to customs, taxes, licenses, legislations, services, etc.

7. Manager's time spent in bureaucratic issues: percentage of managers' time spent in dealing with bureaucratic issues.

8. Dummy for informal competition: variable taking value 1 if the firm competes with informal (no registered) firms.

9. Sales undeclared to taxes: Percentage of total sales undeclared to taxes.

10. Labor costs undeclared: Percentage of workforce undeclared to taxes.

11. Number of inspections: In the last year, total number of inspections (log).

12. Payments to obtain a contract with the government: variable taking value 1 if in plant's sector it is common to pay an extra amount of money in order to obtain a contract with the government.

13. Conflicts with clients: Percentage of conflicts with clients solved in the courts in the last two years.

14. Average duration of conflicts: Average weeks that take to resolve a conflict from the moment the case was brought to court until the moment the court decided the case.

15. Production lost due to absenteeism: Days of production lost due to absenteeism (log).

16. Wait for a construction related permit: Actual delay to obtain a construction related in days (log).

17. Wait for a main operating license: Actual delay to obtain a main operating license in days (log).

18. Wait for a main operating license: Actual delay to obtain a main operating license in days (log).

19. Delay to obtain a phone connection: days it takes to obtain a phone connection from the day the plant submit the (log).

20. Delay to obtain an electricity supply: Actual delay to obtain a water connection in days (log).

21. Delay to obtain a health certification: Actual delay to obtain a health certification in days (log).

22. Dummy for new land or building: variable taking value 1 if the firm acquired or attempted to acquire new land or buildings to expand operations in the previous 3 years.

23. Delay to obtain a land or a building: Total time that took from the moment the firm decided to buy a new land or building to the moment the firm finally got it (Including all the time required for official registration, negotiations with the seller and obtaining all licenses and necessary development\ permits and excluding the time needed for the construction permits).

24. Transaction fees to obtain a land or a building: Total cost related with transaction fees (including registration fees, payments to lawyers, brokers, etc) to obtain a land or a building.

25. Payment to government or private parties to obtain a land or a building: Total cost in informal payments to government officials or private parties to obtain a new land or buildings

26. Dummy for contract enforcement: Dummy variable taking value 1 if the conflict of the firm with clients solved in courts were generally enforced.

27. Dummy for lawsuit: Dummy variable taking value 1 if the firm has been involved in a lawsuit in the last three years.

28. Delayed payments: Percentage of monthly total sales to private customers that were not paid within the agreed time.

29. Sales never repaid: Percentage of monthly total sales to private customers that were never repaid.

Finance

1. Dummy for credit line: Dummy variable taking value 1 if the plant reports that it has a credit line.

2. Dummy for loan: Dummy variable taking value 1 if the plant reports that it has a bank loan.

3. Dummy for loan outstanding: variable taking value 1 if the firm has a loan outstanding from a financial institution.

4. Dummy for loan bank: Dummy variable taking value 1 if the firm has a loan from a domestic private commercial banks.

5. Dummy for loan leasing: Dummy variable taking value 1 if the firm has a loan from a leasing arrangement.

6. Dummy for loan public: Dummy variable taking value 1 if the firm has a loan from a state owned banks.

7. Dummy for loan informal: Dummy variable taking value 1 if the firm has a loan from Informal sources (e.g. money lender).

8. Dummy for loan DOT: Dummy variable taking value 1 if the firm has a loan from the Small and Medium Sized Industry Development Organization of Turkey (Incentive Credit for Export)

9. Dummy for loan Turkish Lira: Dummy variable taking value 1 if the loan is denominated in Turkish Lira.

10. Dummy for loan foreign currency: Dummy variable taking value 1 if the loan is denominated in a foreign currency.

11. Dummy for loan with collateral: Dummy variable taking value 1 if the loan is on collateral.

12. Dummy for loan long term: Dummy variable taking value 1 if the duration of the loan is more than months.

13. Borrows foreign: Percentage of borrows denominated in foreign currency.

14. Dummy for rent land: variable taking value 1 if the plant rents almost all its lands.

15. Dummy for rent buildings: variable taking value 1 if the plant rents almost all its buildings.

16. Dummy for external auditory: variable taking value 1 if the plant has its annual statements engaged in a process of external auditory.

Quality, innovation and labor skills

1. Dummy for quality certification: variable taking value 1 if the plant has a quality certification.

2. Dummy for new product: variable taking value 1 if the plant has developed a new product or product line.

3. Dummy for product upgraded: variable taking value 1 if the plant upgraded an existing product last year.

4. Dummy for new technology purchased: variable taking value 1 if the firm purchased any new technology during last year.

5. Dummy for licensed technology: variable taking value 1 if the firm used a licensed technology of a foreign company in the last year.

6. Dummy for education of the manager: variable taking value 1 if the manager of the plant has a bachelor or higher education degree.

7. Conflicts with employees: times in the last year the firm was taken to court by its current and former employees

8. Duration of conflicts with employees: average weeks that take to resolve a conflict with an employee from the moment the case was brought to court until the moment the court decided the case.

9. Staff-skilled workers: percentage of skilled workers in firm's staff.

10. Staff-unskilled workers: percentage of unskilled workers in firm's staff.

11. Staff-professional workers: percentage of professional workers in firm's staff.

12. Staff-part time workers: percentage of part time workers in firm's staff.

13. Staff-female workers: percentage of female workers in firm's staff.

14. Staff-temporal workers: percentage of temporal workers in firm's staff.

15. Dummy for internal training: variable taking value 1 if the plant provides internal training to its employees.

16. Dummy for external training: variable taking value 1 if the plant provides external training to its employees.

17. Training skilled workers: percentage of skilled workers that received training during last year.

18. Training unskilled workers: percentage of unskilled workers that received training during last year.

19. Weeks of training of skilled workers: number of weeks of training received by the skilled workers during last year.

20. Weeks of training of unskilled workers: number of weeks of training received by the unskilled workers during last year.

21. Staff with university education: percentage of staff with at least one year of university.

22. Staff-middle education: percentage of staff with completed high school (11 years) or complete secondary school (8 years).

23. Staff-basic education: percentage of staff with primary school either completed or not.

Other control variables

1. Dummy for incorporated company: variable taking value 1 if the plant is an incorporated company.

2. Dummy for public: variable taking value 1 if the firm belongs to the government.

3. Dummy for foreign direct investment: variable taking value 1 if any part of the capital of the firm is foreign.

4. Age of the firm: Difference between the year that the plant started operations and current year.

5. Number of competitors: number of competitors in the main market (log).

6. Dummy for exporter: variable taking value 1 if exports are greater than 10%.

7. Dummy for importer: variable taking value 1 if imports are greater than 10%.

8. Percentage of capacity utilization: Average percentage of plant's capacity used during last year.

9. Dummy for holding company: variable taking value 1 if the firm belongs to a holding company.

10. Competitive pressure: categorical variable taking value 1 if the number of competitors in firm's main market has increased during last year.

11. Percentage of workforce unionized: percentage of workers that belongs to a syndicate.

12. Production lost due to strikes: days of production lost due to strikes (log).

13. Dummy for ownership: variable taking value 1 if the firm previously belonged to the government.

14. Dummy for industrial zone: variable taking value 1 if the firm is located in an industrial zone.

15. Dummy for foreign competition: variable taking value 1 if the firm competes with foreign firms.

Appendix II: Tables and figures

Table 1a. Global Competitiveness Report, basic requirements for competitiveness in Spain and comparators

In parenthesis are the ranks of each country within 138 economies, in brackets are the scores.

Rank	Institutions	Infrastructures	Macro stability	Health and primary education	Overall
1	NET (10) [5.76]	**GER (1) [6.65]**	**KOR (4) [6.15]**	FRA (9) [6.35]	**GER (7) [5.96]**
2	**GER (14) [5.65]**	FRA (2) [6.54]	**SPN (30) [5.53]**	NET (11) [6.3]	NET (10) [5.81]
3	**IRL (17) [5.39]**	US (7) [6.1]	NET (36) [5.45]	**IRL (14) [6.28]**	FRA (13) [5.76]
4	FRA (23) [5.1]	JAP (11) [5.8]	**GER (40) [5.42]**	UK (19) [6.17]	**KOR (16) [5.71]**
5	UK (25) [4.99]	NET (12) [5.71]	**IRL (47) [5.33]**	JAP (22) [6.11]	US (22) [5.5]
6	JAP (26) [4.99]	**KOR (15) [5.63]**	UK (58) [5.15]	**GER (24) [6.1]**	UK (24) [5.46]
7	**KOR (28) [4.95]**	UK (18) [5.52]	FRA (65) [5.04]	**KOR (26) [6.1]**	JAP (26) [5.36]
8	US (29) [4.93]	**SPN (22) [5.3]**	US (66) [4.99]	ITL (30) [6.04]	**SPN (27) [5.34]**
9	**PRT (35) [4.75]**	**PRT (26) [5.07]**	**PRT (82) [4.74]**	PRT (33) [6]	**IRL (32) [5.24]**
10	**SPN (43) [4.59]**	**GRE (45) [4.28]**	JAP (98) [4.53]	US (34) [5.97]	**PRT (37) [5.14]**
11	**GRE (58) [4.1]**	**IRL (53) [3.95]**	ITL (100) [4.46]	**SPN (35) [5.96]**	**GRE (51) [4.66]**
12	ITL (84) [3.68]	ITL (54) [3.94]	**GRE (106) [4.37]**	**GRE (40) [5.89]**	ITL (58) [4.53]

Spain (SPN), Germany (GER), Netherland (NET), Ireland (IRL), Portugal (PRT), Greece (GRE), Japan (JAP), United States (US), Italy, (ITL), United Kingdom (UK), France (FRA), Korea (KOR).
In bold and red is Spain. In bold are those countries included in the econometric analysis of the investment climate.
Source: authors' elaboration with data from Global Competitiveness Report 2008, The World Economic Forum.

Table 1b. Global Competitiveness Report, efficiency enhancers in Spain and comparators

In parenthesis are the ranks of each country within 138 economies, in brackets are the scores.

Rank	Higher education and training	Goods markets efficiency	Labor market efficiency	Financial market sophistication	Technologies readiness	Market size	Overall
1	US (5) [5.67]	NET (3) [5.39]	US (1) [5.79]	UK (5) [5.81]	NET (1) [6.01]	US (1) [6.91]	US (1) [5.81]
2	NET (11) [5.52]	US (8) [5.32]	UK (8) [5.19]	**IRL (7) [5.68]**	UK (8) [5.62]	JAP (3) [6.15]	UK (4) [5.45]
3	**KOR (12) [5.51]**	**IRL (9) [5.3]**	JAP (11) [5.09]	US (9) [5.61]	US (11) [5.57]	**GER (4) [5.99]**	NET (7) [5.38]
4	FRA (16) [5.37]	**GER (15) [5.19]**	**IRL (15) [4.95]**	NET (11) [5.57]	**KOR (13) [5.51]**	UK (6) [5.77]	**GER (11) [5.22]**
5	UK (18) [5.27]	JAP (18) [5.13]	NET (30) [4.72]	**GER (19) [5.35]**	**GER (18) [5.22]**	FRA (7) [5.73]	JAP (12) [5.22]
6	**IRL (20) [5.18]**	UK (19) [5.05]	**KOR (41) [4.6]**	FRA (25) [5.19]	FRA (20) [5.16]	ITL (9) [5.65]	**KOR (15) [5.15]**
7	**GER (21) [5.15]**	FRA (21) [5.01]	**GER (58) [4.43]**	**SPN (36) [4.93]**	JAP (21) [5.11]	**SPN (12) [5.47]**	FRA (16) [5.09]
8	JAP (23) [5.08]	**KOR (22) [5]**	**PRT (87) [4.18]**	**KOR (37) [4.85]**	**IRL (24) [4.98]**	**KOR (13) [5.44]**	**IRL (19) [5.05]**
9	**SPN (30) [4.75]**	**SPN (41) [4.63]**	**SPN (96) [4.11]**	JAP (42) [4.75]	**SPN (29) [4.59]**	NET (18) [5.06]	**SPN (25) [4.75]**
10	**PRT (37) [4.59]**	**PRT (45) [4.53]**	FRA (105) [4.05]	**PRT (43) [4.71]**	ITL (31) [4.52]	**GRE (33) [4.52]**	**PRT (34) [4.47]**
11	**GRE (38) [4.52]**	ITL (62) [4.24]	**GRE (116) [3.89]**	**GRE (67) [4.29]**	**PRT (32) [4.51]**	**PRT (43) [4.32]**	ITL (42) [4.38]
12	ITL (44) [4.43]	**GRE (64) [4.22]**	ITL (126) [3.56]	ITL (91) [3.9]	**GRE (59) [3.5]**	**IRL (48) [4.22]**	**GRE (57) [4.16]**

Spain (SPN), Germany (GER), Netherland (NET), Ireland (IRL), Portugal (PRT), Greece (GRE), Japan (JAP), United States (US), Italy, (ITL), United Kingdom (UK), France (FRA), Korea (KOR).
In bold and red is Spain. In bold are those countries included in the econometric analysis of the investment climate.
Source: authors' elaboration with data from Global Competitiveness Report 2008, The World Economic Forum.

Table 1c. Global Competitiveness Report, innovation and sophistication factors in Spain and comparators

In parenthesis are the ranks of each country within 138 economies, in brackets are the scores.

Rank	Business sophistication	Innovation	Overall
1	US (1) [5.84]	US (1) [5.8]	**GER (1) [5.87]**
2	JAP (4) [5.52]	JAP (3) [5.65]	Japan (3) [5.78]
3	**GER (8) [5.22]**	**GER (4) [5.54]**	US (4) [5.75]
4	**KOR (9) [5.18]**	NET (9) [5.2]	NET (8) [5.58]
5	NET (12) [4.82]	**KOR (10) [5.2]**	FRA (9) [5.5]
6	FRA (16) [4.67]	FRA (14) [5.08]	**KOR (16) [5.22]**
7	UK (17) [4.66]	UK (17) [4.93]	UK (17) [5.2]
8	**IRL (21) [4.39]**	**IRL (20) [4.72]**	**IRL (19) [5.05]**
9	**PRT (35) [3.66]**	**SPN (29) [4.25]**	ITL (21) [4.99]
10	**SPN (39) [3.61]**	ITL (31) [4.19]	**SPN (24) [4.89]**
11	ITL (53) [3.38]	**PRT (43) [4.03]**	**PRT (48) [4.39]**
12	**GRE (63) [3.18]**	**GRE (68) [3.65]**	**GRE (66) [4.13]**

Spain (SPN), Germany (GER), Netherland (NET), Ireland (IRL), Portugal (PRT), Greece (GRE), Japan (JAP), United States (US), Italy, (ITL), United Kingdom (UK), France (FRA), Korea (KOR).
In bold and red is Spain. In bold are those countries included in the econometric analysis of the investment climate.
Source: authors' elaboration with data from Global Competitiveness Report 2008, The World Economic Forum.

Table 2a: Spain in the rankings on the ease of doing business

In parentheses are the rankings within 181 economies

Rank	Ease of doing business	Starting a business	Dealing with construction permits	Employing workers	Registering property	Getting credit
1	US (3)	**IRL (5)**	**GER (15)**	US (1)	US (12)	UK (2)
2	UK (6)	US (6)	FRA (18)	JAP (17)	UK (22)	US (5)
3	**IRL (7)**	UK (8)	KOR (23)	UK (28)	NET (23)	JAP (12)
4	JAP (12)	FRA (14)	US (26)	**IRL (38)**	**SPN (46)**	**KOR (12)**
5	**KOR (23)**	**PRT (34)**	**IRL (30)**	ITA (75)	JAP (51)	**GER (12)**
6	**GER (25)**	NET (51)	JAP (39)	NET (98)	**GER (52)**	**IRL (12)**
7	NET (26)	ITA (53)	**GRE (45)**	**GRE (133)**	ITA (58)	NET (43)
8	FRA (31)	JAP (64)	**SPN (51)**	**GER (142)**	**KOR (67)**	FRA (43)
9	**PRT (48)**	**GER (102)**	UK (61)	FRA (148)	**PRT (79)**	**SPN (43)**
10	**SPN (49)**	**KOR (126)**	ITA (83)	**KOR (152)**	**IRL (82)**	ITA (84)
11	ITA (65)	**GRE (133)**	**NET (94)**	**SPN (160)**	**GRE (101)**	**PRT (109)**
12	**GRE (96)**	**SPN (140)**	**PRT (128)**	**PRT (164)**	FRA (166)	**GRE (109)**

In bold and red is Spain. In bold are those countries included in the econometric analysis of the investment climate.
Source: Authors' elaboration with data from Doing Business Report 2009, The World Bank Group, Washington, DC

Table 2b: Spain in the rankings on the ease of doing business

In parentheses are the rankings within 181 economies

Rank	Protecting investors	Paying taxes	Trading across borders	Enforcing contracts	Closing a business
1	US (5)	IRL (6)	**GER (11)**	US (6)	JAP (1)
2	**IRL (5)**	**UK (16)**	**KOR (12)**	**KOR (8)**	**IRL (6)**
3	UK (9)	NET (30)	NET (13)	**GER (9)**	UK (9)
4	JAP (15)	**KOR (43)**	US (15)	FRA (10)	NET (10)
5	**PRT (38)**	US (46)	JAP (17)	JAP (21)	**KOR (12)**
6	ITA (53)	**GRE (62)**	**IRL (18)**	UK (24)	US (15)
7	**KOR (70)**	FRA (66)	FRA (22)	NET (34)	**SPN (19)**
8	FRA (70)	**PRT (73)**	UK (28)	**PRT (34)**	**PRT (21)**
9	**GER (88)**	**GER (80)**	**PRT (33)**	**IRL (39)**	ITA (27)
10	**SPN (88)**	**SPN (84)**	**SPN (52)**	**SPN (54)**	**GER (33)**
11	NET (104)	JAP (112)	ITA (60)	**GRE (85)**	FRA (40)
12	**GRE (150)**	ITA (128)	**GRE (70)**	ITA (156)	**GRE (41)**

In bold and red is Spain. In bold are those countries included in the econometric analysis of the investment climate.
Source: Authors' elaboration with data from Doing Business Report 2009, The World Bank Group, Washington, DC.

Table 3: Summary of the number of establishments surveyed in the sampling frame and number of observations available before and after cleaning missing values and outliers, by country and industry

In parentheses are the percentages with respect to the totals by country

		Mining and Quarrying	Construction	Manufacturing	Transport, storage and communication	Wholesale trade	Real state	Hotels and restaurants	Other services	Total
SPN	Establishments surveyed	3, (0.49)	102, (16.8)	134, (22.1)	55, (9.07)	156, (25.7)	78, (12.8)	67, (11.0)	11, (1.81)	606
	Available before cleaning	0, (0)	89, (19.3)	105, (22.7)	35, (7.59)	116, (25.1)	55, (11.9)	53, (11.4)	8, (1.73)	461
	Available after cleaning	0, (0)	95, (17.4)	112, (20.6)	49, (9.02)	146, (26.8)	69, (12.7)	62, (11.4)	10, (1.84)	543
GER	Establishments surveyed	10, (0.83)	239, (19.9)	221, (18.4)	74, (6.18)	267, (22.3)	244, (20.3)	66, (5.51)	76, (6.34)	1197
	Available before cleaning	0, (0)	219, (20.6)	202, (19.0)	64, (6.02)	232, (21.8)	215, (20.2)	60, (5.64)	70, (6.59)	1062
	Available after cleaning	0, (0)	234, (20.4)	210, (18.3)	73, (6.36)	256, (22.3)	233, (20.3)	65, (5.66)	76, (6.62)	1147
GRE	Establishments surveyed	5, (0.91)	61, (11.1)	98, (17.9)	43, (7.87)	178, (32.6)	54, (9.89)	89, (16.3)	18, (3.29)	546
	Available before cleaning	0, (0)	52, (11.2)	82, (17.7)	38, (8.24)	146, (31.6)	46, (9.97)	80, (17.3)	17, (3.68)	461
	Available after cleaning	0, (0)	54, (11.1)	86, (17.8)	39, (8.07)	158, (32.7)	47, (9.73)	82, (16.9)	17, (3.51)	483
IRL	Establishments surveyed	6, (1.19)	45, (8.98)	175, (34.9)	41, (8.18)	97, (19.3)	72, (14.3)	40, (7.98)	25, (4.99)	501
	Available before cleaning	0, (0)	43, (9.97)	154, (35.7)	35, (8.12)	83, (19.2)	61, (14.1)	34, (7.88)	21, (4.87)	431
	Available after cleaning	0, (0)	45, (9.76)	166, (36.0)	37, (8.02)	88, (19.0)	67, (14.5)	36, (7.80)	22, (4.77)	461
KOR	Establishments surveyed	10, (1.67)	62, (10.3)	215, (35.9)	35, (5.85)	137, (22.9)	56, (9.36)	67, (11.2)	16, (2.67)	598
	Available before cleaning	0, (0)	50, (11.7)	155, (36.2)	21, (4.91)	109, (25.5)	29, (6.79)	52, (12.1)	11, (2.57)	427
	Available after cleaning	0, (0)	55, (10.7)	191, (37.4)	28, (5.49)	128, (25.0)	35, (6.86)	61, (11.9)	12, (2.35)	510
PRT	Establishments surveyed	2, (0.39)	59, (11.6)	132, (26.1)	24, (4.75)	146, (28.9)	55, (10.8)	51, (10.0)	36, (7.12)	505
	Available before cleaning	0, (0)	47, (12.5)	103, (27.3)	18, (4.78)	111, (29.5)	37, (9.84)	36, (9.57)	24, (6.38)	376
	Available after cleaning	0, (0)	47, (11.5)	113, (27.8)	21, (5.17)	120, (29.5)	39, (9.60)	39, (9.60)	27, (6.65)	406
Total	Establishments surveyed	36, (0.91)	568, (14.3)	975, (24.6)	272, (6.88)	981, (24.8)	559, (14.1)	380, (9.61)	182, (4.60)	3953
	Available before cleaning	0, (0)	500, (15.5)	801, (24.8)	211, (6.55)	797, (24.7)	443, (13.7)	315, (9.78)	151, (4.69)	3218
	Available after cleaning	0, (0)	530, (14.9)	878, (24.7)	247, (6.95)	896, (25.2)	490, (13.8)	345, (9.71)	164, (4.61)	3550

Notes:
Sampling frame refers to the number of establishments surveyed. The number of observations available before cleaning does not include both missing values and outliers. The number of observations after cleaning corresponds to the number of observations available after cleaning missing values and outliers.
Source: Authors elaboration with data from the BEEPS.

Table 4: Number of establishments surveyed in Spain by region, industry and size, and number of observations available after and before cleaning

In parentheses are the percentages with respect to the totals by country

By industry:		Madrid	Barcelona	Large cities	Medium cities	Small cities	Total
Construction	Establishments surveyed	28, (27.4)	13, (12.7)	20, (19.6)	19, (18.6)	22, (21.5)	102
	Available after cleaning	24, (26.9)	12, (13.4)	20, (22.4)	15, (16.8)	18, (20.2)	89
	Available before cleaning	27, (28.4)	13, (13.6)	20, (21.0)	15, (15.7)	20, (21.0)	95
Manufacturing	Establishments surveyed	50, (37.3)	20, (14.9)	15, (11.1)	25, (18.6)	24, (17.9)	134
	Available after cleaning	42, (40)	14, (13.3)	13, (12.3)	17, (16.1)	19, (18.0)	105
	Available before cleaning	43, (38.3)	17, (15.1)	14, (12.5)	17, (15.1)	21, (18.7)	112
Transport storage and	Establishments surveyed	27, (49.0)	8, (14.5)	5, (9.09)	10, (18.1)	5, (9.09)	55
	Available after cleaning	23, (65.7)	2, (5.71)	2, (5.71)	6, (17.1)	2, (5.71)	35
	Available before cleaning	26, (53.0)	7, (14.2)	5, (10.2)	7, (14.2)	4, (8.16)	49
Wholesale and retail	Establishments surveyed	38, (24.5)	21, (13.5)	26, (16.7)	30, (19.3)	40, (25.8)	155
	Available after cleaning	35, (30.1)	13, (11.2)	21, (18.1)	22, (18.9)	25, (21.5)	116
	Available before cleaning	38, (26.2)	20, (13.7)	25, (17.2)	26, (17.9)	36, (24.8)	145
Real estate, renting	Establishments surveyed	34, (43.5)	13, (16.6)	11, (14.1)	9, (11.5)	11, (14.1)	78
	Available after cleaning	31, (56.3)	6, (10.9)	10, (18.1)	2, (3.63)	6, (10.9)	55
	Available before cleaning	33, (47.8)	12, (17.3)	11, (15.9)	5, (7.24)	8, (11.5)	69
Hotels and restaurant	Establishments surveyed	20, (30.3)	7, (10.6)	12, (18.1)	8, (12.1)	19, (28.7)	66
	Available after cleaning	19, (36.5)	6, (11.5)	10, (19.2)	5, (9.61)	12, (23.0)	52
	Available before cleaning	20, (32.7)	7, (11.4)	11, (18.0)	6, (9.83)	17, (27.8)	61
Other services	Establishments surveyed	1, (9.09)	1, (9.09)	1, (9.09)	2, (18.1)	6, (54.5)	11
	Available after cleaning	0, (0)	1, (12.5)	1, (12.5)	1, (12.5)	5, (62.5)	8
	Available before cleaning	1, (10)	1, (10)	1, (10)	1, (10)	6, (60)	10
By size		Madrid	Barcelona	Large cities	Medium cities	Small cities	Total
Small	Establishments surveyed	136, (31.1)	56, (12.8)	60, (13.7)	80, (18.3)	105, (24.0)	437
	Available after cleaning	121, (36.5)	39, (11.7)	52, (15.7)	50, (15.1)	69, (20.8)	331
	Available before cleaning	129, (32.9)	55, (14.0)	57, (14.5)	58, (14.7)	93, (23.7)	392
Medium	Establishments surveyed	38, (33.9)	18, (16.0)	21, (18.7)	18, (16.0)	17, (15.1)	112
	Available after cleaning	32, (36.3)	10, (11.3)	17, (19.3)	14, (15.9)	15, (17.0)	88
	Available before cleaning	36, (36)	15, (15)	20, (20)	14, (14)	15, (15)	100
Large	Establishments surveyed	24, (43.6)	9, (16.3)	10, (18.1)	6, (10.9)	6, (10.9)	55
	Available after cleaning	21, (51.2)	5, (12.1)	8, (19.5)	4, (9.75)	3, (7.31)	41
	Available before cleaning	23, (46.9)	7, (14.2)	10, (20.4)	5, (10.2)	4, (8.16)	49
Total	Establishments surveyed	198, (32.7)	83, (13.7)	91, (15.0)	104, (17.2)	128, (21.1)	604
	Available after cleaning	174, (37.8)	54, (11.7)	77, (16.7)	68, (14.7)	87, (18.9)	460
	Available before cleaning	188, (34.7)	77, (14.2)	87, (16.0)	77, (14.2)	112, (20.7)	541

Notes:
Sampling frame refers to the number of establishments surveyed. The number of observations available before cleaning does not include both missing values and outliers. The number of observations after cleaning corresponds to the number of observations available after cleaning missing values and outliers.
Source: Authors elaboration with data from the BEEPS.

Table 5: Robust IC elasticities and semi-elasticities with respect to TFP in Spain

		Restricted estimation			Unrestricted estimation		
		Single step	Two steps		Single step	Two steps	
		Solow residual	Cobb-Douglas	Solow residual	Cobb-Douglas	Solow residual	Cobb-Douglas
IC blocks	**IC variables**	**Coef S.E**	**Coef S.E**	**Coef S.E**	**Coef S.E**	**Coef S.E**	**Coef S.E**
Average real wage		0.204 [0.115]*	0.097 [0.122]	0.091 [0.103]	0.068 [0.134]	0.078 [0.098]	0.133 [0.037]***
Infrastructures	Days to clear customs to export – interaction with dummy for exports (a)	-0.071 [0.041]*	-0.053 [0.043]	-0.016 [0.034]	-0.045 [0.043]	-0.028 [0.043]	-0.035 [0.040]
	Avg duration of phone outages	-0.016 [0.022]	-0.015 [0.018]	-0.021 [0.016]	-0.014 [0.019]	-0.029 [0.016]*	-0.026 [0.013]**
	Dummy for e-mail	0.153 [0.052]***	0.091 [0.052]*	0.081 [0.041]*	0.087 [0.049]*	0.105 [0.058]*	0.058 [0.030]*
	Shipment losses, domestic (b)	-0.001 [0.003]	-0.001 [0.002]	-0.002 [0.002]	-0.004 [0.002]	-0.002 [0.001]	-0.000 [0.002]
Bureaucracy and others	Manager's time spent in bur. issues (b)	-0.006 [0.004]	-0.005 [0.003]*	-0.004 [0.003]	-0.005 [0.003]	-0.006 [0.003]**	-0.003 [0.002]
	Dummy for national laws or regulation affecting	-0.08 [0.061]	-0.051 [0.047]	-0.07 [0.041]*	-0.082 [0.045]*	-0.022 [0.048]	-0.069 [0.043]
	Sales reported to taxes (b)	-0.002 [0.002]	-0.001 [0.002]	-0.002 [0.002]	-0.002 [0.002]	-0.002 [0.002]	-0.002 [0.001]
	Dummy taking 1 if the firm would reduce its current	-0.058 [0.029]*	-0.035 [0.029]	-0.019 [0.027]	-0.028 [0.031]	-0.024 [0.026]	-0.015 [0.031]
Finance	Sales paid before delivery (b)	0.002 [0.001]***	0.002 [0.001]**	0.002 [0.001]***	0.002 [0.001]***	0.002 [0.001]**	0.001 [0.001]
	Working capital financed by internal founds (b)	-0.001 [0.000]	-0.000 [0.000]	-0.000 [0.000]	-0.000 [0.000]	-0.000 [0.000]	-0.000 [0.000]
	Working capital financed by informal founds (b)	-0.003 [0.002]	-0.002 [0.002]	-0.001 [0.001]	-0.002 [0.001]	-0.002 [0.002]	-0.000 [0.001]
	Working capital financed by leasing (b)	0.004 [0.001]***	0.004 [0.002]**	0.004 [0.001]***	0.003 [0.002]	0.005 [0.001]***	0.007 [0.001]***
	Working capital financed by equity (b)	-0.001 [0.001]	-0.001 [0.001]	-0.001 [0.001]*	-0.001 [0.001]**	-0.001 [0.001]	-0.001 [0.001]
	Dummy for trade association	0.041 [0.030]	0.03 [0.026]	0.031 [0.021]	0.017 [0.027]	0.021 [0.024]	0.036 [0.022]
	Dummy for loan with collateral	0.099 [0.098]	0.119 [0.084]	0.136 [0.079]*	0.142 [0.099]	0.116 [0.078]	0.101 [0.056]*
	Interest rate of loan (b)	-0.036 [0.017]**	-0.028 [0.016]*	-0.026 [0.013]*	-0.036 [0.017]**	-0.027 [0.014]*	-0.018 [0.011]
Quality, innovation and labor skills	Dummy for R&D expenditures	0.114 [0.085]	0.132 [0.085]	0.147 [0.083]*	0.163 [0.099]	0.158 [0.076]**	0.094 [0.040]**
	Dummy for brought in house	-0.056 [0.052]	-0.051 [0.041]	-0.06 [0.036]	-0.062 [0.039]	-0.049 [0.034]	-0.038 [0.031]
	University education (b)	0.001 [0.001]	0.001 [0.000]	0.001 [0.000]	0.000 [0.000]	0.000 [0.000]	0.001 [0.000]
Other control variables	Dummy for incorporated company	0.053 [0.033]	0.039 [0.031]	0.022 [0.020]	0.029 [0.031]	0.03 [0.025]	0.028 [0.018]
	Share of exports (b)	0.19 [0.094]**	0.159 [0.095]*	0.106 [0.076]	0.147 [0.100]	0.116 [0.099]	0.098 [0.059]
	Dummy for merged with another firm (b)	-0.04 [0.079]	-0.061 [0.053]	-0.072 [0.040]*	-0.046 [0.057]	-0.073 [0.055]	-0.008 [0.039]
	Dummy for acquired another firm (b)	0.143 [0.046]***	0.078 [0.043]*	0.062 [0.044]	0.117 [0.039]***	0.085 [0.040]**	0.026 [0.042]
	Sector/region/size dummies	Yes	Yes	Yes	Yes	Yes	Yes
	Constant	Yes	Yes	Yes	Yes	Yes	Yes
	Observations	541	541	541	541	541	541
	R-squared	0.46	0.48	0.99	0.99	0.99	0.99

Notes:
Two steps estimation: first, we estimate TFP of equation (1) of section 3 by non-parametric techniques, obtaining two Solow residuals, restricted and unrestricted by industry. Second, we estimate (4) by OLS using as dependent variable TFPs estimated in the first step.
Single step estimation: we estimate (1), (2) and (3) b y OLS where (1) can be a Cobb-Douglas PF or a Translog and with two cases of input-output elasticities, restricted and unrestricted by industry.
Restricted input-output elasticities: equal input-output elasticities for all the firms in the same country.
Unrestricted input-output elasticities: equal input-output elasticities for all the firms in the same industry of each country.
* significant at 10%, ** significant at 5%, *** significant at 1%. Significance is given by robust standard errors [in brackets] correcting for correlation within clusters defined by industry and region.
(a) variable instrumented with the average by industry and region
(a) variable whose missing values are imputed by the expectation conditional on sector/region/size information.
Source: authors' estimations with BEEPS data

Table 6: Robust IC elasticities and semi-elasticities with respect to TFP in Germany

		Restricted estimation			Unrestricted estimation		
		Single step	Two steps		Single step	Two steps	
		Solow residual	Cobb-Douglas	Solow residual	Cobb-Douglas	Solow residual	Cobb-Douglas
IC blocks	**IC variables**	**Coef S.E**	**Coef S.E**	**Coef S.E**	**Coef S.E**	**Coef S.E**	**Coef S.E**
Average real wage		0.382 [0.040]***	0.398 [0.058]***	0.439 [0.070]***	0.414 [0.070]***	0.416 [0.072]***	0.403 [0.076]***
Infrastructures	Power outages (b)	-0.039 [0.014]***	-0.044 [0.013]***	-0.02 [0.014]	-0.025 [0.013]*	-0.025 [0.011]**	-0.025 [0.014]*
Bureaucracy and others	Security expenses (b)	-0.031 [0.008]***	-0.031 [0.007]***	-0.022 [0.006]***	-0.016 [0.006]***	-0.025 [0.006]***	-0.02 [0.008]**
	Dummy for having overdue payments	-0.062 [0.030]*	-0.063 [0.031]*	-0.044 [0.026]	-0.025 [0.026]	-0.05 [0.023]**	-0.032 [0.025]
Finance	Working capital financed by local private banks	0.001 [0.001]*	0.002 [0.001]*	0.002 [0.001]**	0.001 [0.001]	0.001 [0.001]**	0.001 [0.000]
	New fixed assets financed by government (b)	-0.002 [0.001]***	-0.001 [0.001]	-0.001 [0.001]	-0.002 [0.001]*	-0.001 [0.001]	-0.002 [0.001]*
	Percentage of profits reinvested (b)	0.0004 [0.000]	0.0004 [0.000]	0.0004 [0.000]	0.0004 [0.000]	0.0004 [0.000]	0.0004 [0.000]
Quality, innovation and labor skills	Dummy for R&D expenditures	0.062 [0.037]	0.046 [0.036]	0.047 [0.035]	0.025 [0.028]	0.031 [0.039]	0.017 [0.033]
	Dummy for new product technology (b)	0.016 [0.028]	0.042 [0.026]	0.017 [0.021]	0.03 [0.029]	0.031 [0.028]	0.046 [0.033]
	University education (b)	0.0004 [0.000]	0.0004 [0.000]	0.001 [0.000]	0.001 [0.000]*	0.001 [0.000]*	0.0004 [0.000]
Other control variables	Dummy for incorporated company	0.074 [0.039]*	0.041 [0.027]	0.057 [0.029]*	0.031 [0.026]	0.049 [0.027]*	0.046 [0.020]**
	Share of exports (b)	0.042 [0.040]	0.062 [0.042]	0.066 [0.042]	0.049 [0.038]	0.069 [0.044]	0.05 [0.042]
	Share of imports	0.001 [0.001]	0.001 [0.001]	0.001 [0.001]	0.001 [0.001]	0.001 [0.001]	0.001 [0.000]*
	Dummy for acquired another firm (b)	0.17 [0.136]	0.103 [0.137]	0.164 [0.113]	0.148 [0.097]	0.144 [0.129]	0.168 [0.107]
	Percentage of margin (b)	0.008 [0.001]***	0.008 [0.001]***	0.008 [0.001]***	0.008 [0.001]***	0.008 [0.001]***	0.008 [0.001]***
	Sector/region/size dummies	Yes	Yes	Yes	Yes	Yes	Yes
	Constant	Yes	Yes	Yes	Yes	Yes	Yes
	Observations	1147	1147	1147	1147	1147	1147
	R-squared	0.66	0.83	0.97	0.98	0.98	0.98

Notes:
See footnotes in Table 5.
Source: authors' estimations with BEEPS data

Table 7: Robust IC elasticities and semi-elasticities with respect to TFP in Greece

		Restricted estimation			Unrestricted estimation		
		Single step	Two steps		Single step	Two steps	
		Solow residual	Cobb-Douglas	Solow residual	Cobb-Douglas	Solow residual	Cobb-Douglas
IC blocks	**IC variables**	**Coef S.E**	**Coef S.E**	**Coef S.E**	**Coef S.E**	**Coef S.E**	**Coef S.E**
Average real wage		0.394 [0.042]***	0.348 [0.059]***	0.429 [0.055]***	0.394 [0.054]***	0.398 [0.052]***	0.387 [0.062]***
Infrastructures	Days to clear customs to export (interacting with dummy for exporting) (a)	-0.015 [0.012]	-0.016 [0.019]	-0.009 [0.020]	-0.006 [0.012]	-0.019 [0.013]	0.01 [0.012]
	Dummy for e-mail	0.084 [0.021]***	0.048 [0.029]	0.066 [0.019]***	0.049 [0.029]	0.039 [0.030]	0.062 [0.038]
Bureaucracy and others	Cases of commercial arbitration as plaintiff	-0.034 [0.025]	-0.034 [0.012]**	-0.035 [0.016]**	-0.038 [0.013]***	-0.03 [0.013]**	-0.027 [0.012]**
	Manager's time spent in bur. issues (b)	-0.003 [0.001]***	-0.004 [0.001]***	-0.002 [0.001]**	-0.002 [0.001]*	-0.003 [0.001]**	-0.002 [0.002]
	Workforce reported to taxes (b)	-0.001 [0.001]	-0.001 [0.001]	-0.002 [0.001]*	-0.001 [0.001]	-0.001 [0.001]	-0.001 [0.001]**
	Time to fill a vacancy for professional worker	-0.008 [0.015]	-0.01 [0.035]	-0.013 [0.019]	-0.018 [0.020]	-0.034 [0.023]	-0.017 [0.021]
Finance	Working capital financed by internal founds (b)	-0.001 [0.000]**	-0.000 [0.000]	-0.001 [0.000]*	-0.000 [0.000]	-0.000 [0.000]	-0.000 [0.000]
	New fixed assets financed by foreign private banks (b)	0.004 [0.001]***	0.005 [0.002]**	0.004 [0.002]	0.002 [0.002]	0.005 [0.002]**	0.003 [0.002]
	Working capital financed by equity (b)	-0.001 [0.000]	-0.000 [0.000]	-0.001 [0.000]**	-0.000 [0.000]	-0.001 [0.000]	-0.000 [0.000]
	Working capital financed by family/friends (b)	-0.001 [0.001]	-0.004 [0.001]***	-0.002 [0.001]	-0.001 [0.001]	-0.002 [0.001]*	-0.000 [0.001]
	Dummy for international accounting standards (b)	0.112 [0.063]*	0.096 [0.069]	0.124 [0.072]	0.1 [0.071]	0.107 [0.069]	0.089 [0.058]
	Dummy for external auditory (b)	0.055 [0.027]*	0.066 [0.029]**	0.059 [0.026]**	0.052 [0.022]**	0.063 [0.028]**	0.054 [0.024]**
	Largest shareholder (b)	0.000 [0.001]	0.000 [0.000]	0.001 [0.001]	0.000 [0.000]	0.001 [0.000]	0.001 [0.001]
Quality, innovation and labor skills	Dummy for joint venture	-0.115 [0.030]***	-0.046 [0.051]	-0.099 [0.030]***	-0.057 [0.025]**	-0.08 [0.033]**	-0.08 [0.035]**
	Dummy for obtained a new product licensing agreement	0.042 [0.018]**	0.032 [0.029]	0.016 [0.021]	0.031 [0.017]*	0.009 [0.021]	0.034 [0.012]**
	Dummy for brought in house	0.068 [0.031]**	0.046 [0.024]*	0.035 [0.033]	0.026 [0.029]	0.036 [0.025]	0.015 [0.034]
	Staff skilled workers (b)	0.000 [0.000]	0.000 [0.000]	0.000 [0.000]*	0.000 [0.000]***	0.001 [0.000]*	0.001 [0.000]*
Other control variables	Percentage of margin	0.009 [0.001]***	0.009 [0.001]***	0.009 [0.001]***	0.008 [0.001]***	0.008 [0.001]***	0.008 [0.001]***
	Sector/región/size dummies	Yes	Yes	Yes	Yes	Yes	Yes
	Constant	Yes	Yes	Yes	Yes	Yes	Yes
	Observations	476	476	476	476	476	476
	R-squared	0.75	0.86	0.99	0.99	0.99	0.99

Notes:
See footnotes in Table 5.
Source: authors' estimations with BEEPS data

Table 8: Robust IC elasticities and semi-elasticities with respect to TFP in Portugal

		Restricted estimation			Unrestricted estimation		
		Single step	Two steps		Single step	Two steps	
		Solow residual	Cobb-Douglas	Solow residual	Cobb-Douglas	Solow residual	Cobb-Douglas
IC blocks	**IC variables**	**Coef S.E**	**Coef S.E**	**Coef S.E**	**Coef S.E**	**Coef S.E**	**Coef S.E**
Average real wage		0.426 [0.042]***	0.437 [0.054]***	0.476 [0.064]***	0.439 [0.065]***	0.427 [0.062]***	0.41 [0.068]***
Infrastructures	Shipment losses, domestic (b)	-0.001 [0.001]	-0.001 [0.001]	-0.002 [0.001]	-0.001 [0.001]	0 [0.001]	-0.002 [0.002]
Bureaucracy and others	Number of inspections (b)	-0.062 [0.027]**	-0.02 [0.021]	-0.052 [0.024]**	-0.029 [0.014]*	-0.029 [0.013]**	-0.024 [0.014]*
	Time to fill a vacancy for prof	-0.119 [0.062]*	-0.037 [0.027]	-0.058 [0.044]	-0.056 [0.035]	-0.049 [0.032]	-0.043 [0.037]
	Payments to obtain a contract with the government (b)	-0.013 [0.021]	-0.014 [0.013]	-0.013 [0.018]	-0.014 [0.011]	-0.002 [0.012]	-0.012 [0.009]
Finance	Sales paid after delivery (b)	0.001 [0.001]	0.001 [0.000]	0.001 [0.001]	0.000 [0.000]	0.001 [0.001]	0.000 [0.001]
	Working capital financed by local private banks (b)	0.001 [0.001]	0.001 [0.001]	0.001 [0.001]	0.000 [0.000]	0.000 [0.001]	0.000 [0.001]
	Working capital financed by equity (b)	-0.001 [0.001]	-0.002 [0.001]*	-0.001 [0.001]	0.000 [0.001]	-0.001 [0.001]*	-0.001 [0.001]
	Working capital financed by family/friends (b)	-0.002 [0.001]*	-0.001 [0.001]	-0.002 [0.001]***	-0.001 [0.001]	-0.001 [0.001]	-0.001 [0.001]
	Dummy for loan with collateral	0.127 [0.048]**	0.111 [0.042]**	0.071 [0.040]*	0.104 [0.036]***	0.075 [0.037]*	0.091 [0.031]***
	Interest rate of loan (b)	-0.028 [0.010]**	-0.024 [0.006]***	-0.02 [0.009]**	-0.015 [0.006]**	-0.018 [0.004]***	-0.013 [0.007]*
	Loan long term (b)	0.081 [0.043]*	0.066 [0.043]	0.085 [0.043]*	0.068 [0.043]	0.071 [0.033]**	0.042 [0.040]
	Delay to obtain a loan	-0.054 [0.017]***	-0.032 [0.015]**	-0.051 [0.018]**	-0.037 [0.016]**	-0.031 [0.013]**	-0.027 [0.015]*
	Largest shareholder (b)	0.001 [0.001]	0.000 [0.000]	0.001 [0.000]*	0.000 [0.000]	0.001 [0.000]	0.000 [0.000]
	Percentage of profits reinvested (b)	0.001 [0.001]	0.000 [0.000]	0.001 [0.000]	0.000 [0.000]	0.000 [0.000]	0.000 [0.000]
Quality, innovation and labor skills	Dummy for upgrade an existing product	0.039 [0.043]	0.056 [0.030]*	0.026 [0.031]	0.04 [0.029]	0.052 [0.025]*	0.079 [0.021]***
	Dummy for new technology embodied in new machinery or	0.134 [0.072]*	0.033 [0.042]	0.097 [0.027]***	0.045 [0.026]	0.004 [0.030]	-0.013 [0.045]
	Dummy for discontinued at least one product	-0.139 [0.064]**	-0.115 [0.042]**	-0.096 [0.067]	-0.115 [0.054]**	-0.093 [0.047]*	-0.105 [0.046]**
	Staff unskilled workers (b)	-0.001 [0.001]	-0.001 [0.001]	-0.001 [0.001]	-0.001 [0.001]	-0.001 [0.001]	-0.001 [0.000]*
Other control variables	Dummy for FDI	0.11 [0.043]**	0.075 [0.048]	0.093 [0.041]**	0.052 [0.036]	0.044 [0.029]	0.052 [0.035]
	Percentage of margin (b)	0.005 [0.001]***	0.005 [0.001]***	0.006 [0.001]***	0.006 [0.001]***	0.005 [0.001]***	0.005 [0.002]**
	Dummy for merged with another firm (b)	0.129 [0.062]*	0.066 [0.060]	0.116 [0.076]	0.11 [0.052]*	0.079 [0.070]	0.104 [0.058]*
	Losses due to civil protest	-0.026 [0.009]**	-0.023 [0.005]***	-0.05 [0.010]***	-0.036 [0.007]***	-0.022 [0.008]**	-0.019 [0.009]**
	Capacity utilization (b)	-0.003 [0.002]	-0.003 [0.001]**	-0.004 [0.002]	-0.003 [0.001]**	-0.004 [0.001]**	-0.003 [0.001]**
	Sector/región/size dummies	Yes	Yes	Yes	Yes	Yes	Yes
	Constant	Yes	Yes	Yes	Yes	Yes	Yes
	Observations	409	409	409	409	409	409
	R-squared	0.73	0.85	0.99	0.99	0.99	0.99

Notes:
See footnotes in Table 5.
Source: authors' estimations with BEEPS data

Table 9: Robust IC elasticities and semi-elasticities with respect to TFP in Ireland

		Restricted estimation			Unrestricted estimation		
		Single step	Two steps		Single step	Two steps	
		Solow residual	Cobb-Douglas	Solow residual	Cobb-Douglas	Solow residual	Cobb-Douglas
IC blocks	**IC variables**	**Coef S.E**	**Coef S.E**	**Coef S.E**	**Coef S.E**	**Coef S.E**	**Coef S.E**
Average real wage		0.299 [0.076]***	0.225 [0.055]***	0.142 [0.077]*	0.136 [0.065]**	0.126 [0.071]*	0.191 [0.068]***
Infrastructures	Shipment losses, domestic (b)	-0.003 [0.001]**	-0.003 [0.002]	-0.003 [0.001]**	-0.003 [0.001]***	-0.001 [0.002]	-0.004 [0.002]**
Bureaucracy and others	Overdue payments (b)	-0.002 [0.001]	-0.002 [0.001]	-0.002 [0.001]**	-0.001 [0.001]*	-0.001 [0.001]	-0.001 [0.001]
	Cases of arbitrage or comm. Intermediation as plaintiff	-0.074 [0.038]*	-0.076 [0.040]*	-0.075 [0.046]	-0.044 [0.049]	-0.054 [0.051]	-0.041 [0.053]
	Dummy for consultant to deal with public officials (b)	0.066 [0.065]	0.065 [0.069]	0.042 [0.056]	0.08 [0.052]	0.016 [0.070]	0.075 [0.049]
	Sales unreported to taxes (b)	-0.005 [0.002]***	-0.005 [0.001]***	-0.005 [0.001]***	-0.002 [0.001]*	-0.003 [0.001]**	-0.001 [0.001]
Finance	Sales paid before delivery (b)	0.002 [0.002]	0.003 [0.002]	0.002 [0.001]*	0.001 [0.001]	0.001 [0.001]	0.001 [0.001]
	Working capital financed by local private banks (b)	0.002 [0.002]	0.001 [0.002]	0.002 [0.001]	0.001 [0.001]	0.001 [0.001]	0.001 [0.001]
	Dummy for trade association	0.016 [0.056]	0.001 [0.058]	0.039 [0.049]	0.065 [0.030]**	0.076 [0.042]*	0.074 [0.027]**
	Dummy for savings account	0.095 [0.084]	0.098 [0.079]	0.084 [0.076]	0.059 [0.045]	0.121 [0.073]	0.087 [0.054]
	Value of the collateral (b)	-0.0004 [0.000]	-0.0004 [0.000]	-0.0004 [0.000]	-0.0004 [0.000]	-0.0004 [0.000]	-0.0004 [0.000]
Quality, innovation and labor skills	Dummy for outsourcing	0.135 [0.100]	0.14 [0.097]	0.127 [0.107]	0.192 [0.087]**	0.151 [0.088]*	0.187 [0.083]**
Other control variables	Dummy for holding (b)	0.243 [0.128]*	0.233 [0.120]*	0.241 [0.124]*	0.166 [0.065]**	0.183 [0.091]*	0.14 [0.060]**
	Number of competitors in local market	0.034 [0.026]	0.038 [0.025]	0.035 [0.024]	0.025 [0.025]	0.04 [0.019]**	0.03 [0.020]
	Percentage of margin	0.002 [0.000]***	0.002 [0.000]***	0.003 [0.000]***	0.003 [0.000]***	0.003 [0.000]***	0.003 [0.000]***
	Dummy for merged with another firm (b)	0.127 [0.057]**	0.115 [0.065]*	0.111 [0.070]	0.197 [0.055]***	0.124 [0.130]	0.254 [0.092]**
	Sector/region/size dummies	Yes	Yes	Yes	Yes	Yes	Yes
	Constant	Yes	Yes	Yes	Yes	Yes	Yes
	Observations	461	461	461	461	461	461
	R-squared	0.318	0.374	0.923	0.954	0.937	0.97

Notes:
See footnotes in Table 5.
Source: authors' estimations with BEEPS data

Table 10: Robust IC elasticities and semi-elasticities with respect to TFP in Korea

		Restricted estimation			Unrestricted estimation		
		Single step	Two steps		Single step	Two steps	
		Solow residual	Cobb-Douglas	Solow residual	Cobb-Douglas	Solow residual	Cobb-Douglas
IC blocks	**IC variables**	**Coef S.E**	**Coef S.E**	**Coef S.E**	**Coef S.E**	**Coef S.E**	**Coef S.E**
Average real wage		0.348 [0.025]***	0.352 [0.039]***	0.361 [0.034]***	0.351 [0.033]***	0.347 [0.038]***	0.355 [0.037]***
Bureaucracy and others	Cases of arbitrage or comm. Intermediation as defendant	-0.053 [0.036]	-0.044 [0.036]	-0.052 [0.035]	-0.027 [0.034]	-0.032 [0.042]	-0.027 [0.049]
	Payments to deal with bur. issues (b)	-0.075 [0.064]	-0.11 [0.082]	-0.053 [0.036]	-0.048 [0.031]	-0.063 [0.026]**	-0.072 [0.027]**
	Dummy taking 1 if the firm would increase its current level of employees (b)	-0.076 [0.025]***	-0.021 [0.082]	-0.033 [0.036]	-0.064 [0.028]**	0.001 [0.042]	-0.073 [0.052]
	Workforce unreported to taxes (b)	-0.002 [0.001]**	-0.002 [0.001]*	-0.001 [0.001]	-0.001 [0.001]	-0.001 [0.000]	-0.001 [0.001]*
Finance	Working capital financed by internal founds (b)	-0.001 [0.001]*	-0.001[0.001]**	-0.001[0.001]*	-0.001[0.000]*	-0.001[0.000]*	-0.000[0.000]
	Working capital financed by foreign private banks (b)	0.006 [0.001]***	0.006 [0.002]***	0.005 [0.002]***	0.004 [0.002]*	0.004 [0.002]*	0.007 [0.002]**
	Value of the collateral (b)	-0.000 [0.000]	-0.000 [0.000]	-0.000 [0.000]	-0.000 [0.000]	-0.000 [0.000]	-0.000 [0.000]
	Dummy for external auditory (b)	0.034 [0.035]	0.029 [0.035]	0.046 [0.026]*	0.061 [0.024]**	0.041 [0.023]*	0.054 [0.017]***
Quality, innovation and labor skills	Dummy for R&D expenditures	0.056 [0.033]	0.053 [0.050]	0.035 [0.037]	0.045 [0.042]	0.029 [0.038]	0.018 [0.035]
	Dummy for advertising expenditures	0.043 [0.028]	0.049 [0.030]	0.054 [0.029]*	0.044 [0.027]	0.042 [0.024]*	0.041 [0.017]**
	Dummy for upgrade an existing product (b)	0.036 [0.034]	0.06 [0.037]	0.023 [0.032]	0.03 [0.032]	0.036 [0.036]	0.009 [0.023]
Other control variables	Dummy for incorporated company	0.057 [0.054]	0.067[0.054]	0.036[0.046]	0.017[0.044]	0.065[0.036]*	0.034[0.044]
	Dummy for FDI	0.188 [0.111]	0.176 [0.115]	0.153 [0.103]	0.193 [0.104]*	0.208 [0.108]*	0.205 [0.131]
	Dummy for importer	0.122 [0.053]**	0.101 [0.055]*	0.112 [0.057]*	0.091 [0.042]**	0.072 [0.031]**	0.057 [0.031]*
	Percentage of margin	0.005 [0.002]***	0.005 [0.001]***	0.006 [0.001]***	0.006 [0.002]***	0.006 [0.001]***	0.007 [0.001]***
	Sector/region/size dummies	Yes	Yes	Yes	Yes	Yes	Yes
	Constant	Yes	Yes	Yes	Yes	Yes	Yes
	Observations	487	487	487	487	487	487
	R-squared	0.70	0.77	0.99	0.99	0.99	0.99

Notes:
See footnotes in Table 5.
Source: authors' estimations with BEEPS data

Table 11. Percentage contributions of IC variables to aggregate TFP in Spain and peer countries

[Percentage contributions to average TFP in brackets]

	Spain	Germany	Greece	Portugal	Ireland	Korea
	Infrastructures					
Power outages	. [.]	-2.04 [-2.3]	. [.]	. [.]	. [.]	. [.]
Days to clear customs to export (interaction with dummy for exports))	-16.23 [-4.66]	. [.]	-2.96 [-0.65]	. [.]	. [.]	. [.]
Dummy for e-mail	46.67 [34.96]	. [.]	30.24 [16.53]	. [.]	. [.]	. [.]
Shipment losses, domestic	-0.86 [-0.35]	. [.]	. [.]	-3.17 [-2.26]	-0.84 [-1.1]	. [.]
Avg. duration Of phone outages	-2.52 [-1.8]	. [.]	. [.]	. [.]	. [.]	. [.]
	Bureaucracy and others					
Cases of arbitrage or comm. Intermediation as defendant	. [.]	. [.]	. [.]	. [.]	. [.]	-6.72 [-1.02]
Cases of arbitrage or comm. Intermediation as plaintiff	-3.99 [-2]	. [.]	-11.44 [-3.28]	. [.]	-29.76 [-3.38]	. [.]
Dummy for having overdue payments	. [.]	-0.65 [-1.58]	. [.]	. [.]	. [.]	. [.]
Overdue payments	. [.]	. [.]	. [.]	. [.]	-24.79 [-5.81]	. [.]
Payments to deal with bur. Issues	. [.]	. [.]	. [.]	. [.]	. [.]	-1.2 [-1.13]
Payments to obtain a contract with the government	. [.]	. [.]	. [.]	-5.21 [-11.32]	. [.]	. [.]
Manager's time spent in bur. Issues	. [.]	. [.]	-5.24 [-2.59]	. [.]	. [.]	. [.]
Dummy for consultant to deal with public officials	. [.]	. [.]	. [.]	. [.]	12.7 [4.34]	. [.]
Number of inspections	. [.]	. [.]	. [.]	-115.05 [-34.98]	. [.]	. [.]
Sales reported to taxes	-1.59 [-2.29]	. [.]	. [.]	. [.]	-1.72 [-6.29]	. [.]
Workforce reported to taxes	. [.]	. [.]	-1.1 [-3.05]	. [.]	. [.]	-0.51 [-4.51]
Time to find a professional worker	. [.]	. [.]	1.22 [1.07]	-169.64 [-156.12]	. [.]	. [.]
Dummy firm would increase its current level of employees	. [.]	. [.]	. [.]	. [.]	. [.]	-0.24 [-0.75]
Dummy firm would reduce its current level of employees	-8.07 [-2.35]	. [.]	. [.]	. [.]	. [.]	. [.]
Security expenses	. [.]	-5.76 [-6.72]	. [.]	. [.]	. [.]	. [.]
Dummy for national laws or regulation affecting	-5.22 [-1.6]	. [.]	. [.]	. [.]	. [.]	. [.]

"." Non-significant variable and consequently the contribution of the variable is equal to zero.
Results from equation (8) of section 4, where the total contribution of IC variables to *demeaned* aggregate TFP is normalized to be 100%.
Source: authors' estimations with BEEPS data.

Table 11 (cont.). Percentage contributions of IC variables to aggregate TFP in Spain and peer countries

[Percentage contributions to average TFP in brackets]

	Spain	Germany	Greece	Portugal	Ireland	Korea
	Finance					
Working capital financed by internal founds	-9.78 [-11.13]	. [.]	-10.48 [-20.76]	. [.]	. [.]	-28.16 [-30.0]
Working capital financed by equity	-1.9 [-1.39]	. [.]	-0.8 [-1.03]	-6.11 [-0.83]	. [.]	. [.]
Working capital financed by local private banks	. [.]	9.68 [4.74]	. [.]	16.14 [6.56]	4.33 [10.44]	. [.]
Working capital financed by foreign private banks	. [.]	. [.]	4.71 [0.69]	. [.]	. [.]	2.44 [1.44]
Working capital financed by informal founds	-0.33 [-0.57]	. [.]	. [.]	. [.]	. [.]	. [.]
Working capital financed by leasing	0.01 [0.16]	. [.]	. [.]	. [.]	. [.]	. [.]
Working capital financed by family/friends	. [.]	. [.]	-0.04 [-0.26]	-0.28 [-3.36]	. [.]	. [.]
New fixed assets financed by government	. [.]	-0.44 [-0.33]	. [.]	. [.]	. [.]	. [.]
Dummy for loan with collateral	7.4 [12.56]	. [.]	. [.]	24.48 [23.67]	. [.]	. [.]
Interest rate of loan	-8.93 [-17.69]	. [.]	. [.]	-24.88 [-29.7]	. [.]	. [.]
Duration of the loan in months	. [.]	. [.]	. [.]	29.28 [18.02]	. [.]	. [.]
Time to obtain a loan	. [.]	. [.]	. [.]	-96.07 [-38.48]	. [.]	. [.]
Value of the collateral	. [.]	. [.]	. [.]	. [.]	-7.11 [-6]	-4.79 [-4.45]
Dummy for external auditory	. [.]	. [.]	20.74 [10.23]	. [.]	. [.]	10.16 [3.92]
Dummy for international accounting standards	. [.]	. [.]	1.79 [1.68]	. [.]	. [.]	. [.]
Sales paid after delivery	. [.]	. [.]	. [.]	48.43 [35.37]	. [.]	. [.]
Sales paid before delivery	4.39 [4.35]	. [.]	. [.]	. [.]	3.58 [4.86]	. [.]
Percentage of profits reinvested	. [.]	5.07 [4.37]	. [.]	56.42 [70.81]	. [.]	. [.]
Largest shareholder	. [.]	. [.]	10.05 [13.73]	84.26 [87.67]	. [.]	. [.]
Dummy for trade association	11.99 [8.93]	. [.]	. [.]	. [.]	4.7 [3.1]	. [.]
Dummy for savings account	. [.]	. [.]	. [.]	. [.]	21.43 [13.33]	. [.]

"." Non-significant variable and consequently the contribution of the variable is equal to zero.
Results from equation (8) of section 4, where the total contribution of IC variables to *demeaned* aggregate TFP is normalized to be 100%.
Source: authors' estimations with BEEPS data.

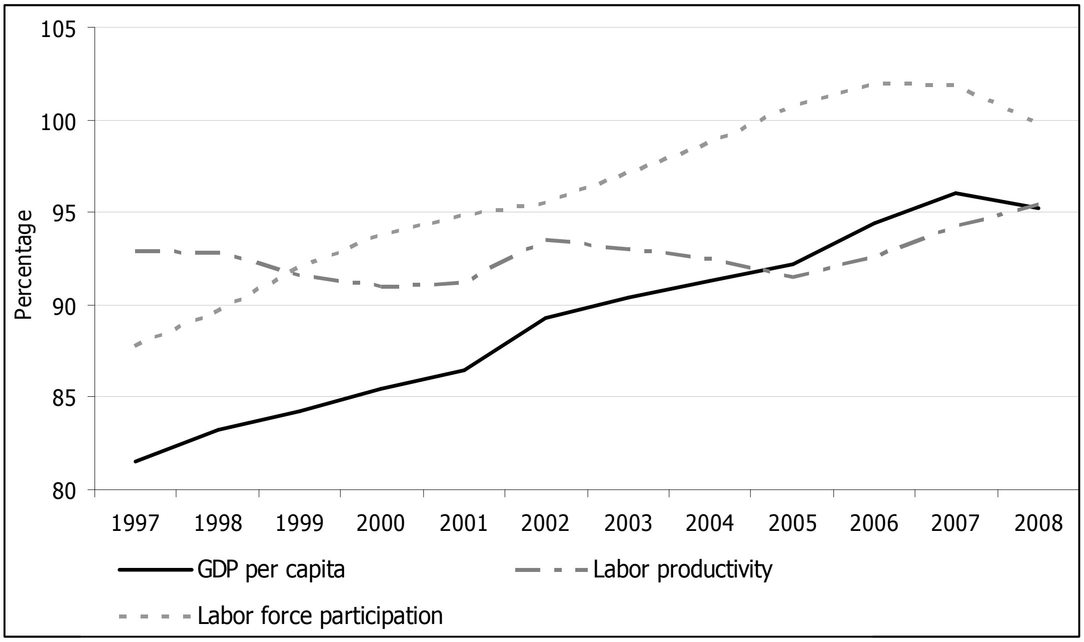
Table 11 (cont.). Percentage contributions of IC variables to aggregate TFP in Spain and peer countries

[Percentage contributions to average TFP in brackets]

	Spain	Germany	Greece	Portugal	Ireland	Korea
Quality, innovation and labor skills						
Dummy for R&D expenditures	22.73 [5.73]	12.28 [3.55]	. [.]	. [.]	. [.]	11.35 [3.65]
Dummy for advertising expenditures	. [.]	. [.]	. [.]	. [.]	. [.]	11.87 [6.25]
Dummy for upgrade an existing product	. [.]	. [.]	. [.]	30.46 [9.07]	. [.]	9.49 [5.95]
Dummy for discontinued at least one product	. [.]	. [.]	. [.]	-72.92 [-9.79]	. [.]	. [.]
Dummy for obtained a new product licensing agreement	. [.]	. [.]	0.98 [1.1]	. [.]	. [.]	. [.]
Dummy for new product technology	. [.]	1.69 [1.31]	. [.]	. [.]	. [.]	. [.]
Dummy for new tech embodied in new mach. or equip	. [.]	. [.]	. [.]	55.44 [23.68]	. [.]	. [.]
Dummy for outsourcing	. [.]	. [.]	. [.]	. [.]	18.15 [4.83]	. [.]
Dummy for brought in house	-1.78 [-0.99]	. [.]	2.14 [0.95]	. [.]	. [.]	. [.]
Dummy for joint venture	. [.]	. [.]	-8.02 [-3.33]	. [.]	. [.]	. [.]
University education	3.34 [2.86]	1.79 [2.18]	. [.]	. [.]	. [.]	. [.]
Staff skilled workers	. [.]	. [.]	2.73 [3.04]	. [.]	. [.]	. [.]
Staff unskilled workers	. [.]	. [.]	. [.]	-19.62 [-18.55]	. [.]	. [.]
Other control variables						
Dummy for holding	. [.]	. [.]	. [.]	. [.]	52.12 [11.37]	. [.]
Dummy for incorporated company	15.41 [10.62]	. [.]	. [.]	. [.]	. [.]	16.45 [7.62]
Number of competitors in the local market	. [.]	. [.]	. [.]	. [.]	30.48 [32.35]	. [.]
Dummy for importer	. [.]	. [.]	. [.]	. [.]	. [.]	32.26 [8.46]
Share of exports	34.15 [11.44]	. [.]	. [.]	. [.]	. [.]	. [.]
Dummy for FDI	. [.]	. [.]	. [.]	58.82 [14.31]	. [.]	25.33 [5.66]
Dummy for merge with other firm	-2.45 [-0.4]	. [.]	. [.]	61.24 [6.3]	3.03 [1.31]	. [.]
Dummy for acquired other firm	17.53 [3.29]	. [.]	. [.]	. [.]	. [.]	. [.]
Losses due to civil protests	. [.]	. [.]	. [.]	-0.01 [-0.38]	. [.]	. [.]
Margin over costs	. [.]	78.38 [69.78]	65.48 [71.18]	147.99 [138.45]	13.7 [21.37]	22.26 [32.52]
Total of all IC blocks of variables	**100 [47.67]**	**100 [74.98]**	**100 [85.24]**	**100 [128.14]**	**100 [84.71]**	**100 [33.6]**

"." Non-significant variable and consequently the contribution of the variable is equal to zero.
Results from equation (8) of section 4, where the total contribution of IC variables to *demeaned* aggregate TFP is normalized to be 100%.
Source: authors' estimations with BEEPS data.

Figure 1: Decomposition of Spanish GDP per capita with respect to EU (15) countries



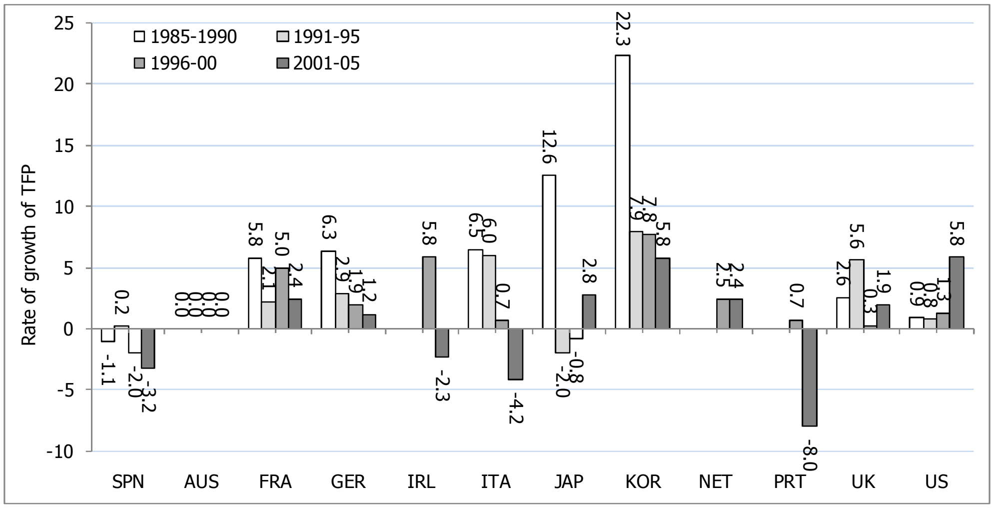
Notes:

a) European Union includes: Austria, Belgium, Denmark, Finland, France, Germany, United Kingdom, Greece, Ireland, Italy, Luxemburg, Netherlands, Spain, Sweden and Portugal.

b) Per capita income (Y/P) is decomposed into labor productivity (Y/L) and the employment-population rate (L/P) by following the next expression: (Y/P)= (Y/L)*(L/P); relative to the EU(15) the expression becomes: (YUS/PUS)]= [(YSP/LSP)/(YEU/LEU)]*[(LSP/PSP)/ (LEU/PEU)]

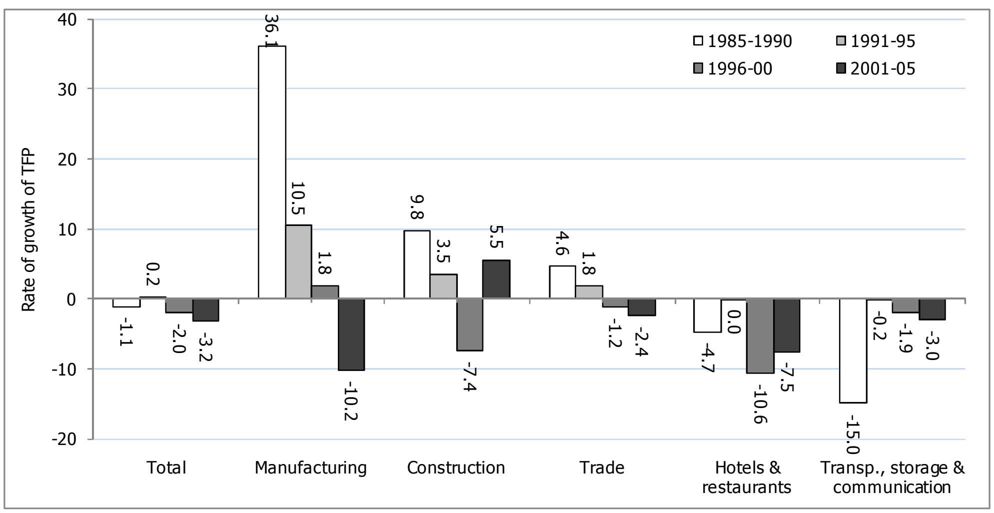
Source: Authors' elaboration with data from Eurostat.

Figure 2: Rates of growth of TFP in Spain and peer countries, 1985-2005



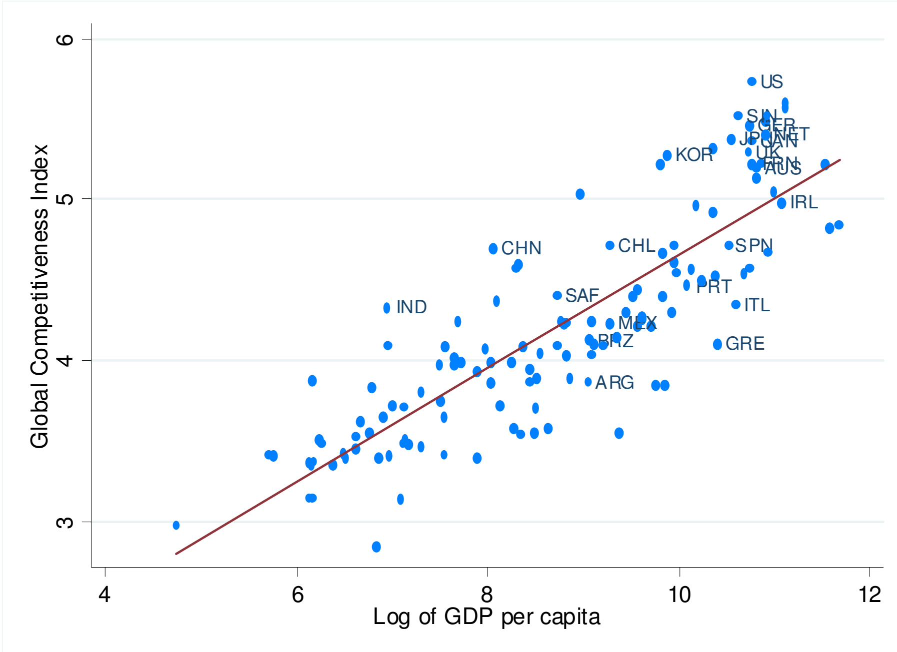
Notes:
TFP based on gross output.
Source: Authors' elaboration with data from EU-KLEMS database, 2008, Groningen Growth and Development Centre, University of Groningen

Figure 3. Rates of growth of TFP in Spain by activity sectors, 1985-2005



Notes:
TFP based on gross output.
Source: Authors' elaboration with data from EU-KLEMS database, 2008, Groningen Growth and Development Centre, University of Groningen

Figure 4. Cross-plot between Global Competitiveness Index (2009) and GDP per capita

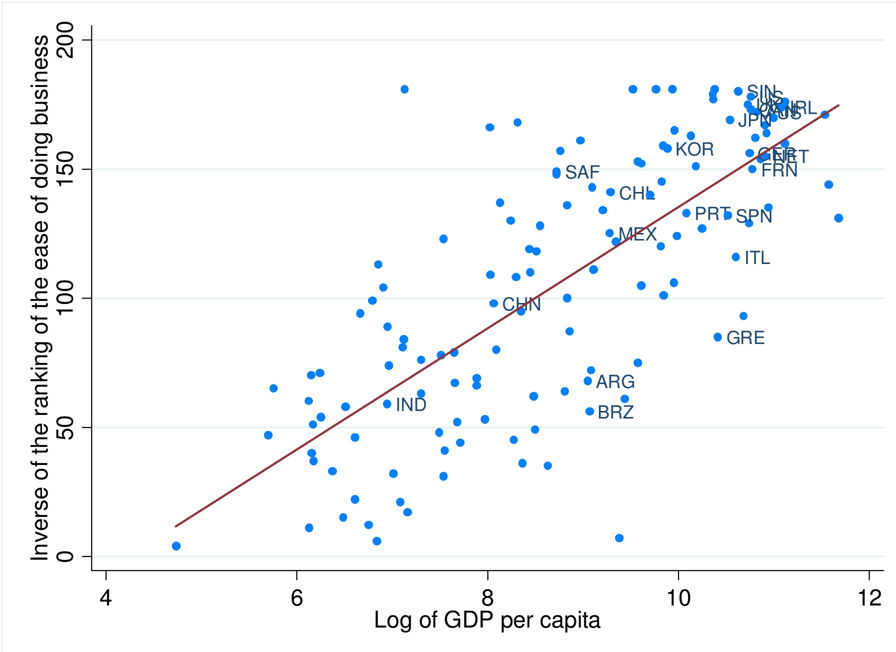


Notes:

The global Competitiveness Index is computed as a weighted average of the 12 fundamental pillars for competitiveness. The stage of development of each economy is taken into account when computing the weights of each pillar.

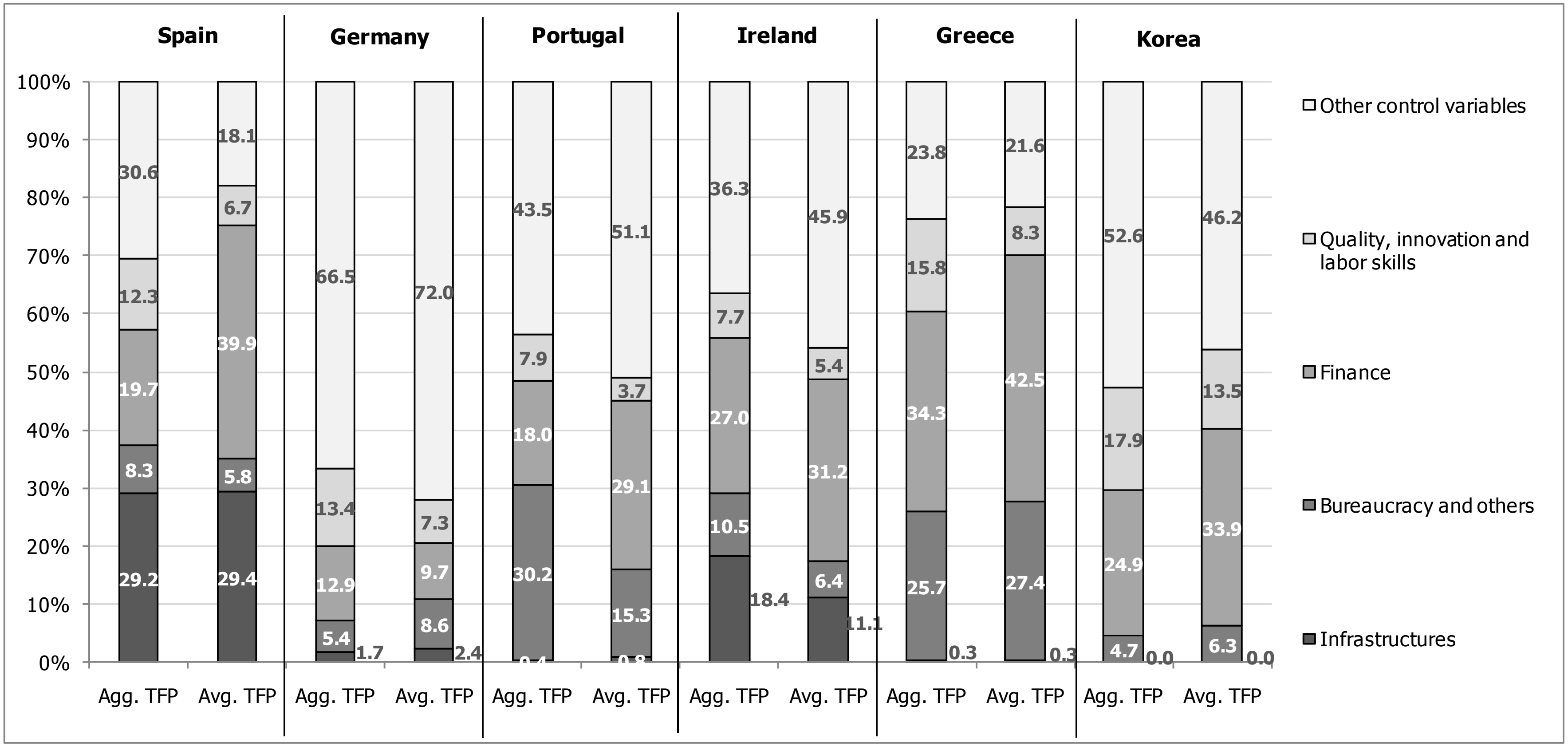
Source: Authors' elaboration with data from the World Economic Outlook 2009, IMF; y Global Competitiveness Report 2008, The World Economic Forum.

Figure 5. Cross-plot between the (inverse) rankings of the ease of doing business (2008) and GDP per capita



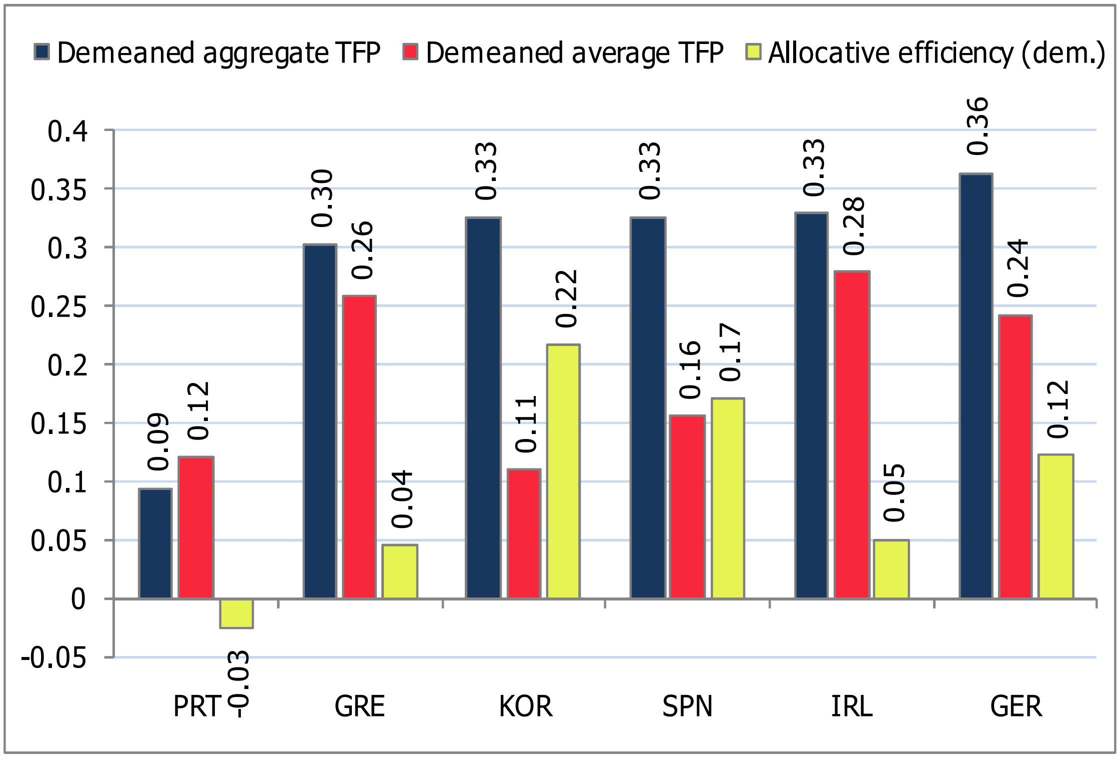
Notes:
The ranking of the ease of doing business is the result of a weighted average of each one of the rankings of the basic aspects of doing business. In the vertical axis it is the inverse of the ranking, say number of countries (181) minus the ranking of each country. The higher the inverse the easier it is doing business.
Source: Authors' elaboration with data from the World Economic Outlook 2009, IMF and Doing Business Report 2009, The World Bank Group, Washington, DC.

Figure 6. Percentage absolute contributions to *demeaned* aggregate and average TFPs by groups of IC variables in Spain and peer countries



Results from equation (8) of section 4 of the main text, IC percentage contributions to demeaned TP in absolute values are accumulated by groups of variables.
Source: authors' estimations with BEEPS data.

Figure 7. Olley and Pakes decompositions of demeaned aggregate TFP in Spain and peer countries



Results from equation (8) of section 4.
Source: authors' estimation with BEEPS data.

CHAPTER II

ASSESSING INFRASTRUCTURE QUALITY ON FIRM PRODUCTIVITY IN AFRICA

Assessing the Impact of Infrastructure Quality on Firm Productivity in Africa*

Cross-Country Comparisons Based on Investment Climate Surveys from 1999 to 2005

by

Alvaro Escribano,[†] J. Luis Guasch[‡] and Jorge Pena[§]

February, 2009

Abstract

This paper provides a systematic, empirical assessment of the impact of infrastructure quality on the total factor productivity (TFP) of African manufacturing firms. This measure is understood to include quality in the provision of customs clearance, energy, water, sanitation, transportation, telecommunications, and information and communications technology (ICT). We apply microeconometric techniques to investment climate surveys (ICSs) of 26 African countries carried out in different years during the period 2002–6, making country-specific evaluations of the impact of investment climate (IC) quality on aggregate TFP, average TFP, and allocative efficiency. For each country we evaluated this impact based on 10 different productivity measures. Results are robust once we control for observable fixed effects (red tape, corruption and crime, finance, innovation and labor skills, etc.) obtained from the ICSs. We ranked African countries according to several indices: per capita income, ease of doing business, firm perceptions of growth bottlenecks, and the concept of demeaned productivity (Olley and Pakes 1996). We divided countries into two blocks: high-income-growth and low-income-growth. Infrastructure quality has a low impact on TFP in countries of the first block and a high (negative) impact in countries of the second. We found heterogeneity in the individual infrastructure elements affecting countries from both blocks. Poor-quality electricity provision affects mainly poor countries, whereas problems dealing with customs while importing or exporting affects mainly faster-growing countries. Losses from transport interruptions affect mainly slower-growing countries. Water outages affect mainly slower-growing countries. There is also some heterogeneity among countries in the infrastructure determinants of the allocative efficiency of African firms.

Key words: Africa, Infrastructure, Total Factor Productivity, Investment Climate, Competitiveness.

---

* This study is part of the Africa Infrastructure Country Diagnostic (AICD), a project designed to expand the world's knowledge of physical infrastructure in Africa. Financing for AICD is provided by a multi-donor trust fund to which the main contributors are the Department for International Development (United Kingdom), the Public Private Infrastructure Advisory Facility, Agence Française de Développement, and the European Commission. Inquiries concerning the availability of datasets should be directed to vfoster@worldbank.org. The authors are indebted to Vivien Foster and her group of researchers at the World Bank for providing data sets and further insight into the African region. All errors in this report are the sole responsibility of the authors. A version of this paper is available in the Policy Research Working Paper Series of the World Bank.

[†] Telefónica-UC3M chair in the economics of telecommunications, Department of Economics, Universidad Carlos III de Madrid; alvaroe@eco.uc3m.es

[‡]The World Bank and the University of California, San Diego; jguasch@worldbank.org.

[§] Department of Economics, Universidad Carlos III de Madrid; jpizquie@eco.uc3m.es.

1. The wide-ranging effects of infrastructure quality

For Africa's awaited growth resurgence to occur, a broad range of factors—political, institutional, and economic—must be improved. The World Bank's landmark Africa Competitiveness Report (ACR) (2004 and 2007) focuses on problems that, in the words of Artadi and Sala-i-Martin (2003), constitute the most important growth tragedy of the twentieth century—a phenomenon that has received special attention in recent growth literature, such as that of Ndulu and O'Connell (2005). It is agreed that improving Africa's infrastructure is a crucial step toward penetrating international markets and meeting the goals of continuous growth and poverty reduction.

Infrastructure quality has a pervasive influence on all areas of an economy. Low-quality infrastructure and limited transport and trade services increase logistical and transaction costs, rendering otherwise competitive products uncompetitive, as well as limiting rural production and people's access to markets—with adverse effects on economic activity and poverty reduction. A large number of empirical studies illustrate the impact of infrastructure on economic performance, including those of Calderón et al. (2003a and b), Calderón and Serven (2003), Canning (1998), Reinikka and Svensson (1999), Prud'homme (2004), Escribano and Guasch (2005), Escribano et al. (2005), and Guasch (2004). All suggest that Africa's infrastructure gap is an important growth bottleneck with a negative impact on productivity and the overall competitiveness of the region. Furthermore, several studies using the methodology of Escribano and Guasch (2005, 2008) and Escribano et al. (2008a and b and 2009) have found empirical evidence—in cases such as Brazil, Chile, Costa Rica, Mexico, Turkey, and several southeast Asian countries—that improvements in investment climate (IC) conditions in general, and in infrastructure quality in particular, may lead to important gains in productivity and in other economic performance measures: employment, real wages, exporting activities, and foreign direct investment (FDI) inflows.

Disentangling the ways that infrastructure affects Africa's economic growth poses several difficulties because of the special characteristics of the African region. The comprehensive analysis found in Estache (2005) takes stock of the basic characteristics of infrastructure in Sub-Saharan Africa and the impact of 1990 reforms, pointing out that the impact of infrastructure in Africa may be different than in other regions. As Brunel (2004) signals, the colonial period has had a lasting effect on the use of space in the region, resulting in a productive structure that consists, in most cases, of coastal cities connected inland by railways designed to carry raw materials to main ports. This and other factors that are progressively modifying the continent's productive structure—such as continuous urbanization, the movement of economic activity from the agricultural to manufacturing and service sectors, and the increasing openness of African economies—has caused both a quantitative and qualitative mismatch between the current supply of infrastructure and ever-increasing demand. Factors such as inequality across income levels (affecting the affordability of infrastructure services), large and unoccupied areas, and regional variations in climate are increasingly becoming a concern for African policy makers managing infrastructure.

In addition to furthering the regional integration needed to support infrastructure investment, African governments made important contributions to infrastructure development in the decades following independence. The majority of African state monopolies were, however, characterized by inefficient bureaucracies. These became increasingly unable to satisfy customer demands, with increasing deficits. By the beginning of the 1980s, most African countries embarked on infrastructure sector reforms, with the aim of increasing private sector participation in provision. Despite attempts to introduce more competition and to attract private investors, Africa continues to trail the world in both the quantity and quality of its infrastructure, with bottlenecks particularly in the management of current stock.

Figure 1.1 of the appendix on tables and figures shows the geographical distribution of the countries considered in this study, both in North and Sub-Saharan Africa. The countries studied are divided into five main geographical areas, identified in some cases by the major multilateral organization of each region: (a) the North African region, or Maghreb, includes Morocco, Algeria, and Egypt; (b) the Economic Community of West African States (ECOWAS) includes Mauritania, Mali, Niger, Senegal, Burkina Faso, Benin, Cameroon, and Cape Verde; (c) the Horn of Africa region is composed of two countries, Eritrea and Ethiopia; (d) the East African Community (EAC) includes Tanzania, Kenya, Uganda, and Burundi; and (e) the South African Development Community (SADC), for which we have data for Malawi, Zambia, Namibia, Botswana, Swaziland, Lesotho, and Madagascar. South Africa and Mauritius are the last two individual countries included in the report.

[FIGURE 1.1 ABOUT HERE]

The 26 countries show enormous heterogeneity due to (a) geographical factors, such as whether a nation is landlocked (Cape Verde, Madagascar, and Mauritius), tropical (with landmass for the most part covered by rainforests), or dominated by deserts (such as the North African countries Mauritania and Namibia); (b) social or political factors, such as civil wars, armed conflicts, early democracies, dictatorships, and colonial heritage; and (c) economic factors, which this paper discusses for all countries, from the most affluent (Mauritius) to the poorest (Eritrea).

Figure 1.2 clarifies the different evolutions of per capita income across the countries included in this analysis. Out of the 26 African countries analyzed, Mauritius was, in 1950, the country with the highest per capita income (measured in terms of per capita gross domestic product, GDP), followed closely by South Africa, and, by a wider gap, Namibia and Algeria. But the per capita income situation in 2003 was somewhat different; Mauritius was still ranked first, followed by Swaziland, South Africa, and Botswana—and, by a wider gap, Algeria, Cape Verde, Egypt, and Morocco. Panel B of figure 1.2 shows the five-year growth rate of per capita income. Mauritius and Botswana are the countries that have experienced the highest, sustained per capita income growth during the recent years. Lesotho is the median country, splitting the cross-section into two blocks. The first block comprises countries with faster and steadier growth rates (Mauritius, Swaziland, South Africa, Botswana, Namibia, and Lesotho in the south; Algeria, Morocco, and Egypt in the north; and Cape Verde and Cameroon in central Africa). In the second are countries with lower and more irregular growth rates (Mauritania, Senegal, Benin, Mali, Niger, and Burkina Faso in the central west; Uganda, Kenya, Zambia, Tanzania, Malawi, Burundi, Madagascar, Ethiopia, and Eritrea from the central east), periods of positive expansion fluctuate with those of persistent reductions in per capita income.

[FIGURE 1.2 ABOUT HERE]

These per capita income rankings are correlated with the rankings obtained from the World Bank's 2007 Doing Business Report (DBR), presented in panel C of figure 1.2. In 2007 Mauritius, Swaziland, South Africa, Botswana, and Namibia rank 32nd, 76th, 29th, 48th, and 42nd in the world based on the ease-of-doing-business indicators. This index considers questions such as the number of days required to start a business and the ease of dealing with licenses, registering a property, trading across borders, employing workers, and so on. Other 2007 rankings include 83rd for Kenya, 97th for Ethiopia, 165th for Egypt, and 170th for Eritrea.

To better understand the convergence or divergence of trends, we plotted the per capita income of each African country relative to the per capita income of the United States (see panel A of Figure 1.3). Convergence is observed only in Mauritius, Swaziland, and Botswana. For all other study countries, including South Africa, per capita income is diverging from the United States, while in a few (Egypt, Morocco, and Cape Verde) the ratio has remained stable. While persistently positive GDP growth allowed Mauritius's per capita income to reach 45 percent of the United States in 2003, this is clearly the exception (together with Swaziland and Botswana). For the rest of the countries, including South Africa, relative per capita income was much lower in 2003 than in 1960 (indicating divergence). In fact, the 2003 per capita income of several countries was no larger than 5 percent of the per capita income of the United States. As expected, labor productivity is the main factor explaining this divergence in per capita income in Africa (panel B of Figure 1.3), given that labor force participation has a steady influence (panel C of Figure 1.3).[1] Since TFP is usually a key factor explaining the evolution of labor productivity, in this paper we seek to use investment climate surveys (ICSs) to identify the main infrastructure-related TFP bottlenecks in Africa.

[1] The *per capita income* of country J (Y^J/P^J) is decomposed into labor productivity (Y^J/L^J) and the employment-population rate (L^J/P^J) by following the expression: $(Y^J/P^J) = (Y^J/L^J)*(L^J/P^J)$, where Y is GDP, L is total labor force, and P is total population.

[FIGURE 1.3 ABOUT HERE]

Figure 1.4 shows the percentage of firms that *perceive* telecommunications, electricity, customs clearance, and transport as major obstacles to their economic performance. Only in Benin, Kenya, and Zambia do more than 50 percent of firms identify telecommunications as a severe obstacle. The quality of electricity provision is a major problem for more than 50 percent of firms in more than half of the countries in our sample. In Burundi, Cameroon, Benin, Burkina Faso, and Cape Verde, the percentage of firms considering electricity as a severe or very severe obstacle exceeds 80 percent; on the other hand, only 20 percent of firms in Morocco, South Africa, Botswana, and Namibia consider electricity a severe obstacle. Customs clearance is considered an acute problem in Benin, Kenya, Madagascar, Senegal, and Algeria. Finally, transportation is considered a severe obstacle by more than 70 percent of firms in Burkina Faso and Benin.

[FIGURE 1.4 ABOUT HERE]

Figure 1.5 offers another view of the state of infrastructure in Africa. The World Bank's ACR (2007) evaluates a wide range of factors related to economic activity, infrastructure among them. Once again there are clearly different performance levels across the two blocks of countries. While in Namibia, South Africa, Botswana, Egypt, and Morocco the quality of infrastructure exceeds the approval level; in the remaining countries this quality is rated low in most cases. The same holds for the disaggregated results, including the number of telephone lines and the quality of ports, air transport, and electricity supply.

[FIGURE 1.5 ABOUT HERE]

The difference between the two blocks becomes even clearer in figure 1.6, where the cross-plots between GDP per capita relative to the United States and firms' perceptions are presented. A preliminary analysis of the cross-plots suggests two points: first, that there is an intuitive and negative relation between income level and infrastructure constraints; and, second, that the diversion of the two blocks of countries remains intact, showing now the largest dispersion in the constraint perceptions of figure 1.6 for the lowest per capita income group.

[FIGURE 1.6 ABOUT HERE]

The objective of this paper is to assess the impact of the quality of existing infrastructure on the TFP of African firms. This measure is understood to include quality in the provision of the following services: customs clearance, energy, water, sanitation, transportation, telecommunications, and information and communications technology (ICT). We also want to identify infrastructure factors with statistically significant impacts on TFP, country by country. In the econometric evaluation we use 10 different measures of TFP and show that the results are robust—no matter what measure of TFP is used—if we follow the econometric methodology of Escribano and Guasch (2005, 2008), and Escribano et al. (2008).

For the empirical analysis of infrastructure's constraints on TFP, we go down to the firm level since infrastructure is one of the key elements of a country-specific IC, and a significant component of country competitiveness. To provide reliable and robust estimates of the impact of infrastructure on economic performance is not a straightforward task. As we will see later on, we have to deal with the heterogeneity of the countries included in our sample, and the endogeneity of explanatory variables (inputs and IC variables) in several dimensions due to unobservable fixed effects, measurement errors, missing observations, and so on. To solve these problems, we take advantage of the useful and rich firm-level information provided by the ICSs undertaken by the World Bank in Africa from 2002 to 2006. These surveys capture firm-level information in a range of areas related to economic performance: infrastructure, financing, governance, corruption, crime, regulation, tax policy, labor relations, conflict resolution, supplies and marketing, quality, technology, and training, among others. These surveys offer information on the production function (PF) variables over one, two, or three years, depending on the African country. But for infrastructure and other IC and plant control (C) variables they only provide information for a single year. We will see how we can use this valuable information to evaluate how firms operate in Africa and to identify the main obstacles to productivity improvements.

Section 2 of this report clarifies the link between this type of empirical work and existing literature on infrastructure and productivity. The properties and quality of the ICSs are analyzed in section 3. Why we classify the IC factors in broad categories or groups will also be discussed, together with the infrastructure variables (INFs) used. In section 4 we present the econometric methodology we use to estimate the impact of infrastructure and other IC variables and C characteristics on TFP. Once we have estimated the infrastructure and other IC elasticities and semi-elasticities on productivity, we evaluate the effects of infrastructure on aggregate productivity and on allocative efficiency, using the Olley and Pakes (O&P, 1996) decomposition. The main empirical results are described in the remaining sections. In particular, section 5 focuses on the relative importance of infrastructure in the IC of each country. Section 6 presents the empirical results country by country, and section 7 includes the main conclusions. Most of the tables and figures are included in the appendix at the end of the paper.

2. How does infrastructure quality affect economic performance?

Much literature discusses the different ways that infrastructure affects growth and other development outcomes at the macroeconomic level. For example, the World Bank's landmark

World Development Report (1994) highlighted multiple links between infrastructure and development and emphasized how policy can improve not only the quantity, but also the quality, of infrastructure services in developing countries.

As Straub (2008) signals, macrolevel literature has too often sought to obtain the elasticity of infrastructure capital and compare it with the elasticity of private capital. Few papers go beyond measures of infrastructure spending and infrastructure stocks to consider the issue of infrastructure efficiency. Since the seminal paper of Aschauer (1989) found that infrastructure capital has a large impact on aggregate TFP, this finding has been replicated by a number of earlier studies: Munnell (1990a, 1990b, 1992) for the United States, Mitra et al. (2002) for that of India, and Easterly and Rebelo (1993) for cross-sectional country data. Loayza, Fajnzylber, and Calderón (2002) find that a telecommunications indicator is robustly related to growth in a large panel data set that includes both industrial and developing countries.

For the case of Africa, studies exploring the relation between infrastructure and growth are scarce.[2] Traditionally, infrastructure services have been viewed as public goods in Africa, with their provision entrusted to government monopolies. The overall performance of government-owned providers of infrastructure in Africa has been very poor. This sector is characterized by high inefficiency, a lack of technological dynamism, and very poor service provision. In addition, the provision of infrastructure-related services in most African countries is characterized by high prices and long waits between the time of application for services and actual connection. Many African economies are also endowed with adverse natural and geographical attributes, such as lack of access to sea ports and tropical climates.

The Economic Commission for Africa (ECA) Report (2005) and Sachs et al. (2004) have explored the African need for new investments in infrastructure, but without a properly systematic cross-country analysis. Estache et al. (2005) makes one of the first attempts to conduct a more systematic, quantitative assessment of the importance of Sub-Saharan Africa's infrastructure. The main finding of this paper is that electricity, water, roads, and telecommunications are crucial factors in promoting growth, with colonial and postcolonial histories also being important factors explaining some of the differences among countries. On the other hand, Esfahani and Ramirez (2003) estimate that Sub-Saharan Africa's poor growth performance is, in part, related to underinvestments in electricity and telecommunications infrastructure. Estache (2005) estimates that if Africa had enjoyed Korea's quantity and quality of infrastructure, it would have raised its annual growth per capita by about 1 percent. Hulten (1996) finds that differences in the effective use of infrastructure resources explain one-quarter of the growth differential between Africa and East Asia, and more than 40 percent of the growth differential between low- and high-growth countries.

[2] Estache (2005) points out the two main reasons for ignoring the role of infrastructure as one of the most important drivers of economic growth in Africa: (a) econometric focus on human capital, and (b) low quality of available data.

Empirical explorations of infrastructure's effect on growth and productivity, however, have been characterized by ambiguous results with little robustness. The possible endogeneity of infrastructure measures has been advanced as a reason for contradictory findings of the impact of public capital on long-run economic development indicators. Literature has signaled that endogeneity in this context might come from three sources: (a) measurement errors stemming from the use of public capital figures as proxies for infrastructure; (b) omitted variables, which may arise when there is a third variable, unobserved, that affects the infrastructure and growth measure; and (c) the fact that under the simultaneous determination of infrastructure and productivity or output, the bias and inconsistency of standard estimators would follow where infrastructure provision itself positively responds to productivity gains.[3] Possible reasons for such feedback would arise with increased reliance on the private sector for the provision of infrastructure services, or with successful lobbying by industry interest groups that experience either positive productivity gains or constraints on performance due to infrastructure provision.

Various panel data and country studies have tried to address these issues. Röller and Waverman (2001) explicitly model and estimate the impact of telecommunications under simultaneity. In a cross-country panel estimation, Calderón and Serven (2003, 2005) employ generalized method of moments (GMM) panel estimation methods to control for the possibility of endogeneity, reporting significant improvements in results. Dessus and Herrera (1999) allow for simultaneity in a panel data set for 28 countries. Country-specific time series also confirm the presence of simultaneity between output and infrastructure measures—see Frutos et al. (1998) for Spain, and Fedderke et al. (2005) for South Africa. Also for South Africa, Fedderke and Bogetic (2006)—controlling for the potential endogeneity of infrastructure in estimation—robustly eliminate nearly all evidence of possible overinvestment in infrastructure. Indeed, controlling for the possible endogeneity of infrastructure measures renders the impact of infrastructure capital positive. Romp and Haan (2005) indicate that when simultaneity is taken into consideration, the elasticity estimates found in earlier studies considerably decrease.

Another possibility behind the ambiguous results obtained from empirical studies of public capital impacts on output might simply be that aggregate measures of infrastructure hide the productivity impact of infrastructure at a more disaggregated level. A second batch of studies, focusing mainly on microdata, reveals the existence of the possible indirect impact of infrastructure on economic growth and economic performance beyond the effect of the simple accumulation of capital. Thus, for instance, Shioji (2001) finds that the positive impact of infrastructure arises in panel data on U.S. and Japanese industries once public capital is properly disaggregated. Agénor and Moreno-Dodson (2006) point out that improvement in the stock of infrastructure can reduce the adjustment costs of private capital by (a) lowering the logistical cost

[3] Notice that we avoid using the terms *causality* or *reverse causality*, since there is no control group to compare against and the temporal dimension is not large enough to consider Granger-causality concepts. Therefore, we use the terms *simultaneity* and *identification*, which are more appropriate for ICSs.

of the investment in private capital, and (b) allowing the replacement of unproductive private investments such as electricity generators or boreholes and wells with more productive investments in machinery and equipment. This assumption has been tested in the context of investment climate assessments (ICAs) with firm-level information. Reinikka and Svensson (1999) show that improvements in the infrastructure stock in Uganda make infrastructure services more reliable, reducing the necessity of investing in less productive substitutes (such as generators) in order to avoid potential service interruptions, and thus freeing funding of private productive investments.

Relationships at a more disaggregated level tend to be obscured by aggregated data, and are unobservable with country-level data. Another channel of infrastructure impact is via improvements in labor productivity through (a) improved transport between home and work, and (b) more efficient work processes. Another way that better infrastructure might increase labor productivity is through improvements in health and education, making existing human capital more efficient, and promoting successive investments in human capital (Galiani et al., 2005).

The effect of infrastructure on firms' international integration has also been tested. Recent literature affirms that improvements in transportation services and infrastructure can lead to improvements in export performance. Thus, for instance, Francois and Manchin (2006) explore the role that infrastructure plays (among other factors such as policy reforms, institutional development, colonial history, development assistance, and general north-south differences) in the different trade performances observed in the so-called *globalizer* countries such as India and China, as well as other developing countries (many located in Africa and with a very different story to tell regarding the integration of the global economy). Limão and Venables (2001) show that infrastructure is quantitatively important in determining transport costs, concluding that poor infrastructure accounts for much of the different transport costs observed in coastal and landlocked countries. Bougheas et al. (1999), in the context of gravity models, find evidence in the European economy of a positive relationship between the level of infrastructure and the volume of trade. Wilson et al. (2004) consider ports, customs, regulations, and e-businesses as proxies of trade-facilitation efforts, finding that the scope and benefit of unilateral trade-facilitation reforms are very large, and that the gains fall disproportionately to exports.

In a world where governments compete to attract more FDI inflows through a variety of investment and tax incentives and other policy preferences, the availability of good-quality physical infrastructure could also increase the inflow of FDI by subsidizing the cost of total investment by foreign investors and thus raising the rate of return. The favorable role of physical infrastructure in influencing patterns of FDI inflows has been corroborated by recent studies, such as those of Loree and Guisinger (1995), Mody and Srinivasan (1996), and Kumar (2001), among others. Multinational enterprises may consider the quality of available infrastructure especially important while deciding to relocate export-platform production undertaken for efficiency considerations. In other words, the quality of physical infrastructure could be an important consideration for multinationals in their location choices, for FDI in general, and for efficiency-seeking production in particular.

As has been pointed out, the main concern of this paper is to offer a robust assessment of the various channels through which infrastructure quality may impact TFP. Thus, instead of the quantity of macrovariables, we use, as an explanatory variable, the quality of existing infrastructure stock. Instead of aggregate infrastructure measures usually included in macromodels, such as kilometers of paved roads or total number of telephone lines, we incorporate measures that allow us to identify direct relationships between infrastructure and economic performance at a more disaggregated level. Additionally, by going down to the firm level, we avoid the endogeneity problems of the macrolevel variables. Nevertheless, microlevel data have specific endogeneity problems, and several variables cannot be considered to be exogenously determined; for instance, public investment decisions are likely to be affected by expected returns on investment, and firms faced with different quality and availability of infrastructure services would choose different technologies. The solutions proposed in this methodology allow us to obtain a robust assessment of the impact of infrastructure quality on TFP.

3. Country-level data and their treatment in the study

Produced by the World Bank, ICSs of private enterprises explore the difficulties that firms located in developing countries encounter in starting and running businesses. More precisely, the surveys capture firms' experiences in a range of areas related to economic performance: financing, governance, corruption, crime, regulation, tax policy, labor relations, conflict resolution, infrastructure, supplies and marketing, quality, technology, training, and so on. For that purpose, we classify IC factors in five categories to evaluate the impact of each group on economic performance. In the first group—infrastructure—we include all related variables such as customs clearance, power and water supply, telecommunications (including phone connection and information technology, IT), and transportation. In the second group—red tape, corruption, and crime—we include IC factors relating to tax rates, conflict resolution, crime, bureaucracy, informalities, corruption, and regulations. The next group comprises financial and corporate governance and includes factors related to management, investments, informalities in sales and purchases, access to finance, and accountability (or auditing). The last group of IC variables includes quality, innovation, and labor skills, as well as quality certifications, technology usage, product and process innovation, research and development (R&D), quality of labor, training, and managers' experience and education. The last group—other C variables—are not properly a group of IC factors, but a group of other firms' control characteristics. We classify in this group all the factors that may have an important impact on economic performance but are not considered IC factors: exports and imports, age, FDI, number of competitors, firm size, and so on. Table A.2 (see appendix 2) includes the whole list of IC and C variables, as well as a description of how each is measured. Likewise, not all surveys provide the same information on ICSs, although there is a common group of variables in each group that is available for all the countries; although the regressions among them are slightly heterogeneous, we can use this common group as a benchmark for comparison purposes.

The ICSs provide information on TPF variables, output (sales), employment, intermediate materials, capital stock, and labor costs. Table A.1 (see appendix) includes information on these variables and indications of how they were measured. The ICSs do not provide information on prices at the firm level, so the production function (PF) variables were deflated by using the World Bank's country-specific consumer price index (CPI), base 2000. The information on the net book value of the capital stock (NBVC) is not available for Algeria, Kenya, Mali, Senegal, and Uganda; in these cases the NBVC is substituted by the replacement cost of machinery and equipment, which, in the surveys, is only available for a single year. We thus recursively estimate the missing values of the NBVC from the information on the replacement cost of and the net investment in machinery and equipment by using the permanent inventory method, according to which the capital stock at moment t is given by $K_{it}=K_{it-1}(1-\delta)+I_{it}$. By inverting this formula we can obtain the value of the capital at moment $t-1$ as $K_{it-1}=(K_{it}-I_{it})/(1-\delta)$ where K_{it} is approximated with the replacement cost of machinery and equipment, I_{it} is the net investment in machinery and equipment, and δ is the depreciation rate of the machinery and equipment.[4]

In this paper we focus on the manufacturing sector, and while classifying the establishments by their international standard of industrial classification (ISIC) code, we end up with establishments from the next eight sectors: (a) food and beverages; (b) textiles and apparels; (c) chemicals, rubber, and plastics; (d) paper, printing, and publishing; (e) machinery and equipment/metallic products; (f) wood and furniture; (g) nonmetallic products; and (h) other manufacturing.

Classification of countries by geographical area

For the classification of countries by groups used in the regression analysis, we take into account the following facts: (a) the surveys provide different information on PF variables and on IC and C variables; (b) the surveys were carried out in different years during the period 2002–6; (c) the quality of the data varies across surveys; and (d) not all the surveys provide panel data information (recall data) for the PF variables. Thus, we end up with two types of country databases. For those countries with a large enough number of observations available for regression analysis (see column 6 of table 3.1) and with panel data information for the PF

[4] The depreciation rate used is 15 percent, a standard percentage commonly applied in other works. Other percentages were also used in order to check robustness. Alternatively, to check whether the results were robust for other ways of constructing the NBVC, we used the next formula $K_{it-1}=K_{it}(1-\Delta I_{it})$, where K_{it} is approximated by the replacement cost and $\Delta I_{it}=(I_{it}-I_{it-1})/I_{it-1}$ is the rate of growth of the net investment in machinery and equipment. In both cases the main results were maintained.

variables (for more than one year), we carry out the analysis country by country. For the countries in which surveys were collected in 2006 (which only offer one year of information for PF variables) and the number of firms surveyed was lower than in the previous surveys, we follow the estimation strategy of pooling the information according to the similarity of geographical and economic factors—thus gaining efficiency in the parametrical estimation of the IC parameters (with more observations in the regressions) at the cost of having common IC parameters for some countries.

We end up with two pools of 2006 countries: (a) ECOWAS countries, such as Mauritania, Cameroon, Niger, and Burkina Faso; and (b) SADC countries, such as Botswana, Namibia, and Swaziland. Finally, since Eritrea has only 179 observations available, we consider this country as a special region of Ethiopia and carry out a joint analysis of the two, constituting the third pool of countries considered in the analysis.

Table A.3 offers an initial overview of the data we use in the analysis. We have data for 26 countries from five different geographical regions. Cape Verde, Lesotho, and Burundi are special cases. The PF information for Lesotho is rather poor and it is impossible to make reliable statistical inferences with only 79 observations. We did not group Lesotho with the pool of SADC countries because the survey of this country is from 2003 and the information on the IC and C variables is quite different. Burundi presents similar problems—the information on the PF for this country is for a single year (2005), and the number of observations is only 101. Although Burundi belongs to the EAC—along with Kenya, Uganda, and Tanzania—we did not pool Burundi with any of these countries because the information on the PF and the IC comes from different years and with different information on the IC and C variables. Cape Verde is another country with information for a single year (2006) and with only 47 observations available for regression analysis. Because of its obvious difference from the rest of the ECOWAS countries—different per capita income and its condition as an insular state, as well as other geographical considerations—we did not include Cape Verde in this pool. As a result, no regression analyses were conducted for Cape Verde, Lesotho, and Burundi.

[TABLE A.3 ABOUT HERE]

By running the regressions country by country we can use as many infrastructure and other IC and C variables as are available. This allows us to gain heterogeneity estimating the impact of infrastructure on productivity. In addition, we can use more variables as proxies for firm-level, unobservable fixed effects, and we do not have to constrain ourselves to the subset of IC variables common to all the countries.

Cleaning the data

The IC databases are, in some respects, troublesome. From table B.1 (see appendix) it is clear that out of the total number of establishments surveyed there are a considerable number of observations with at least one PF variable missing, and/or with outlier observations in the PF variables.[5] This problem becomes more acute for some countries—such as Algeria, Senegal, Eritrea, Tanzania, and Mauritius—where more than half the observations are missing for the regression analysis (see the upper panel of table B.1), which results in the sample representativity being lost. To reduce the effects of this sample selection bias, we apply a preliminary data-cleaning process that allows us to retrieve a considerable number of establishments for the analysis; it is based on a robust simple version of the EM-algorithm of Dempster et al. (1977) (for more details see Little and Rubin, 1987; Escribano et al., 2008). First, we exclude those plants with missing values in all the PF variables—sales, materials, capital stock, and labor cost. We convert outlier observations of PF variables into missing observations, then proceed as follows: (a) we replace the missing values by the corresponding (cells) industry-region-size median of the variables keeping from 15 to 20 observations in each cell; (b) if we do not have enough observations in some cells we replace them with the corresponding industry-size medians; (c) if we still do not have enough observations in those cells we replace them with the region-size medians; and (d) if still necessary, in the last step we compute the medians only by size and/or by industry to replace those missing values. Table B.1 shows that the number of available observations in all the countries considerably increases with the application of this data-cleaning process. Tables B.2.1 and B.2.2 (see appendix 2) show the distribution of the observations (by country and year, and by country and industry, respectively) before and after the cleaning process. From these tables it is clear that this process does not alter much of the original representativity of the ICSs.

The importance of infrastructure among IC variables

As has been previously pointed out, we classify the IC factors in several categories to evaluate the impact of each group on economic performance. The infrastructure group of variables (INFs) is intended to be part of the country-specific IC. Within the infrastructure group we consider the next list of IC variables: customs clearance, energy, water, telecommunications, ICT, and transportation. Table B.4 describes the main INFs used in the empirical analysis.

[TABLE B.4 ABOUT HERE]

The variables listed in table B.4 are common to almost all the countries considered, and are therefore intended to be a benchmark for comparison purposes; however, there are other country-specific variables not listed. For a description of the complete set of variables, along

[5] By *outlier*s we mean those observations with ratios of materials to sales and/or labor cost to sales greater than 1.

with the countries for which they are available and the response rate of the variables, see table B.3 of the appendix.

Within each infrastructure subgroup we consider different factors. Thus, in the *customs clearance group* the factor considered is the time required to clear customs for exports and imports, and the time to get an import license. In the *energy group* we consider variables that describe the quality of power provision (number and average duration of power outages, and subsequent losses), the use of a generator as a substitute for the public provision of power, the price of energy either from the public grid or from private generators, and the average time waiting to be hooked up for electricity supply. Similarly, for the *group of water* we consider provision quality, price, the use of alternative supplies of water (such as private wells or boreholes), and the time to get water supply. In the *telecommunications and ICT group*, the variable considered is the quality of the phone provision and the time to obtain a phone connection, as well as the use of ICT technologies (such as Internet or e-mail) in communications with clients and suppliers. The *transport group* mainly incorporates a description of the quality of transportation services and dummy variables for the use of own-transport services (roads, transportation for workers, and so on).

From the econometric point of view we use three types of variables: (a) variables in logs, whose coefficients are interpreted as elasticities; (b) variables in percentages, whose coefficients can be interpreted as semi-elasticities; and (c) dummy variables, for which coefficients from the regressions are interpreted as semi-elasticities.

Finally, some of the variables in the same group are likely to be correlated since they provide similar information; for instance, the number and average duration of power outages and subsequent losses. In order to avoid multicolinariety problems we do not simultaneously use all variables in the regressions, but in the final model specification we test for possibly omitted variables.

4. Evaluating the impact of infrastructure on total factor productivity (TFP)

Escribano and Guasch (2005, 2008) relate infrastructure and other IC and C variables with firm-level productivity (TFP) according to the following observable fixed-effects system of equations:

$$\log Y_{it} = \alpha_L \log L_{it} + \alpha_M \log M_{it} + \alpha_K \log K_{it} + \log TFP_{it} \quad (4.1a)$$

$$\log TFP_{j,it} = a_i + \alpha'_D D_i + \alpha_P + w_{it} \quad (4.1b)$$

$$a_i = \alpha'_{INF} INF_i + \alpha'_{IC} IC_i + \varepsilon_i \quad (4.1c)$$

where, Y is firms' output (sales), L is employment, M denotes intermediate materials, K is the capital stock, INF is a time-fixed vector of observable infrastructure variables, IC is a time-fixed effect vector of other investment climate and other control variables, and *D* is a vector of industry and year dummies.

The usually unobserved time fixed effects (a_i) of the TFP equation (4.1b) is here proxy by the set of observed time-fixed components **INF** and **IC** variables of (4.1c) and a remaining unobserved random effect (ε_i). The two random error terms of the system, ε_i and w_{it}, are assumed to be conditionally uncorrelated with the explanatory L, M, K, **INF** and **IC** variables[6] of equation (4.2):

$$\log Y_{it} = \alpha_L \log L_{it} + \alpha_M \log M_{it} + \alpha_K \log K_{it} + \alpha'_{INF} INF_i + \alpha'_{IC} IC_i + \alpha'_D D_i + \alpha_P + u_{it}\,. \tag{4.2}$$

Therefore, the regression equation (4.2) is representing the *conditional expectation* plus a composite RE error term equal to $u_{it} = \varepsilon_i + w_{it}$.

Providing reliable and robust estimates of the impact of infrastructure on productivity is not a straightforward task. First, because the functional form of the PF is not observed and there is no available single salient TFP measure. Second, there is an identification issue separating TFP from PF. When any PF inputs are influenced by unobserved common causes affecting productivity—such as a firm's fixed effects—there is a simultaneous equation problem in equation (4.1a). Third, we could expect that several IC variables have at least some degree of endogeneity, questioning therefore the conditional lack of correlation of (4.2). In what follows of this section, we briefly review the solutions to these questions suggested in Escribano and Guasch (2005, 2008) and Escribano et al. (2008).

Estimating infrastructure's impact on productivity

TFP or multifactor productivity refers to the effects of any variable different from the inputs—labor (L), intermediate materials (M), and capital (K)—affecting the production (Y) process. Since there is no single salient measure of productivity (or $logTFP_i$), any empirical evaluation of the productivity impact of INFs might critically depend on the particular productivity measure used. Escribano and Guasch (2005, 2008) suggested—following the literature on *sensitivity analysis* of Magnus and Vasnew (2006)—to look for empirical results (elasticities) that are robust across several productivity measures. This is also the approach we follow in this paper.

In particular, we want the elasticities of INFs on productivity (TFP) to be robust (with equal signs and similar magnitudes) for the 10 different productivity measures used. Alternative productivity measures come from considering:

- Different functional forms of the PFs (Cobb-Douglas and Translog)

[6] Under this formulation (and other standard conditions) the OLS estimator of the productivity equation (4.2) with robust standard errors is consistent, although a more efficient estimator (GLS) is given by the random effects (RE) estimator that takes into consideration the particular covariance structure of the error term, $\varepsilon_i + w_{it}$, which introduces a certain type of heteroskedasticity in the regression errors of (4.2).

- Different sets of assumptions (technology and market conditions) to get consistent estimators based on Solow's residuals, ordinary least squares (OLS), or random effects (RE), and so on
- Different levels of aggregation in measuring input-output elasticities (at the industry level or at the aggregate country level)

Box 4.1 Summary of productivity measures and estimated investment climate (IC) elasticities

1. Solow´s Residual	Two-step estimation	1.1 Restricted coefficient 1.2 Unrestricted coefficient	1.1.a OLS 1.1.b RE 1.2.a OLS 1.2.b RE	2 (P_{it}) measures 4 (IC) elasticities
2. Cobb-Douglas	Single-step estimation	2.1 Restricted coefficient 2.2 Unrestricted coefficient	2.1.a OLS 2.1.b RE 2.2.a OLS 2.2.b RE	4 (P_{it}) measures 4 (IC) elasticities
3. Translog	Single-step estimation	3.1 Restricted coefficient 3.2 Unrestricted coefficient	3.1.a OLS 3.1.b RE 3.2.a OLS 3.2.b RE	4 (P_{it}) measures 4 (IC) elasticities
Total				10 (P_{it}) measures 12 (IC) elasticities

Source: Author's estimations.
Note: Restricted coefficient = equal input-output elasticities in all industries.
Unrestricted coefficient = different input-output elasticities by industry.

Box 4.1 summarizes the productivity measures used for the IC evaluation. The two-step estimation starts from the nonparametric approach based on cost shares from Hall (1990) to obtain Solow's residuals in logs under two different assumptions:[7] (a) the cost shares are constant for all plants located in the same country (restricted Solow residual), and (b) the cost shares vary among industries in the same country (unrestricted by industry Solow residual). Once we have estimated the Solow residuals ($logTFP_i$) in the first step, in the second step we can estimate equation (4.3) by OLS with robust standard errors for the countries that have a single year of data (2006) on PF variables. For the remaining 14 countries and for the blocks of countries described in section 3, we can also estimate (4.3) by RE to obtain the corresponding IC elasticities and semi-elasticities,

$$\log TFP_{it} = \alpha'_{INF} INF_i + \alpha'_{IC} IC_i + \alpha'_D D_i + \alpha_P + \varepsilon_i + w_{it} \qquad (4.3a)$$

where *INF* is the observable fixed effects vector of infrastructure variables, and *IC* is the observable fixed effects vector of other IC and control variables listed in table A.2 of the appendix. In all the panel data regressions, we always control for several sector-industry

[7] The advantage of the Solow residuals is that they require neither the inputs (L, M, K) to be exogenous nor the input-output elasticities to be constant or homogeneous (Escribano and Guasch, 2005 and 2008). The drawback is that they require having constant returns to scale (CRS) and, at least, competitive input markets.

dummies (D_j, j = 1, 2, . . ., q_D), and in the cases having more than one year of observations we also include a set of time (D_t, t = 1, 2, . . ., q_T) dummy variables and always a constant term (α_P).

For cross-country comparisons based on TFP we use the concept of demeaned TFP,[8] which gets rid of the constant term as well as the constant effects by industry and by year, concentrating therefore on the part of TFP that is influenced by INF, IC, and the other plant-level control variables,

$$Demeaned\ \log TFP_{it} = \alpha'_{INF} INF_i + \alpha'_{IC} IC_i \, . \tag{4.3b}$$

In the single-step estimation approach, we consider the parametric estimation by OLS and RE of the extended PF (4.2). To address the well-known problem of the endogeneity of inputs, we follow the approach proposed by Escribano and Guasch (2005, 2008). That is, we proxy the usually unobserved firm-specific fixed effects (which are the main cause of inputs' endogeneity) by a long list of observed firm-specific fixed effects coming from the ICSs. Controlling for the largest set of IC variables and plant characteristics, we can get—under standard regularity conditions—consistent and unbiased least squares estimators of the parameters of the PF and the INF and IC elasticities. Furthermore, we use two different functional forms of the PF—Cobb-Douglas and Translog—under two different assumptions on the input-output elasticities: equal input-output elasticities in all industries (restricted case) and different input-output elasticities by industries (unrestricted case).

Notice that even if we are only interested in assessing the impact of infrastructure on TFP, we do not limit the scope of the control analysis to only this subset of IC variables. We include (and therefore control for) all the IC factors because of the crucial role IC variables play as proxies for the unobserved fixed effects; this is the key feature of this methodology in order to provide robust empirical results. If we tried to estimate only the impact of infrastructure, without controlling for the other blocks of IC variables, we might get different signs on certain coefficients because of the omitted variables problem (Escribano and Guasch, 2008).

Another econometric problem we have to face when estimating the parameters of the INF and IC variables—either from the two-step or single-step procedure—is the possible endogeneity of some of these variables. That is, many INF or IC variables are likely to be determined simultaneously along with any TFP measure. With these productivity equations, the traditional instrumental variable (IV) approach is difficult to implement, given that we only have information for one year, and therefore we cannot use the natural instruments for inputs, such as those provided by their own lags. As an alternative correction for the endogeneity of the INF and IC variables, we use the region-industry-size average of firm-plant-level INF and IC variables

[8] Notice that the demeaned TFP concept of equation (4.3b) corresponds to the observable part of the fixed effects equation (4.1c).

instead of the crude variables,[9] which is a common solution in panel data studies at the firm level[10] (see Veeramani and Goldar, 2004, for other use of industry-region averages with IC variables).

Using industry-region-size averages also mitigates the effect of having certain missing individual INF and IC observations at the plant level, which—as mentioned in section 3—represent one of the most important difficulties of using ICSs. As an alternative, we also follow a second strategy to deal with the missing values of some INF and IC variables. In order to keep as many observations in the regressions as possible to avoid losing efficiency, when the response rate of the variables is large enough, we decided to replace those missing observations with the corresponding industry-region-size average.[11] Thus, we gain observations, efficiency, and representativity at the cost of introducing measurement errors into some variables.[12]

The econometric methodology applied for the selection of the variables (INF and IC) goes from the general to the specific. The otherwise omitted variables that we encounter—starting from a too-simple model—generate biased and inconsistent parameter estimates. We start the selection of variables with a wide set compounded by up to 90 variables (depending on the country). We avoid simultaneously using time variables that provide the same information and are likely to be correlated, mitigating the problem of multicollinearity that could otherwise arise. We then start removing the less significant variables from the regressions one by one, until we obtain the final set of variables, all significant in at least one of the regressions and with parameters varying within a reasonable range of values. Once we have selected a preliminary model we test for omitted INF and IC variables.

The robust coefficients of the INF and IC variables in productivity, along with their level of significance, are available upon from request. The parameters estimated in the two step procedure with restricted input-output elasticites can be found in figures 6.1 to 6.23.

Infrastructure assessment based on O&P decompositions

According to the O&P (1996) decomposition, aggregate productivity for a given country, industry, or region may be decomposed into two terms: (a) average productivity, and (b) a covariance term measuring whether the economy is able to efficiently reallocate resources from

[9] Because of the low number of available regions in most of the countries, we had to use the industry-region-size variables instead of the region-industry averages. For the creation of cells a minimum number of firms are imposed—there must be at least 15 to 20 firms in each industry-region-size cell to create the average, otherwise we apply the region-industry averages. If the problem persists, we apply the industry-size or the region-size average.

[10] This two-step estimation approach is a simplified version of an instrumental variable (IV) estimator (two-stage least squares, 2SLS).

[11] Notice that this replacement strategy has a straightforward weighted least squares interpretation since we are giving a greater weight to those observations with more variance (Escribano et al., 2008).

[12] The measurement error introduces a downward bias in the parameters that depends on the ratio between the variances of the variables and the measurement error. Since those explanatory variables are constant within regions, sizes, and industries we expect their variances will be small.

less productive establishments to more productive ones. Once we have estimated a robust set of parameters for the IC factors with statistically significant impacts on firms' productivity, we exploit the exact relation, proposed by Escribano et al. (2008a), between the terms of the O&P decomposition, and the IC factors affecting productivity. The IC infrastructure variables affect both the average productivity of African establishments (or firms) as well as their allocative efficiency component. It is well known that competitive markets efficiently allocate resources under certain conditions. But in a world of imperfect information a turbulent IC introduces distortions into markets, and, as a result, affects the efficiency of the economy as a whole. The allocative efficiency term of the O&P decomposition should therefore reflect those imperfections.

In the second part of this analysis—taking advantage of the robustness of the INF, IC, and C elasticities estimated—our aim is to concentrate on the TFP measure that comes from the restricted Solow's residuals in order to evaluate the infrastructure effects on average productivity and on allocative efficiency based on the O&P decomposition of aggregate productivity in levels,

$$TFP_q = \overline{TFP}_q + N_q \,\hat{\text{cov}}(s^Y_{q,it}, TFP_{q,it}) \,. \tag{4.4a}$$

Furthermore, we want to exploit the log-linear properties of the following mixed[13] O&P decomposition for each of the African countries considered in order to obtain closed form O&P decompositions in terms of IC and C variables:

$$\log TFP_q = \overline{\log TFP}_q + N_q \,\hat{\text{cov}}(s^Y_{q,it}, \log TFP_{q,it}) \,. \tag{4.4b}$$

Aggregate log productivity of country *q*, say ($\log TFP_q$), is equal to the sum of the sample average log productivity of the establishments of country *q*, and the covariance between the share of sales (s^Y_q) and log productivity of that country (allocative efficiency of country *q*). The index *q* could also indicate a particular industry, region, size, and so on. The useful additive property of equation (4.2) in logarithms allows us to obtain an exact closed form solution of the decomposition of aggregate log productivity according to equation (4.5). Following Escribano et al. (2008), we can express aggregate log productivity as a weighted sum of the average values of the IC, dummy D variables, the intercept, and the productivity average residuals ($\hat{\omega}$) from (4.2), and the sum of the *covariances* between the share of sales and investment climate variables IC, dummies D, and the productivity residuals ($\hat{\omega}$):

$$\begin{aligned} \log TFP_q = {} & \hat{\alpha}_{INF}\overline{INF}_q \hat{\alpha}_{IC}\overline{IC}_q + \hat{\alpha}_D \overline{D}_q + \hat{\alpha}_p + \bar{\hat{u}}_{q,it} + N_q \hat{\alpha}_{INF}\,\hat{\text{cov}}(s^Y_{q,it}, INF_{q,i}) \\ & + N_q \hat{\alpha}_{IC}\,\hat{\text{cov}}(s^Y_{q,it}, IC_{q,i}) + N_q \hat{\alpha}_{Ds}\,\hat{\text{cov}}(s^Y_{q,it}, D_i) + N_q\,\hat{\text{cov}}(s^Y_{q,it}, \hat{u}_{q,it}) \end{aligned} \tag{4.5}$$

[13] It is called a *mixed* Olley and Pakes (O&P) decomposition because in the original O&P decomposition both TFP and the share of sales were in levels, while TFP in (4.4b) is in logs (log P).

where the set of estimated parameters used comes from the two-step TFP estimation, having the restricted Solow's residual as a dependent variable in (4.2).

From equation (4.5) each INF and IC variable may affect the aggregate log productivity through both its average and covariance (with respect to the share of sales). This complements the information provided by the marginal effects (INF and IC elasticities). Suppose that an INF variable with a low impact—in terms of marginal effects (elasticities)—affects most of the firms in a given country; the impact of such an IC variable in terms of average productivity could be very high. It is therefore very important for policy analysis to combine the empirical evidence from the estimated IC elasticities on productivity with their corresponding INF impact on the two components of O&P decompositions: average productivity and allocative efficiency.

A variable with a negative marginal effect on average productivity (or logTFP) may have either a positive or a negative effect on allocative efficiency. If the covariance of that IC variable and the market share is positive, then the greater proportion of sales in the hands of establishments with high levels of that variable, the larger the negative impact on aggregate productivity will be, therefore decreasing the allocative efficiency. In contrast, a negative covariance means that those establishments with the highest levels of the IC variable have the lowest market shares, and therefore the negative effect of the IC variable on average productivity is somehow compensated through the effect on the reallocation of resources among firms.

By operating in (4.5) Escribano et al. (2008) obtained the next expression, which allows us to obtain a direct decomposition of the impacts of each INF and IC variable on aggregate productivity ($\log TFP_q$):

$$\begin{aligned}100 = \frac{100}{\log TFP_q}[\hat{\alpha}_{IC}\overline{INF}_q + \hat{\alpha}_{IC}\overline{IC}_q + \hat{\alpha}_D\overline{D}_i + \hat{\alpha}_p + \overline{\hat{u}}_i \; + N_q\hat{\alpha}_{INF}\hat{c}ov(s^Y_{q,i}, INF_{q,i}) \\ + N_q\hat{\alpha}_{IC}\hat{c}ov(s^Y_{q,i}, IC_{q,i}) + N_q\hat{\alpha}_D\hat{c}ov(s^Y_{q,i}, D_i) + N_q\hat{c}ov(s^Y_{q,i}, \hat{u}_{q,i})].\end{aligned} \tag{4.6}$$

There are several advantages of using equation (4.6). First, we can compare net contributions by isolating the impact of INF and other IC variables from the impact of industry dummies, the intercept, and the residuals. Second, we can express what portion of aggregate productivity is explained by INF, IC, and C variables (demeaned log TFP), and what proportion is due to the constant term, industry dummies, and so on. To make cross-country comparisons based on IC impacts on TFP and to avoid the problem of comparing apples and oranges, it is desirable to create an index (demeaned TFP). After subtracting the mean (that is, the constant term, time effects, industry effects, and country-specific effects) from aggregate productivity we can concentrate on the contributions of IC variables to the demeaned TFP.

Similarly, we can construct the demeaned counterparts of expressions (4.5) and (4.6) and compute the percentage contribution of each INF variable or block of IC variables—as in equations (4.7) and (4.8), respectively—obtaining the following *demeaned mixed O&P decomposition*:

$$Demean \log TFP_q = \hat{\alpha}_{INF} \overline{INF}_q \hat{\alpha}_{IC} \overline{IC}_q + N_q \hat{\alpha}_{INF} \,\hat{\text{c}}\text{ov}(s^Y_{q,it}, INF_{q,i}) + N_q \hat{\alpha}'_{IC} \,\hat{\text{c}}\text{ov}(s^Y_{q,it}, IC_{q,i}) \quad (4.7)$$

$$100 = \frac{100}{Demean \log TFP_q} [\hat{\alpha}'_{IC} \overline{INF}_q + \hat{\alpha}'_{IC} \overline{IC}_q + N_q \hat{\alpha}_{INF} \hat{c}ov(s^Y_{q,i}, INF_{q,i}) + N_q \hat{\alpha}_{IC} \hat{c}ov(s^Y_{q,i}, IC_{q,i})]. \quad (4.8)$$

So far, we have exploited the linear properties of the logarithm form of the mixed O&P decomposition of TFP. But the original O&P decomposition is based on TFP and the share of sales (in levels), and is therefore also capturing nonlinear relations between market shares and IC variables coming from (4.3a). To know to what extent these nonlinear terms are affecting this relation, we perform simulation experiments[14] on INF, IC, and C variables, and evaluate the consistency of the results with the ones obtained from the previous mixed O&P decomposition—see (4.4b). The IC simulations are done variable by variable (one at a time) keeping the rest of the variables constant; that is, we propose a scenario in which the level of one of the IC variables improves by 20 percent in all establishments (20 percent less power outages, 20 percent less shipment losses, etc). We compute the corresponding rate of change of aggregate productivity, average productivity, and allocative efficiency caused by such a 20 percent improvement. We repeat the same experiment for the rest of the IC and C variables, and, for comparative purposes, we also evaluate the relative group of IC variables.

5. The contribution of infrastructure to the investment climate (IC) of Africa

In section 4 we described the econometric methodology used in section 6 to assess the impact of infrastructure on productivity. We suggested three key elements of empirical evaluations of infrastructure and other IC and C impacts on productivity: the marginal productivity effects, the percentage contributions of infrastructure to aggregate log productivity (mean and efficient components), and the simulations of infrastructure improvements on aggregate productivity (in levels).

In this section we focus on presenting the results of infrastructure contributions to aggregate productivity from simulation experiments. In addition to the results of the econometric analysis, we consider African firms' perceptions of the main obstacles to economic performance. In the first subsection the objective is to assess how African firms perceive infrastructure quality. This is followed in the next subsections by the results of the econometric analysis, focusing on the infrastructure's effect on productivity after controlling for other IC factors. Finally, to complement the robustness of the results we check the consistency of the conclusions obtained from both the IC contributions to average log TFP and from the TFP simulation experiments.

Do African firms perceive infrastructure as an obstacle to growth?

[14] We are indebted to Ariel Pakes for this suggestion.

In ICSs, firms are asked to rate a number of IC factors as obstacles to economic performance. The survey options offered are *no obstacle*, *minor obstacle*, *moderate obstacle*, *major obstacle*, or *very severe obstacle* on a broad range of IC aspects: infrastructure, red tape, corruption and crime, finance, and labor skills.

Figure 5.1 (see appendix 3) shows the degree to which each group of IC factors is perceived by firms as an obstacle to economic development. These perceptions are sorted in descending order by their perceived contribution to the total, after being normalized to 100. For example, in panel A of figure 5.1, we observe that in Cape Verde 25 percent of firms believe infrastructure to be a major or very severe obstacle to economic growth; 40 percent find red tape, corruption, and crime as a major or very severe obstacle; 23 percent finance; and 10 percent a lack of labor skills. The countries in which infrastructure is perceived as an especially great obstacle to growth are—in descending order—Cape Verde, Burundi, Burkina Faso, Mauritania, Cameroon, Ethiopia, and Niger. Countries where a relatively low number of firms perceive infrastructure as a major constraint are Egypt, Mauritius, Morocco, South Africa, and Botswana.

[FIGURE 5.1 ABOUT HERE]

The factor group with the largest number of subfactors is red tape, corruption, and crime. So it is not surprising that this has the largest percentage of all. In order to isolate this effect, we normalize after computing the mean of each IC factor type. For example, in the infrastructure group there are four IC subfactors (telecommunications, customs, electricity, and transportation), whereas the red tape group contains eleven subfactors. But the importance of infrastructure is very similar across countries, although obviously the relative contribution of the red-tape group is more balanced, gaining relatively more relevance to the other IC groups.

In the next subsection the results of the econometric analysis—estimating the relative impact of infrastructure on average productivity using simulations—are compared with firms' perceptions of obstacles to growth. The question of interest is: are the econometric results consistent with firm perceptions?

Impact of infrastructure on productivity

The impact of infrastructure factors on productivity are evaluated here in terms of their effect on the O&P decomposition. Infrastructure's contribution to the aggregate productivity of each country's manufacturing sector is decomposed into its contribution to (a) average productivity and (b) allocative efficiency (the ability of markets to reallocate resources from less productive to more productive establishments).

[FIGURE 5.2 ABOUT HERE]

Figure 5.2 presents the two alternative country-by-country O&P decompositions given by equations (4.4a) and (4.4b), sorted by aggregate productivity in descending order. The productivity measure used to calculate the O&P decomposition is the restricted Solow residual obtained from the two-step estimation approach (see section 4). We present two sets of results with O&P decompositions. Panel A of figure 5.2 shows the O&P decomposition with the restricted Solow residual in levels, and panel B shows the mixed O&P decomposition with the restricted Solow residual, in logs, weighted by the share of sales. This is important because the results of the simulations are associated with the O&P decomposition in levels, and the results from the percentage contributions to the average use the convenient additive property of the TFP equation in logs. Notice that both panels A and B preserve the rankings of average productivity, but this is not necessary true for aggregate productivity. The reason is clear: aggregate productivity is simply the sum of average productivity and allocative efficiency, and this second term depends somewhat on whether we use TFP in levels or in log form.[15] From panel A of figure 5.2 we observe a positive reallocation of output. That is, output is moving from less productive establishments to more productive ones, since in all the countries the allocative efficiency is positive, with the greatest effects found in Benin, Burkina Faso, Cameroon, Niger, and Eritrea.

Notice that we avoid direct comparisons of TFP across countries but we suggested in section 4 to compare demeaned productivity decompositions (see figure 5.2, panels A and B, in the appendix). Remember that the demeaned productivity (or demeaned TFP) at the firm level is simply firm-level productivity minus the constant term of the productivity equation and the industry and year dummies (see equation 4.7).[16] We assume that all the productivity differences resulting from units of measurement, different deflators, and so on are contained in the constant, industry, and time-fixed effects, and therefore what is left in the productivity measure are only the TFP effects of the infrastructure, IC, and other C variables.

[FIGURE 5.3 ABOUT HERE]

We obtain the O&P decomposition using the demeaned productivity either in levels or in logs. This demeaned TFP set of cross-country comparisons is presented in figure 5.3 (see appendix 3). Panel A shows the decomposition of the demeaned productivity in levels; it is interpreted as the productivity that stems from IC conditions after controlling for all the other elements. The results are not at all surprising since they are basically consistent with those provided by the per capita income and by the DBR (2007). Rankings based on demeaned productivity are topped by South Africa and Mauritius, closely followed by Botswana, Algeria, Egypt, Namibia, and Swaziland. The lowest-ranked countries are those with the most

[15] For a deeper discussion of this issue see Escribano et al. (2008).

[16] Obviously, the year dummies are only subtracted from the productivity measure of the countries with panel data.

antiproductive IC, in other words, those whose IC conditions pose difficulties to economic development. These countries are Tanzania, Malawi, Uganda, Benin, Mauritania, and Zambia. Symmetrically, as for the regular O&P decompositions, the contribution of the IC to aggregate demeaned productivity is decomposed into its contributions to average demeaned productivity and the allocative demeaned efficiency term (see equation 4.7). Notice that, in Africa, the allocative efficiency component is always lower than the effect of average productivity. Nevertheless, in Madagascar, Botswana, Mauritius, and other countries, the IC has a considerable effect on the efficient reallocation of resources among establishments.

Alternatively, this demeaned productivity may be interpreted as a sum of pro- and antiproductive infrastructure, as well as other IC and C factors. Examples of proproductive infrastructure factors are the use of e-mail and websites. Negative or antiproductive infrastructure factors include the number of power outages, the average duration of water outages, and so on. As a consequence, productivity will decrease as the importance of antiproductive factors becomes larger and larger; this picture becomes even clearer in panel B of figure 5.4 (see appendix 3). The demeaned O&P decomposition of TFP in logs (see panel B) shows how aggregate productivity may be negative (in Tanzania, Benin, Malawi, and so on) when the negative TFP aspects of IC dominate over the positive (proproductive IC factors weigh more than the negative ones), as in the case of South Africa, Mauritius, Egypt, Botswana, and so on.

In sum, in African countries the IC has important effects on the aggregate productivity of the manufacturing industry, and this net effect may be positive or negative depending on which IC aspect matters more—the proproductive or the antiproductive. The aim now is to know to what extent such decreases or increases in productivity are due to infrastructure or other IC groups.

[FIGURE 5.4 ABOUT HERE]

Figure 5.4, panel A, provides the decomposition of demeaned productivity in levels using simulations of improvements to IC variables. When the IC factor improves by 20 percent it could mean, for example, that 20 percent more firms are using e-mail, or that there is a 20 percent reduction in power outages, and so on—which implies decreases in the negative IC factors and increases in the positive ones. The total effect of improving each IC by 20 percent, maintaining the rest of IC factors constant, implies that aggregate productivity could increase in South Africa by 55 percent, in Mauritius by 30 percent, and so on. From panel A of figure 5.4 it is clear that there are some economies that are more likely to be affected by the IC. These are therefore more sensitive to changes in IC conditions. This is the case in Kenya and Benin, where the aggregate productivity could increase by 70 percent and 85 percent, respectively. At the other extreme are Egypt, Morocco, and Eritrea. Lastly, improvements to aggregate productivity come in almost all countries via improvements to average productivity, and, to a lesser extent, allocative efficiency, with the exception of Algeria, Kenya, and Benin. The role of infrastructure in the composition of aggregate productivity is considerable in all the countries, but is the greatest in Uganda, Benin, Malawi, Cameroon, and Zambia. This suggests that these countries are the most sensitive to changes in infrastructure quality.

Panel B of figure 5.4 shows a more static interpretation using the O&P decomposition in logs by group of variable. In particular, panel B offers information on the actual and current situation of the IC and its effect on aggregate productivity; in other words, gains and losses generated by the average IC conditions (O&P decomposition of TFP in logs decomposed by groups of INF, IC, and C variables). For example, in South Africa, aggregate *demeaned* TFP is 0.83 (see panel B of figure 5.3). Out of this level of productivity -0.9 is explained by the overall contribution of the infrastructure factors; 0.95 by red tape, corruption, and crime; and the rest by the remaining IC and C variables. Notice that in panel B the contributions of the different groups are not in absolute value, so the positive effect of the proproductive factors compensate for the negative effect of the antiproductive ones. Even taking this into account, the overall contributions of the infrastructure group are negative in all the countries, implying that the proproductive infrastructure IC factors never compensate for the negative IC effects, with the exception of Kenya (where it is slightly positive) and Madagascar, Ethiopia, and Algeria (where the contribution of the infrastructure group is close to zero and almost negligible). As expected, the largest and most negative infrastructure effect is found in Benin, followed by Malawi, Uganda, Mauritania, Cameroon, and Zambia.

[FIGURE 5.5 ABOUT HERE]

Continuing with the same idea in figure 5.5, we are interested in obtaining the weight of the infrastructure group relative to the IC as a whole. Thus, by normalizing to 100 the contribution of the IC to aggregate productivity, average productivity, and allocative efficiency, we find via simulations that the relative 20 percent improvement of infrastructure in Malawi reaches 58 percent, in Eritrea 50 percent, and in Uganda 45 percent (as panel A of figure 5.5 shows). The same holds for average productivity (panel B) where the rankings do not change, and for allocative efficiency (panel C), where, once again, Malawi, Senegal, and Uganda show the largest contributions of the infrastructure group.

[FIGURE 5.6 ABOUT HERE]

A similar picture is provided by figure 5.6, where, instead of simulations, we consider the relative contributions by groups of variables to average demeaned log productivity and to the demeaned efficiency term—see equation (4.8). In this case we sum up the different contributions of the INF, IC, and C factors of equation (4.8), but in absolute value so that the negative effects do not compensate for the positive ones and vice versa, and we compute the relative contribution of each group within the IC group as a whole. Therefore, the relative contribution of the infrastructure group is the sum in absolute value of all individual infrastructure variables divided by the total absolute contribution of all INF and IC variables—multiplied by 100. The largest relative effects of infrastructure on aggregate log productivity are found in Malawi (60 percent), Uganda (50 percent), Benin (50 percent), Zambia (47 percent), and Ethiopia (46 percent). The lowest contributions are in Kenya, Swaziland, and Botswana. A similar ranking is provided by panel B, where the effects on the average log productivity are isolated from those from the allocative efficiency, as seen in panel C. Once again, Malawi, Benin, Senegal, Uganda, and Ethiopia lead a ranking closed by Mauritius, Egypt, Swaziland, Botswana, and Namibia. Panel C offers the results for allocative efficiency. In Malawi, Senegal, Namibia, and Algeria, the effects of infrastructure on the efficient reallocation of results among firms appear to be very significant, reaching the relative contributions of 54 percent, 48 percent, and 46 percent, respectively.[17]

Cross-country comparisons

Table C.1 summarizes the empirical results discussed in previous sections. The first column shows ranking of African countries based on per capita income, the second based on the DBR (2007), the third column based on quality of overall infrastructure (1 minimum, 7 maximum) given in the ACR (2007), the fourth column the demeaned aggregate productivity, and the fifth column shows the ranking of firms' perceptions of the quality of infrastructure (from 23rd being the poorest quality to 1st being the best quality in our sample). Columns 6 and 7 show the percentage of absolute contributions of infrastructure to average log productivity and to allocative efficiency, with TFP in logs, while columns 8 and 9 show the percentage absolute contributions of infrastructure to average productivity in levels and to allocative efficiency via simulations.

[TABLE C.1 ABOUT HERE]

The rankings presented in the first five columns are very consistent. In particular, the ranking based on demeaned aggregate productivity (column 4) shows a clear positive correlation to per capita GDP and with the ranking based on the DBR rankings. From the results of the

[17] It is useful to clarify that the differences between the rankings of the contributions to the aggregate productivity via simulations of panel A of figure 5.5 and the rankings of the contributions to the aggregate log productivity of panel A of figure 5.6 come mainly from the role of the allocative efficiency. Notice that in figure 5.6 the allocative efficiency term based on log TFP does not have the same scale as the efficiency term when TFP is in levels. As a result, since aggregate productivity is simply the sum of the average productivity and the efficiency term in levels, the role of the efficiency term with TFP in levels will increase with respect to its counterpart in logs and therefore, could alter the rankings of countries based on the two alternative measures of aggregate productivity (weighted productivity).

rankings obtained from the first five columns we find two groups of African countries, as was suggested by looking at the growth rates of per capita GDP (see figure 1.2, panel B). That is: (a) countries in the north and south of Africa are relatively more successful, and (b) countries from the central-east and central-west regions of Africa are relatively less successful.

The last four columns show two alternative measures of the percentage absolute contribution of infrastructure to productivity, along with the ranking in parentheses. In particular, column 6 shows a negative correlation between the ranking based on the contribution of IC to average log TFP and per capita GDP and also with the ranking based on the DBR, indicating that low infrastructure quality is one of the key growth bottlenecks in Africa. The results show a great homogeneity among the rankings in the first four columns and the results from the econometric analysis. Thus, for instance, Mauritius is ranked 1st in terms of per capita income and quality of overall infrastructure in the ACR (2007), 2nd according to the DBR (2007) and firms' perceptions, 19th (out of 23) according to the impact of the INF variables on the average log productivity, and 18th (out of 23) with respect to the allocative efficiency in logs. In these cases the correlation with firm growth is negative, signaling again that infrastructure quality is an important growth bottleneck in Africa.

Egypt and Morocco are interesting cases. Both countries show a relatively high quality of infrastructure according to the ACR and perception rankings, and the results of the econometric analysis confirm this. Egypt is 4th and Morocco is 5th in the rankings based on demeaned aggregate productivity, and both have one of the lowest contributions of infrastructure to TFP.

Countries with the poorest infrastructure quality, according to ACR (2007) and firms' perceptions, are Cameroon, Burkina Faso, Niger, and Ethiopia. Once again this is consistent with the econometric analysis done for these countries, showing a great negative influence of infrastructure on productivity. Cameroon and Burkina Faso are ranked among the countries with the highest contribution of infrastructure to average productivity. The influence of infrastructure on Ethiopia's manufacturing sector productivity is also very high, with a relative contribution equal to 52.6 percent of the total IC effect.

The following set of figures provides some additional evidence on the relation between measures of countries' economic performance and TFP based on our econometric analysis. The previous conclusions become more apparent by looking at the cross-plots. Figure 5.7 shows a clear positive correlation between GDP per capita and demeaned aggregate productivity, with a correlation coefficient equal to 0.81. Notice that this positive relationship has a decreasing dispersion as per capita income grows; that is, for those African countries with a per capita GDP lower than 10 percent of that of the United States, demeaned aggregate productivity presents a more heterogeneous behavior. The conclusion obtained from figures 5.8 and 5.9 are similar. Those countries that are high in the ranking based on ease of doing business (DBR) also have a large demeaned aggregate productivity, with a correlation coefficient of 0.77. The same is true for the positive relationship between the quality of overall infrastructure and the ACR (2007)—

the more productive the manufacturing firm is, the higher the contribution of overall infrastructure quality to TFP (correlation coefficient equal to 0.76).

A question of interest is whether those countries with the lowest demeaned aggregate productivity levels are also those with the greatest impact of infrastructure on firm's perceptions, on average productivity, and on allocative efficiency. Figures 5.10 and 5.12 provide clear answers to these questions. Figure 5.10 shows the negative correlation between the mixed demeaned aggregate productivity and firm's perceptions of growth bottlenecks, with a correlation coefficient of 0.76. The absolute contribution of infrastructure to both average log productivity and average productivity via simulations decreases as the demeaned aggregate productivity increases. This relation is stronger in the case of the absolute percentage contribution to the average log productivity since the corresponding coefficients of correlations are -0.60 (figure 5.11) and -0.49 (figure 5.12).

[FIGURES 5.7 TO 5.12 ABOUT HERE]

Figures 5.13 and 5.14 show the linear correlations between demeaned aggregate productivity and percentage absolute contribution to allocative efficiency TFP in logs and allocative efficiency via simulations, correspondingly. There is a negative relation in both figures. But the linear correlation is smaller in the case of the average log productivity (correlation coefficient equal to -0.31) than in the case of allocative efficiency with TFP in levels (correlation coefficient equal to -0.49).

Finally, figures 5.15 and 5.16 show the strong linear positive relation between the infrastructure contributions to the two components of the O&P decomposition based on TFP in logs and TFP in levels. Their corresponding coefficients of correlation are 0.69 and 0.77, respectively.

[FIGURES 5.13 TO 5.16 ABOUT HERE]

All effects of infrastructure are not supposed to be negative in all cases. There are positive factors intended to stimulate productivity and economic activity, such as the use of ICT or a firm's own electricity generator. A question of interest is to what extent the impacts listed in table 5.1 are due to positive factors that enhance economic performance or negative factors that constrain economic activity. Since the absolute percentage contributions are constructed based on absolute values, at this point we still cannot say anything about the direction of the effect of infrastructure on economic performance. But it is easy to analyze the effect of the individual INF factors, and this is one of the aims of the following section.

6. Country-by-country results

In the preceding section we evaluated the relative weight of infrastructure among IC variables. In this section the objective is to present a summary of the main results, country by country, focusing on the impact of the individual infrastructure factors or variables. We measure the strength of infrastructure's impact on TFP through three different procedures: (a) elasticities or semi-elasticities, (b) simulations, and (c) evaluation of the IC regressions impact on the sample mean of the variables.

These three sets of results provide complementary information. The elasticities and semi-elasticities measure the impact of a change in an independent variable (infrastructure factors or other IC and C variables) on the dependent variable (productivity). But elasticities and semi-elasticities are not directly comparable.

On the other hand, simulations measure how the dependent variable changes from scenario A, in which the infrastructure and other IC and C factors are as observed by the survey, to scenario B, in which one of the infrastructure factors improves by, say, 20 percent. From this we can make the following assertions: *If the number of power outages suffered by firms in country* X *is reduced by 20 percent, then the average productivity (or the allocative efficiency) could increase by* Y *percent, holding everything else constant*.

Finally, the evaluation at the sample means of the regression variables, as opposed to the simulations, is a static exact decomposition of the terms of the mixed O&P decomposition. We can evaluate the contributions of all the INF, IC, and C factors to the sample mean of average log productivity, identifying the relative importance of each infrastructure variable (for example, losses due to the number of water outages or transport failures) in net terms or in absolute terms.

The next subsection focuses on the results of each of the 23 countries. The results are presented in a series of country-specific figures 6.1–6.23. The figures also report the results for the productivity equation. The first panel of each figure shows the elasticities and semi-elasticities; the second and third panels focus on the relative contribution of the INF variables to the average log productivity and to the allocative efficiency in logs; and the fourth and fifth panels present the results of the simulations (that is, how much the average productivity and the allocative efficiency would increase if we improve the INF variables).

Note that the results on the elasticities and semi-elasticities are not comparable since they use different measurement scales. For purposes of comparison we should rely on the simulations and on the results of the contributions to the average.

In the interest of space, we focus only on the major results for each country.

[FIGURES 6.1 TO 6.23 ABOUT HERE]

6.1 Infrastructure impacts on TFP by country

High-growth countries in southern Africa

Mauritius (MUS). Mauritius is the top-ranked in terms of per capita GDP and demeaned aggregate productivity, and the second according to the DBR (2007) and firms' perceptions of the quality of infrastructure (table 5.1). The relative contribution of infrastructure to average log productivity is 26.6 percent—one of the lowest among the African countries considered (see figure 5.5, panel B, and figure 6.14, panel B). The most important constraint on productivity comes from the number of days to clear customs for exports (17 percent on average) (see figure 6.14); 5.2 percent is due to the use of IC technologies (positive factor); and low-quality provision of electricity and water accounts for only 2 percent of the average log productivity.

Swaziland (SWZ). Swaziland is ranked second in terms of per capita GDP (table 5.1). Productivity of firms is negatively affected by shipment losses in customs, the number of power outages, and the average duration of transport by road (see figure 6.20). These results are common to Namibia and Botswana since the countries are pooled together for estimation. The use of generators has a positive sign, meaning that it stimulates productivity. Country-specific results for Swaziland show that the largest contribution to average log productivity comes from problems in customs during the export process (10 percent), and from the number of power outages (9 percent).

South Africa (ZAF). South Africa is ranked third based on per capita GDP and demeaned aggregate productivity (table 5.1). Productivity is negatively affected by the days to clear customs to import, the sales lost due to power outages, the number of water outages, the time waiting for an electricity supply, and sales lost due to delivery delays. Therefore, the low quality of the customs services, electricity services, and water affects productivity performance at the firm level in South Africa. The contribution to average log TFP of electricity provision is 6.9 percent, and the contribution of water provision is 5.7 percent. Time wasted in customs while importing accounts for 9.4 percent of the average log productivity. Lastly, problems in transport services represent 5.7 percent of average log productivity (see figure 6.19).

Botswana (BWA). Botswana is ranked fourth based on per capita income (DBR, 2007) and on perceptions and sixth in terms of demeaned aggregate productivity. The productivity of firms located in Botswana is affected by shipment losses in customs while exporting (negative), power outages (negative), the percentage of electricity that comes from firm's own generators (positive), and the average duration of transport by road (negative) (see panel A of figure 6.3 and panel B of figure 5.5). These marginal effects are common to Namibia and Swaziland since the countries are pooled together for estimation. Country-specific results show that the largest contributor to both average log productivity and to allocative efficiency is shipment loss in customs while exporting (panel B). Simulations show that the largest productivity improvement comes from reduction in power outages (panel D of figure 6.3). That is, according to our simulations, if the number of power outages suffered by firms in Botswana were reduced by 20 percent, average productivity could increase by 2.1 percent.

Namibia (NAM). Namibia is ranked sixth in terms of per capita GDP and ACR (2007) (table 5.1). Productivity of firms is negatively affected by shipment losses in customs while exporting, the number of power outages, and the average duration of transport by road (see figure 6.16). These results are common to Botswana and Swaziland since the countries are pooled together for estimation. Country-specific results show that the impact of infrastructure on the productivity of manufacturing firms in Namibia mainly comes from problems in customs while exporting—this factor represents 9 percent of the average log productivity. Problems from electricity provision (power outages) and from use of alternative power infrastructure (such as a generator) represent 2.2 percent and 3 percent of the average log productivity, respectively, in absolute terms (figure 6.16).

High-growth countries in North Africa

Algeria (DZA). **Algeria is ranked fifth in terms of per capita GDP, and seventh in terms of firm perceptions and demeaned aggregate productivity**. The results on the **productivity impact** of infrastructure (see figure 6.1) show the total effect on absolute value is as large as 48.6 percent of average log TFP. The quality of infrastructure variables affecting TFP are: cost of exports, having an own generator, number of power outages, losses due to water outages, having an own well, the cost of water from the public system, having e-mail, and low-quality supplies. The largest and most positive effect comes from having e-mail, which could represent 14.5 percent of average log TFP.

Egypt (EGY). Egypt is ranked seventh based on per capita GDP, sixth in terms of ACR (2007), fourth in terms of demeaned aggregate TFP, and third in firms' perceptions of infrastructure quality. From the econometric analysis, the contribution of infrastructure to the average log productivity in Egypt is only 26 percent. The main infrastructure factors affecting firms' productivity are the average duration of water and power outages (both with negative effects), the percentage of firms with their own generator (positive effect), the dummy for own transportation (positive), shipment losses in exports (negative), and days of inventory of the main supply (negative) (figure 6.6).

Morocco (MAR). The perceptions of the managers of the Moroccan firms suggest that infrastructure is not a major concern when compared to other IC constraints; it is ranked first in table 5.1. But according to figure 6.15, the contribution of infrastructure to average log productivity is 31.3 percent, with the largest impacts coming from (a) the average time to clear customs to import, and (b) the time wasted to obtain a phone connection. A 20 percent reduction of average customs delays for imports could increase average productivity by 1.6 percent and allocative efficiency by 0.4 percent. Notice, that the ranking based on the econometric analysis (demeaned aggregate productivity) is consistent with the ranking based on per capita GDP (see table 5.1), which establishes that Morocco is in eighth position, not first.

Low-growth countries in central-west Africa

Cameroon (CMR). Cameroon is rank ninth in term of per capita GDP, which is somehow surprising if we compare it with the results of the rest of the rankings based on DBR (2007), ACR (2007), firms' perceptions, and demeaned aggregate productivity. We found that the productivity of manufacturing firms in Cameroon is reduced by the following factors: number of days required to clear customs for imports, average duration of power and of water outages, shipment losses, and time waiting for a phone connection (panel A of figure 6.5). These factors are common to Mauritania, Burkina Faso, and Niger since the countries are pooled together for estimation purposes. Country-specific results show that the largest contributions to average log productivity come from the number of days waiting to clear customs, duration of power outages, and from the time waiting for a phone connection (panel B of 6.5). Infrastructure represents 41.6 percent of average log TFP.

Mauritania (MRT). Mauritania ranks 10th in terms of per capita GDP, 13th in term of the total absolute contribution of infrastructure to average log productivity, 17th in terms of firm perceptions, and 15th in terms of demeaned aggregate productivity. These results are common to Burkina Faso, Cameroon, Niger, Ethiopia, and Eritrea since those countries are pooled together for estimation purposes. Delays in customs while importing represent 22 percent of average log productivity and 12 percent of allocative efficiency. In terms of simulation, a 20 percent improvement in this variable could cause a 13.9 percent increase in average productivity and a 7.8 percent one in allocative efficiency. Low-quality provision of electricity and water and its indirect costs also reduce average productivity in Mauritania (see figure 6.13).

Senegal (SEN). Senegal reveals a high infrastructure impact on the TFP of manufacturing firms. The percentage contribution of infrastructure to the average log productivity of this country is 58.5 percent; the indirect costs stemming from the low-quality provision of electricity represent 9.3 percent of this (see figure 6.18). The use of own-power infrastructure partially alleviates the negative impact of the low quality of electricity provision. The relative importance of problems in transport services (such as low-quality supplies) is very high; this variable represents 23.4 percent of average log productivity and 14 percent of allocative efficiency. A 20 percent reduction in the percentage of low-quality supplies received may cause a 3.2 percent increase in average productivity. Notice that the simulation of a 20 percent improvement in the percentage of low-quality supplies received causes a decrease in allocative efficiency of -0.4 percent. The reason for this phenomenon is clear: the allocative efficiency is simply the covariance between productivity at the firm level and share of sales. Therefore, a negative rate of change of the allocative efficiency indicates that the firms receiving a larger share of low-quality supplies are those with the largest market shares.

Benin (BEN). Benin is ranked 12th in per capita GDP, 11th in terms of demeaned aggregate productivity, and 13th in terms of DBR (2007). The time waiting for phone connections and to clear customs in order to export are the two factors that most negatively contribute to average log productivity (see panel B of figure 6.2). An independent 20 percent improvement in these two variables could increase average productivity by 3.8 percent and 4.3 percent, respectively (panel D of figure 6.2). The same holds for the allocative efficiency term (panels C and E).

Mali (MLI). Mali is ranked 14th in per capita GDP terms, in demeaned aggregate productivity, and in the ACR (2007). The total contribution of infrastructure to average log productivity in Mali is 42.7 percent. The low quality of electricity, water, and phone provision accounts for almost 32 percent of the average log productivity. The use of firms' own roads is a factor that increases productivity (figure 6.12).

Burkina Faso (BFA). Burkina Faso ranks 16th and 12th in terms of per capita GDP and demeaned aggregate TFP, respectively. The main infrastructure problems are clearing customs while importing, the average duration of power and water outages, shipment losses, and time waiting to obtain a phone connection. These results are common to Mauritania, Cameroon, and Niger since these countries are pooled for estimation. In particular, for Burkina Faso, all of these factors reduce productivity at the firm level (see figure 6.4) and can contribute to 35 percent of average log TFP.

Niger (NER). Niger is one of the poorest countries in our sample: it ranks 19th based on per capita income, and managers' perceptions show a great concern regarding quality of current infrastructure (20th in the rank). The absolute contribution of infrastructure to average log productivity is 34.7 percent, with 20.7 percent due to problems clearing customs while importing, 9.4 percent due to the average time wasted in obtaining a phone connection, and 4.5 percent due to the cumulated negative effect of the low-quality provision of electricity and water and the poor transport system (see figure 6.17).

Low-growth countries in central-east Africa

Kenya (KEN). Kenya ranked 13th in per capita GDP, 6th in terms of DBR (2007), and 8th in terms of demeaned aggregate productivity. The results of the productivity equation in Kenya show multiple interrelationships between productivity at the firm level, on one hand, and infrastructure, on the other. Factors such as the use of generators, the cost of electricity from the public grid, water outages, sales lost due to power outages, and so on, reduce productivity at the firm level (see figure 6.9). Although the contribution to the average log productivity of any of these factors is over 6 percent, infrastructure as a whole accounts for 30 percent of average log productivity, and for 19.9 percent of allocative efficiency (which illustrates the major influence infrastructure has on Kenyan firms' productivity).

Uganda (UGA). Uganda is ranked 15th in per capita GDP and 17th in demeaned aggregate productivity. Uganda's average log productivity is strongly influenced by infrastructure conditions, representing 58.4 percent. The two main factors affecting average productivity are the time to clear customs and the provision of electricity (figure 6.22, panel D).

Zambia (ZMB). Zambia ranks 17th in terms of per capita GDP, 14th in terms of firm perceptions, and 18th in terms of the ACR (2007) .The contribution of the cost of electricity from the public grid to average log productivity in Zambia is 32.5 percent. The contribution of the average duration of power outages to average log TFP is 9.1 percent (see figure 6.23). The total contribution of infrastructure to average TFP is 50.6 percent.

Tanzania (TZA). Tanzania is 18th based on per capita GDP, 14th in terms of DBR (2007), and 15th in terms of firm perceptions. The absolute contribution of infrastructure to average log productivity in Tanzania is 34.1 percent. Out of this, 14.8 percent is due to time wasted waiting for water supply and 5.5 percent is due to the number of transport outages (figure 6.21).

Malawi (MWI). Malawi is ranked 20th in per capita GDP terms and 22nd in terms of ACR (2007). The econometric evidence shows that the aggregate productivity of Malawian's manufacturing firms are dramatically affected by infrastructure quality (83 percent of average log TFP). Delays in clearing customs while importing account for 25 percent of the average log productivity. A 20 percent improvement in this variable could increase the average productivity by 6 percent and the allocative efficiency by 1.5 percent (see panels D and E of figure 6.11). Symmetrically, productivity decreases as the number of power outages increases; the percentage contribution of this variable to average log productivity is 9.2 percent and to the allocative efficiency is 3.5 percent. The cost of electricity from firms' own generators is another antiproductive factor. Water provision also impacts the productivity of Malawian firms (figure 6.11).

Madagascar (MDG). Madagascar is ranked 21st in per capita GDP terms, and 16th and 17th in terms of demeaned aggregate productivity and DBR (2007), respectively. How infrastructure may impact firm-level productivity is clear from the results obtained in figure 6.10. The factors related to electricity supply are intimately linked to productivity. Water costs and the number of phone outages also reduce productivity at the firm level. The total contribution of infrastructure to average log productivity in Madagascar is 31 percent and to allocative efficiency is 28 percent.

Ethiopia (ETH). Ethiopia is ranked 22nd in terms of per capita GDP and 21st in terms of firms' perceptions. The productivity of Ethiopian manufacturing firms is negatively affected by the days to clear customs to import, the cost of electricity from the public grid, shipment losses while in transit, and the percentage of supplies that are of lower than agreed-upon quality. Positive effects on productivity come from the percentage of electricity from firms' own generators and from the days of inventory of the main supply. These results are common to Eritrea since the countries are pooled together for estimation. The largest contribution to the average log productivity is by the days of inventory of main supply, days to clear customs to import, the cost of electricity from the public grid, and the electricity from firms' own generators (figure 6.8). In total, infrastructure represents 32 percent of average log TFP, which seems too low. This might indicate that pooling with Eritrea for estimation of IC elasticities might not be a good idea in this case.

Eritrea (ERI). Eritrea is the last country in the ranking, in 23rd position in terms of per capita GDP. The total number of days waiting to clear customs to import, the cost of electricity from the public grid, shipment losses, and low-quality supplies are the factors with negative effects on firms' productivity (see figure 6.7). Remember that these results are common to Ethiopia since the countries are pooled together for estimation. The cost of energy from the public grid accounts for a 22.9 percent of average log productivity, almost 50 percent more than the second factor (days to clear customs to export) in order of importance. In total, infrastructure represents 48 percent of average log TFP.

6.2 Summary of the main empirical results

The aim of this subsection is to summarize the main empirical results obtained, country by country, including the absolute percentage contribution of the infrastructure group of variables to the sample means of productivity in logs. The *customs clearance* subgroup includes those variables related to the ease or difficulty of clearing customs when exporting or importing. Within *provision of electricity* we have grouped all the variables related to low-quality provision of electricity (number of power outages, power fluctuations, cost of electricity, and so on). *Use of power infrastructure* is intended to enclose all the variables related to the use of alternative sources of energy, such as generators. Similarly, the subgroups *provision of water* and *provision of phone connections* includes all the variables related to the quality of the provision of these utilities, whereas *use of water infrastructure* includes the use of firms' own wells or boreholes to replace the public provision of water. Obviously, *use of ICT* takes into account the use of ICT in firms' commercial operations. Lastly, *transport services* contains all variables relating to the quality of transportation services, such as shipment losses in transit, transport delays, delivery delays, and so on. *Own transport infrastructure* includes the use of own transportation for products or workers.

Figure 6.24 shows the prominent influence of low-quality electricity provision on average log productivity in the different countries considered in this report. The percentage absolute contribution of this group of variables to average log productivity ranges from 34.1 percent in Zambia to 0.3 percent in Morocco, being a negative effect in all cases. Only in Tanzania was the low quality of electricity provision not statistically significant, probably due to the significant and very influential effect of water provision in this country. The low quality of electricity and the continuous outages are partially alleviated by the use of own-power infrastructure, as the positive effect of the group *use of power infrastructure* shows.

Another group of variables with a statistically high impact on average log productivity is *customs clearance*. The contributions of this group are negative and very large in most countries, indicating a clear and pervasive constraining effect of the time wasted in customs when importing or exporting.

[FIGURES 6.24 ABOUT HERE]

Regarding *provision of water*, the relative importance of this group of variables is lower when compared to the *provision of electricity*; nevertheless, there are some cases (such as Tanzania, Mali, and Kenya) where the contribution of this group of variables is very high, even compared to the provision of electricity. As with the provision of electricity, the use of alternative water infrastructure such as boreholes or wells has a positive impact on plants' productivity—an effect that only appears to be significant when there is a negative effect of water provision, suggesting the existence of a replacement effect between the public provision of water and alternative supplies of water.

The poor quality of phone provision is negatively related to productivity in 14 countries; nevertheless, the quantitative contribution of this group of variables is, in general, lower than the impact of the electricity provision. Benin is an exception. The contribution to the average log productivity of telephone provision in this country is 40.8 percent—more than 20 times the contribution of electricity provision, which is only 2 percent. The use of IC technologies is positively related to productivity, but the use of these technologies was only significant in the productivity regressions of six countries: Malawi, Algeria, Tanzania, Kenya, Morocco, and Mauritius, with the largest impacts in Algeria (14.5 percent) and Mauritius (5.2 percent).

Problems with product transport are negatively related to productivity in all the cases, with the exception of Botswana, Swaziland, and Namibia, for which no variables for this group were significant in productivity regressions. The largest impacts of this subgroup of variables were seen in Senegal, Tanzania, Madagascar, South Africa, and Zambia; nevertheless, the contributions of transport services to average log productivity are relatively lower than the impact of the *provision of electricity* or the *customs clearance* subgroups. On the other hand, the use of own-transport infrastructure stimulates productivity growth; in all the cases in which any variable belonging to this subgroup was significant in the productivity regressions, it appeared with a positive sign. But the positive effects on productivity of these factors were concentrated in only seven countries: Malawi, Benin, Senegal, Eritrea, Tanzania, Kenya, and Egypt.

Figure 6.25 provides similar information, but, in this case, we have grouped the different infrastructure factors in only five groups: *customs clearance, electricity, water, telecommunications and ICT,* and *transportation*. From this figure the high influence of *electricity* factors become even clearer. When we include the *provision of electricity* and *use of generators* or *power infrastructure* into a single group, the resulting block of electricity factors can explain more than half of average log productivity in Uganda, Ethiopia, Zambia, Eritrea, Swaziland, and Botswana. The water group is relatively important in Tanzania, Kenya, and

Egypt. The customs clearance group gains importance in those countries with a more patent export orientation, such as Mauritius, Botswana, and Namibia. Finally, as has been signaled, the transportation subgroup explains more than half of the whole infrastructure impact on average log productivity in Senegal and Madagascar.

[FIGURES 6.25 ABOUT HERE]

Figure 6.26 reports the absolute percentage contribution of infrastructure by key factors via simulations. The results are fully consistent with the ones provided by figure 6.24. The relative weights of the electricity factors dominate in more than half the countries, and the water and telecommunications and ICT subgroups tend to play a secondary role in explaining average productivity when compared to electricity. The main difference with respect to figure 6.24 is the lower relative weight of the customs clearance group. Once again, it should be pointed out that the information provided by the simulations complements the results from the evaluation of the sample average of log productivity. In this case, we are talking about a cumulative effect, all other things being equal, since we evaluate the change in the average productivity when one of the INFs changes.

[FIGURES 6.26 ABOUT HERE]

The summary of results is complemented by the analysis of allocative efficiency. Figures 6.27 and 6.28 show that the impact of infrastructure on allocative efficiency is equally distributed among the different infrastructure factors. The impact of water is, in this case, larger than in the case of average productivity, while the impact of energy provision is considerably less—gaining relative importance with the use of own generators. The positive effect of the use of own generators on allocative efficiency indicates that those firms that accumulate a larger proportion of market sales are also the firms that use their own generators. The same holds for the use of IC technologies and the use of own-water infrastructure. The customs clearance group has important implications for the allocative efficiency of Namibia, Mauritania, and Botswana. Finally, the quality of transport services plays a secondary role in explaining the behavior of allocative efficiency in the different countries included in our sample.

[FIGURES 6.27 ABOUT HERE]

Figure 6.28 organizes the different subgroups of infrastructure factors into five key groups. From this figure, the important contribution of the electricity subgroup becomes even clearer. Transportation explains more than 50 percent of the allocative efficiency in logs of

Morocco. Once again, the relative importance of the water, telecommunications, and ICT subgroups is lower when compared to the contribution of electricity and customs clearance.

[FIGURES 6.28 ABOUT HERE]

7. Conclusions

For Africa's awaited growth resurgence to occur, a broad range of factors—political, institutional, and economic—must be improved. The World Bank's landmark Africa Competitiveness Reports (2004 and 2007) focus on problems that inhibit economic growth. It is agreed that improving Africa's infrastructure is a crucial step toward penetrating international markets and meeting the goals of continuous growth and poverty reduction.

Infrastructure quality has a pervasive influence on all areas of an economy. Low-quality infrastructure and limited transport and trade services increase logistical and transaction costs, rendering otherwise competitive products uncompetitive, as well as limiting rural production and people's access to markets—with adverse effects on economic activity and poverty. A large number of empirical studies illustrate the impact of infrastructure on economic performance. All suggest that Africa's infrastructure gap is an important growth bottleneck, with a negative impact on productivity and the overall competitiveness of the region. Using the methodology of Escribano and Guasch (2005, 2008) and Escribano et al. (2008), several studies have found empirical evidence—in cases such as Brazil, Chile, Costa Rica, Mexico, Turkey, and several Southeast Asian countries—that improvements in investment climate conditions in general, and in infrastructure quality in particular, may lead to important gains in productivity and in other economic performance measures: employment, real wages, exporting activities, and inflows of foreign direct investment (FDI).

Approach and methods

This paper provided a systematic, empirical assessment of the impact of infrastructure quality on the TFP of African manufacturing firms. We applied microeconometric techniques to investment climate surveys of 26 African countries to gauge the impact of infrastructure quality on TFP.

For each country we estimated, by regression techniques, the impact of infrastructure quality based on 10 different productivity measures and showed that the results were robust once we controlled for other observable fixed effects (red tape, corruption and crime, finance, innovation and labor skills, and so on) obtained from the investment climate surveys (see Escribano and Guasch, 2005, 2008).

We pooled data from the investment climate surveys only for the few African countries for which we did not have sufficient observations for estimation purposes. Otherwise, we performed a country-by-country estimation to reveal firm and industry information by country. After pooling the data from several countries, the econometric results were then suitably disaggregated following the method of Olley and Pakes (1996) (Escribano et al., 2008), which allowed us to make country-specific evaluations of the impact of investment-climate quality on aggregate TFP, average TFP, and allocative efficiency.

We ranked the African countries in the study according to several aggregate indices: per capita income, ease of doing business, firm perceptions of growth bottlenecks, and the recent concept of *demeaned productivity (demeaned TFP)*, which overcame the problem of comparing apples and oranges when doing TFP cross-country comparisons (Escribano et al., 2008). We found the concept of demeaned productivity very useful because it is highly correlated with per capita income, ease of doing business indices, firm's perceptions of growth bottlenecks, and the results of the Africa Competitiveness Reports. Furthermore, the information obtained from the investment-climate determinants of demeaned TFP provided a much deeper insight into the firm-level investment-climate infrastructure elements that are constraining productivity growth in African countries.

We distinguished two clear blocks of countries in Africa.

The first block comprised countries with faster, steadier growth rates. These are mainly in the south, including Mauritius, Swaziland, South Africa, Botswana, Namibia, and Lesotho. The block also included Algeria, Morocco, and Egypt from the north, and Cape Verde and Cameroon from Central Africa. In southern Africa Botswana, Namibia, and Swaziland, were pooled for estimation purposes.

In the second block were Mauritania, Senegal, Benin, Mali, Niger, and Burkina Faso in the central-west; and Uganda, Kenya, Zambia, Tanzania, Malawi, Burundi, Madagascar, Ethiopia, and Eritrea in the central-east. These countries have experienced lower and more irregular growth rates, with periods of both positive increase and persistent decrease in per capita income. Pooled for estimation purposes were the West African states (Burkina Faso, Cameroon, Mauritania, and Niger) and Eritrea and Ethiopia.

Out of the 26 African countries analyzed, Mauritius was, in 1950, the country with the highest per capita income (measured in terms of per capita gross domestic product, GDP), followed closely by South Africa, and, by a larger gap, Namibia and Algeria. But the per capita income levels in 2003 were somewhat different; Mauritius was still ranked first, followed by Swaziland, South Africa, and Botswana—and, by a wider gap—Algeria, Cape Verde, Egypt, and Morocco. Mauritius and Botswana experienced the highest sustained per capita income growth during recent years. Lesotho's rate is the median, splitting the study into two blocks.

To better understand the convergence or divergence of trends, we plotted the per capita income of each African country relative to the per capita income of the United States. Convergence was observed only in Mauritius, Swaziland, and Botswana. For all other study countries, including South Africa, per capita income was found to be diverging from the United States, while, in a few (Egypt, Morocco, and Cape Verde) the ratio was stable. While persistently positive GDP growth allowed Mauritius's per capita income to reach 45 percent of the United States' in 2003, this was clearly the exception (together with Swaziland and Botswana). For the rest of the countries, including South Africa, relative per capita income was much lower in 2003 than in 1960 (indicating a divergence). In fact, the 2003 per capita income of several countries was no larger than 5 percent of the per capita income of the United States. As expected, labor productivity was the main factor explaining this divergence, given that labor force participation has a steady influence. Since total factor productivity (TFP) is usually a key factor explaining the evolution of labor productivity, in this paper we used investment climate surveys to identify the main infrastructure-related TFP bottlenecks in Africa.

The per capita income rankings were correlated with the rankings obtained from the World Bank's 2007 *Doing Business* report. In 2007 Mauritius, Swaziland, South Africa, Botswana, and Namibia ranked 32nd, 76th, 29th, 48th, and 42nd in the world based on the ease-of-doing-business indicators. This index considers questions such as the number of days required to start a business and the ease of dealing with licenses, registering a property, trading across borders, employing workers, and so on. Other 2007 rankings include 83rd for Kenya, 97th for Ethiopia, 165th for Egypt, and 170th for Eritrea.

We showed the percentage of firms that *perceived* telecommunications, electricity, customs clearance, and transport as major obstacles to their economic performance. Only in Benin, Kenya, and Zambia did more than 50 percent of firms identify telecommunications as a severe obstacle. Meanwhile, the quality of electricity provision is a major problem for more than 50 percent of firms in more than half of the countries in our sample. In Burundi, Cameroon, Benin, Burkina Faso, and Cape Verde, the percentage of firms considering electricity as a severe or very severe obstacle exceeded 80 percent; on the other hand, only 20 percent of firms in Morocco, South Africa, Botswana, and Namibia considered electricity a severe obstacle. Customs clearance was considered an acute problem in Benin, Kenya, Madagascar, Senegal, and Algeria. Finally, transportation was considered a severe obstacle by more than 70 percent of firms in Burkina Faso and Benin.

The World Bank's 2007 Africa Competitiveness Report evaluated a wide range of factors related to economic activity, with infrastructure among them. Once again there were clearly different performance levels across the two blocks of countries. While in Namibia, South Africa, Botswana, Egypt, and Morocco, the quality of infrastructure exceeded the approval level, in the remaining countries this quality was rated low in most cases. The same held for the disaggregated results, including the number of telephone lines and the quality of ports, air transport, and electricity supply.

The difference between the two blocks becomes even more apparent when looking at the cross-plots between GDP per capita relative to the United States and firms' perceptions. A preliminary analysis of the cross-plots suggests two points: first, that there is an intuitive and negative relation between income level and infrastructure constraints; and, second, that the division of the two blocks of countries remains intact, showing now the largest dispersion in the constraint perceptions for the lowest per capita income group.

Findings

Among the countries of the high-income-growth block, infrastructure has a low impact on TFP (see panel B of figures 5.5 and 5.6 and panel A of figure 6.24). Red, tape, corruption, and crime dominate over infrastructure in countries such as Mauritius, Egypt, and South Africa (figure 5.5, panel B). Infrastructure quality has a high impact on TFP in the countries of the low-income-growth block (see panel B of figures 5.5 and 5.6 and panel A of figure 6.24), but the impact is very negative (see panel B of figures 5.3 and 5.4), identifying important bottlenecks for TFP growth.

We found much heterogeneity among individual infrastructure factors affecting countries in both the high- and low-growth blocks (see figure 6.26). Among related factors that most influence the average productivity TFP of African firms are:

- Poor-quality *electricity* provision, which affects mainly poor countries, such as Eritrea, Ethiopia, Mali, Senegal, Uganda, Zambia, and Kenya. It also affects countries that are growing faster, in relative terms, such as Botswana, Namibia, and Swaziland.
- Problems dealing with *customs* during importing or exporting affects mainly fast-growing countries, such as Mauritius, Morocco, and Swaziland. But low quality of customs also affects slow-growing countries, such as Niger, Mauritania, Cameroon, Malawi, Burkina Faso, and others.
- Losses from *transport interruptions* affect mainly slower-growing countries, such as Madagascar, Kenya, Tanzania, and Senegal.
- *Water outages* affect mainly slower-growing countries, such as Tanzania, Kenya, Burkina Faso, Mauritania, Niger, and Mali. But it also affects some of the faster-growing countries, such as Egypt.

Of the infrastructure determinants that most influence the *allocative efficiency* of African firms there is also some heterogeneity across countries.

- Poor-quality *electricity* provision affects the allocative efficiency of mainly poor countries, such as Zambia, Mali, Uganda, Eritrea, and Kenya.
- Problems dealing with *customs* while importing or exporting affects mainly slow-growing countries, such as Mauritania, Niger, and Cameroon. But it also affects the allocation efficiency of countries that are growing fast, such as Morocco, Namibia, and Mauritius.
- *Transport services* affects the allocative efficiency of mainly slower-growing countries, such as Madagascar, Senegal, and Tanzania.
- *Water provision* affects the allocative efficiency of mainly slower-growing countries, such as Tanzania, Kenya, and Mali.

References

Agenor, P. R., and B. Moreno-Dodson (2006): "Public Infrastructure and Growth: New Channels and Policy Implications," Policy Research Working Paper 4064, *World Bank*, Washington DC.

Artadi, E. V., and X. Sala-i-Martin (2003): "The Economic Tragedy of the XXth Century: Growth in Africa," *National Bureau of Economic Research* (NBER) Working Paper 9865, Cambridge, MA.

Aschauer, D. A. (1989): "Is Public Expenditure Productive?" *Journal of Monetary Economics* 23(2): 177–200.

Bougheas, S., P. Demetriades, and E. Morgenroth (1999) "Infrastructure, Transport Costs and Trade," *Journal of International Economics* 47: 169–89.

Brunel, S. (2004) "L'Afrique—Un Continent en Réserve de Développement," *Editions Bréal*, Rosny-sous-Bois.

Calderon, C., and L. Serven (2003): "The Growth Cost of Latin America's Infrastructure Gap," In William Easterly and Luis Serven (eds.), *Adjustment Undetermined? Public Deficits and Growth in Latin America, 1980–2000*. Princetown, N.J.: Princetown University Press.

——— (2005): "The Effects of Infrastructure Development on Growth and Income Distribution," Mimeo, World Bank, Washington, DC.

Calderon, C., W. Easterly, and L. Serven (2003a): "Infrastructure Compression and Public Sector Solvency in Latin America," In William Easterly and Luis Serven (eds.), *Adjustment Undetermined? Public Deficits and Growth in Latin America, 1980–2000*. Princetown, N.J.: Princetown University Press.

——— (2003b): "Latin America's Infrastructure in the Era of Macroeconomic Crisis," In William Easterly and Luis Serven (eds.), *Adjustment Undetermined? Public Deficits and Growth in Latin America, 1980–2000*. Princetown, N.J.: Princetown University Press.

Canning, D. (1998) "A Database of World Infrastructure Stocks, 1950–95," Policy Research Working Paper 1929, World Bank, Washington, DC.

Center for International Comparisons of Production, Income and Prices (2009) "Penn World Table," http://pwt.econ.upenn.edu/cic_main.html.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977): "Maximum Likelihood Estimation From Incomplete Data Via the EM Algorithm," *Journal of Royal Statistical Society* B39: 1–38.

Dessus, S., and R. Herrera (1999): "Public Capital and Growth: An Econometric Study on Panel Data," *Revue Economique* 50(1): 113–26.

Easterly, W., and S. Rebelo (1993): "Fiscal Policy and Economic Growth," *Journal of Monetary Economics* 32: 417–58.

Economic Commission for Africa (ECA) (2005): "Assessing Regional Integration in Africa," ECA Policy Research Paper.

Escribano, A., and J. L. Guasch (2008): “Robust Methodology for Investment Climate Assessment on Productivity: Application to Investment Climate Surveys from Central America,” Working Paper # 08-19 (11), Universidad Carlos III de Madrid.

——— (2005): “Assessing the Impact of Investment Climate on Productivity Using Firm Level Data: Methodology and the Cases of Guatemala, Honduras and Nicaragua.” Policy Research Working Paper 3621, World Bank, Washington, DC.

Escribano, A. and J. Pena (2008): “Empirical Econometric Evaluation of Alternative Methods of Dealing with Missing values in Investment Climate Surveys,” Unpublished manuscript, Universidad Carlos III de Madrid.

Escribano, A., J. L. Guasch, M. de Orte and J. Pena (2008a): “Investment Climate Assessment Based on Demean Olley and Pakes Decompositions: Methodology and Applications to Turkey’s Investment Climate Survey,” Working Paper # 08-20 (12), Universidad Carlos III de Madrid.

Escribano, A., J. L. Guasch, M. de Orte and J. Pena (2008b): “Investment Climate and Firm’s Performance: Econometric and Applications to Turkey’s Investment Climate Survey,” Working Paper # 08-20 (13), Universidad Carlos III de Madrid.

Escribano, A., J. L. Guasch, M. de Orte and J. Pena (2009): “Investment Climate Assessment in Indonesia, Malaysia, Philippines and Thailand: Results from Pooling Firm Level Data,” Singapore Economic Review, Special Issue on the Econometric Analysis of Panel Data, Vol. 54, #3, August 2009.

Escribano, A., N. Peltier, L. Garrido, and H. Sing (2005): “The Impact of Infrastructure on Competitiveness in Latin America: A Firm Level Analysis Based on Investment Climate Assessments,” Paper presented at the Joint Conference of the Inter-American Development Bank (IADB) and the World Bank (WB), June 6–7.

Esfahani, H. S., and M. T. Ramirez (2003): “Institutions, Infrastructure, and Economic Growth,” *Journal of Development Economics*, 70: 443–77.

Estache, A. (2005): “What Do We Know About Sub-Saharan Africa’s Infrastructure and the Impact of Its 1990 Reforms?” Mimeo, World Bank, Washington, DC.

Estache, A., B. Speciale, and D. Veredas (2005): “How Much Does Infrastructure Matter to Growth in Sub-Saharan Africa?” Mimeo, World Bank, Washington, DC.

Fedderke, J. W., P. Perkins, and J. M. Luiz (2005): “Infrastructural Investment in Long-Run Economic Growth: South Africa 1875–2001,” *World Development*.

Fedderke, J. W., and Z. Bogetic (2006): “Infrastructure and Growth in South Africa: Direct and Indirect Productivity Impacts of 19 Infrastructure Measures,” Working Paper 3989, World Bank, Washington, DC.

Francois, J., and M. Manchin (2006): “Institutional Quality, Infrastructure, and the Propensity to Export,” Unpublished manuscript, The World Bank, Washington D.C.

Frutos, R. F., M. Gracia-Diez, and T. Perez-Amaral (1998): “Public Capital Stock and Economic Growth: An Analysis of the Spanish Economy,” *Applied Economics* 30(8): 985–94.

Galiani, S., P. Gertler, and E. Schargrodsky, (2005): “Water for Life: The Impact of Privatization of Water Services on Child Mortality,” *Journal of Political Economy* 113: 83–120.

Guasch, J. L. (2004): “Granting and Renegotiating Infrastructures Concessions, Doing It Right,” World Bank Institute (WBI) Development Studies, Report 28816, World Bank, Washington, DC.

Hall, R. (1990): “Invariance Properties of Solow’s Productivity Residual.” In P. Diamond (ed.), *Growth, Productivity, Unemployment. Essays to Celebrate Bob Solow’s Birthday*. Cambridge, MA: Massachusetts Institute of Technology (MIT) Press.

Hulten, C. R. (1996): “Infrastructure Capital and Economic Growth: How Well You Use It May Be More Important than How Much You Have,” NBER Working Paper 5847, Cambridge, MA.

Kumar, N. (2001): “Infrastructure Availability, Foreign Direct Investment Inflows and Their Export-orientation: A Cross-Country Exploration,” Research and Information System for Developing Countries, New Delhi, India.

Limao, N., and A. Venables (2001): “Infrastructure, Geographical Disadvantage, Transport Costs and Trade.” *World Bank Economic Review* 15: 451–79.

Little, R. J. A., and D. B. Rubin (1987): “Statistical Analysis With Missing Data,” *Wiley Series in Probability and Mathematical Statistics*. John Wiley and Sons (eds).

Loayza, N., P. Fajnzylber, and C. Calderón (2002): “Economic Growth in Latin America and the Caribbean,” Mimeo, World Bank, Washington, DC.

Loree, D. W., and S. E. Guisinger (1995): “Policy and Non-Policy Determinants of U.S. Equity Foreign Direct Investment,” *Journal of International Business Studies* 26(2), Second Quarter 1995.

Magnus, J. R., and A. L. Vasnev (2007): “Local Sensitivity and Diagnostic Tests,” *Econometrics Journal* (10): 166–92.

Mitra, A., A. Varoudakis, and M. Véganzonès-Varoudkis (2002): “Productivity and Technical Efficiency in Indian States’ Manufacturing: The Role of Infrastructure,” *Economic Development and Cultural Change* 2002: 395–426.

Mody, A., and K. Srinivasan (1996): “Japanese and United States Firms as Foreign Investors: Do They March to the Same Tune?” Mimeo, World Bank, Washington, DC.

Munnell, A. H. (1990a): “Why Has Productivity Growth Declined? Productivity and Public Investment,” *New England Economic Review* January/February 1990: 3–22.

——— (1990b): “How Does Public Infrastructure Affect Regional Economic Performance?” *New England Economic Review* September/October 1990: 11–32.

——— (1992): “Infrastructure Investment and Economic Growth.” *Journal of Economic Perspectives* 6(4): 189–98.

Ndulu, B. J., and A. O’Connell (2005): “Sub-Saharan Africa: Growth Econometrics and Country Experiences.” Paper prepared for the African Economic Research Consortium (AERC)/Harvard workshop on, “Explaining African Economic Growth,” Weatherland Center, Harvard University.

Olley, G. S., and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica* 64 (6): 1263–97.

Prud'Homme, R. (2004): "Infrastructure and Development," Paper presented at the Annual Bank Conference on Development Economics, World Bank, Washington DC.

Reinikka, R., and J. Svensson (1999): "How Inadequate Provision of Public Infrastructure and Services Affect Private Investment?" Policy Research Working Paper 2262, World Bank, Washington, DC.

Roller, L. H., and L. Waverman (2001): "Telecommunications Infrastructure and Economic Development: A Simultaneous Approach," *American Economic Review*, American Economic Association, 91(4): 909–23.

Romp, W., and J., de Haan (2005): "Public Capital and Economic Growth: a Critical Survey," *European Investment Bank Papers* 10(1).

Sachs, J., J. McArthur, G. Schmidt-Traub, M. Kruk, C. Bahadur, M. Faye, and G. McCord (2004): "Ending Africa's Poverty Trap," Brookings Papers on Economic Activity, 117–240.

Shioji, E. (2001): "Public Capital and Economic Growth: A Convergence Approach," *Journal of Economic Growth* 6(3): 205–27.

Solow, R. (1957): "Technical Change and the Aggregate Production Function," *Review of Economics and Statistics* 39: 312–20.

Straub, S. (2008): "Infrastructure and Growth in Developing Countries: Recent Advances and Research Challenges," Policy Research Working Paper Series 4460, World Bank, Washington, DC.

Veeramani, C., and G. Goldar (2004): "Investment Climate and Total Factor Productivity in Manufacturing: Analysis of Indian States," Working paper #127, Indian Council for Research on International Economic Relations, New Delhi, India.

Wilson, J. S., C. L. Mann, and T. Otsuki (2004): "Assessing the Potential Benefit of Trade Facilitation: A Global Perspective." Working Paper WPS 3224, World Bank, Washington, DC.

World Bank (1994): "*World Development Report 1994: Infrastructure for Development,*" New York: Oxford University Press.

——— (2004): "*Africa's Competitiveness Report 2004,"* Washington, DC: World Bank.

——— (2006): "*Doing Business in 2007: How to Reform,"* Washington, DC: World Bank.

——— (2007): "*Africa's Competitiveness Report 2007,"* Washington, DC: World Bank.

Appendix: Tables and figures

Table A.1 General information on firms and industries and on production function (PF) variables

General information at firm level	**Industrial classification**	(a) Food and beverages; (b) textiles and apparels; (c) chemicals, rubber, and plastics; (d) paper, printing, and publishing; (e) machinery and equipment/metallic products; (f) wood and furniture; (g) nonmetallic products; and (h) other manufacturing.
	Size classification	Small firms: less than 20 employees; medium firms in between 20 and 99 employees; large firms more than 99 employees.
	Country/Region classification	**1) Algeria**: Region A, Region B, Region C, Region D **2) Benin**: South (coastal), rest of country (rainforest) **3) Botswana**: Francistown, Gaborone **4) Burkina Faso**: Ouagadougou, rest of country **5) Burundi**: Bujumbura **6) Cameroon**: Bafoussam, Douala, Yaounde **7) Cape Verde**: Mindelo, Praia **8) Eritrea**: Eritrea **9) Ethiopia**: Addis Ababa, Awasa, Bahir Dar, Dire Adwa, Mekele, Nazareth, Gondar, Adigrat, Harar, Adwa, rest of country **10) Kenya**: Nairobi, rest of country **11) Madagascar**: Antananarivo, rest of country **12) Malawi**: Blantyre, Lilongwe, rest of country **13) Mali**: Bamako, rest of country **14) Mauritania**: Noauadhibou, Nouakchott **15) Mauritius**: Port Louis, Beau Bassin, Vacoas Phoenix, Curepipe, Quatre Bornes, other **16) Morocco**: Settat, Nador, Casablanca, Rabat, Fes, Tanger **17) Namibia**: Walvis Bay, Windhoek **18) Niger**: Maradi, Niamey **19) Senegal**: Dakar, rest of country **20) South Africa**: Gauteng, Kwazulu, Natal, Western Cape, Eastern Cape **21) Swaziland**: Matsapha, Manzini, Mbabane **22) Tanzania**: Dar es Salaam, Kilimanjaro, Tanga/Arusha, Lake Victoria, South, Zanzibar; **23) Uganda**: Central, North East, South West **24) Lesotho**: Maseru, rest of country **25) Egypt**: Cairo, Suez Channel, Qualyubia, Menoufiya, Alexandria, Nile Delta, Sharkiya, Lower Egypt **26) Zambia**: Lusaka, Ndola, Kitwe, rest of country
PF variables (productivity)	**Sales**	Used as the measure of output for the PF estimation. Sales are defined as total annual sales. The series are deflated by using the consumer price index (CPI), base 2000.
	Employment	Total number of permanent (full-time) and temporal (full-time) workers.
	Total hours worked per year	Total number of employees multiplied by the average hours worked per year.
	Materials	Total costs of intermediate and raw materials used in production (excluding fuel). The series are deflated by using the CPI deflator, base 2000.
	Capital stock	Net book value of machinery and equipment (NBVC); for those countries which the net book value is not available it is replaced by the replacement cost of machinery and equipment. The series are deflated by using the CPI deflator, base 2000.
	User cost of capital	The user cost of capital is defined in terms of the opportunity cost of using capital; it is defined as the 15 percent of the value of the capital stock.
	Labor cost	Total expenditures on personnel, deflated by using the CPI deflator, base 2000.

Source: IC data.

Note: All figures are in U.S. dollars.

Table A.2.1 Definition of investment climate (IC) variables; infrastructure

Name of the variable	Description of the variable
Days to clear customs to import	Average number of days to clear customs when importing.
Longest number of days to clear customs to import	Longest number of days to clear customs when importing.
Days to clear customs to export	Average number of days to clear customs when exporting directly.
Longest number of days to clear customs to export	Longest number of days to clear customs when exporting directly.
Cost to clear customs to export	Total cost to clear customs for a typical consignment as a percentage of the consignment value (including payments to clearing agents, storage fees, container handling fees, and gifts or informal payments to customs officials).
Inspections in customs	Percentage of establishment's exports that were physically inspected during last financial year (LFY).
Shipment losses in customs to export	Percentage of the consignment value of the products shipped to be exported that was lost while in transit because of breakage or spoilage.
Dummy for profit from export facilities	Dummy taking value 1 if the plants enjoy a export facility such as customs duty drawback, duty suspension on imported inputs, profit tax exemption, and so on.
Cost of exports	Percent of the value of export earnings was transport costs.
Dummy for public mechanism to cover risks in exports	Dummy taking value 1 if the firm has a public mechanism to cover risk of nonpayment of exported products.
Dummy for outside clearing agent for imports	Dummy variable taking value 1 if the firms uses an agent to facilitate customs clearance for imports.
Average number of days to clear an outgoing container through port	Average time of clearing an outgoing container through a port clear (including preshipment inspection).
Cost to clear an outgoing container through port	Average cost of clearing an outgoing container through a port clear (including preshipment inspection).
Average number of days to clear an incoming container through port	Average time of clearing an incoming container through a port clear (including preshipment inspection).
Cost to clear an incoming container through port	Average cost of clearing an incoming container through a port clear (including preshipment inspection).
Dummy for own power infrastructure	Dummy taking value 1 if the firm provides its own power infrastructure, excluding generators.
Dummy for own generator	Dummy variable taking value 1 if the firm has its own power generator.
Electricity from own generator	Percentage of the electricity used by the plant provided by a own generator.
Cost of electricity from generator	Estimated annual cost of generator fuel as percentage of annual sales.
Cost of electricity from public grid	Average cost per kilowatt-hour (KwH) when using power from the public grid.
Dummy for equipment damaged by power fluctuations	Dummy taking value 1 if any machine or equipment was damaged by power fluctuations.
Equipment damaged by power fluctuations	Value of the losses of machinery and equipment damaged by power fluctuations as a percentage of the net book value of machinery and equipment (NBVC).
Power outages	Total number of power outages suffered by the plant in LFY.
Average duration of power outages	Average duration of power outages suffered in hours, conditional on the pant reports having power outages.
Power fluctuations	Total number of power fluctuations suffered by the plant in LFY.
Average duration of power fluctuations	Average duration of power fluctuations suffered in hours, conditional on the plant reports having power fluctuations.
Sales lost due to power outages	Losses due to power outages as a percentage of total annual sales, conditional on the plant reports having power outages.
Water outages	Total number of water outages suffered by the plant in LFY.
Average duration of water outages	Average duration of water outages suffered in hours, conditional on the plant reports having water outages.
Sales lost due to water outages	Losses due to water outages as a percentage of total annual sales, conditional on the plant reports having power outages.
Dummy for own well or water infrastructure	Dummy taking value 1 if the plant has its own or shared borehole or well, or builds its own water infrastructure.
Water from own well or water infrastructure	Percentage of firm's water supply from its own or shared well.
Cost of water from own well	Total annual cost of self-provided water as a percentage of total annual sales.
Cost of water from public system	Unit cost of using water from the public water system.
Phone outages	Total number of phone outages suffered by the plant in LFY.
Average duration of phone outages	Average duration of phone outages suffered in hours, conditional on the plant reports having water outages.
Losses due to phone outages	Losses due to phone outages as a percentage of total annual sales, conditional on the plant reports having power outages.
Transport failures	Total number of transport failures suffered by the plant in LFY.
Average duration of transport failures	Average duration of transport failures suffered in hours, conditional on the plant reports having water outages.
Sales lost due to transport failures	Losses due to transport failures as a percentage of total annual sales, conditional on the plant reports having power outages.

Source: IC data.

Table A.2.1 Definition of IC variables; infrastructure (cont.)

Name of the variable	Description of the variable
Average duration of transport	Time in hours that it takes to ship the inputs transported by road from the point of origin to the establishment.
Public postal service interruptions	Total number of public postal service interruptions suffered by the plant in LFY.
Average duration of public postal service interruptions	Average duration of public postal service interruptions suffered in hours, conditional on the plant reports having water outages.
Sales lost due to public postal service interruptions	Losses due to public postal service interruptions as a percentage of total annual sales, conditional on the plant reports having power outages.
Dummy for own roads	Dummy taking value 1 if the firm provides its own roads.
Dummy for own transportation for workers	Dummy taking value 1 if the firm provides its transportation for workers.
Dummy for own waste disposal	Dummy taking value 1 if the firm provides its own waste disposal.
Dummy for contract with transportation company	Dummy taking value 1 if the firm arranges transport services for the delivery of finished products, or raw materials by direct contract with transportation company.
Dummy for own transportation	Dummy taking value 1 if the firm arranges transport services for the delivery of finished products, or raw materials with its own transportation.
Products with own transport	Percentage of products delivered with firm's own transport.
Transport delay, outgoing domestic merchandise	Percentage of times that transport services are late in picking up sales for domestic markets at the plant for delivery.
Transport delay, outgoing export merchandise	Percentage of times that transport services are late in picking up sales for exports at the plant for delivery.
Transport delay, incoming domestic merchandise	Percentage of times that transport services are late in dropping off supplies from domestic sources at the plant for delivery.
Transport delay, incoming export merchandise	Percentage of times that transport services are late in dropping off direct imports at the plant for delivery.
Shipment losses, domestic	Percentage of the consignment value of the products shipped for domestic transportation lost while in transit because of theft, breakage, or spoilage.
Shipment losses, exports	Percentage of the consignment value of the products shipped for international transportation lost while in transit because of theft, breakage, or spoilage.
Dummy for e-mail	Dummy variable taking value 1 if the plant mainly uses e-mail to communicate with clients and suppliers.
Dummy for Web page	Dummy variable taking value 1 if the plant uses its own Web page to communicate with clients and suppliers.
Wait for phone connection	Number of days waiting to obtain a phone connection.
Dummy for gifts to obtain a phone connection	Gifts expected or requested to obtain a phone supply.
Wait for electric supply	Number of days waiting to obtain an electricity supply.
Dummy for gifts to obtain a electric supply	Gifts expected or requested to obtain an electrical connection.
Wait for a water supply	Number of days waiting for a water supply.
Dummy for gifts to obtain a water supply	Gifts expected or requested to obtain a water supply.
Wait for an import license	Number of days waiting for an import license.
Dummy for gifts to obtain an import license	Gifts expected or requested to obtain an import license.
Low quality supplies	Percentage of domestic inputs/supplies that are of lower than agreed-upon quality.
Sales lost due to delivery delays, domestic	Percentage of domestic sales lost due to delivery delays from suppliers in LFY.
Sales lost due to delivery delays, imports	Percentage of exports lost due to delivery delays from suppliers in LFY.
Transport delays in domestic sales	Percentage of domestic sales lost due to delays in transportation services in LFY.
Transport delays in international sales	Percentage of exports lost due to delays in transportation services in LFY.
Illegal payments to obtain public utilities	Amount (as a percentage of total annual sales) spent by a typical establishment in "unofficial payments" for public utilities (that is, power, water and sewage, and telephone).
Days of inventory of main supply	Average number of days (measured in production days) that the main input is available on stock.
Days of inventory of finished goods	Average number of days (measured in production days) that the main output is available on stock.

Source: IC data.

Table A.2.2 Definition of IC variables; red tape, corruption, and crime

Name of the variable	Description of the variable
Manager's time spent in bureaucratic issues	In typical week percentage of manager's time spent dealing with bureaucratic issues.
Payments to deal with bureaucratic issues	Total payments as a percentage of total annual sales to "speed up" bureaucratic issues.
Illegal payments to obtain licenses	Amount (as a percentage of total annual sales) spent by a typical establishment in "unofficial payments" for licenses from government institutions, for example, a city council.
Illegal payments to tax administration	Amount (as a percentage of total annual sales) spent by a typical establishment in "unofficial payments" to tax administration.
Wait for a construction permit	Days waiting to obtain a construction permit.
Dummy for gifts to obtain a construction permit	Gifts expected or requested to obtain a construction permit.
Wait for an operating license	Days waiting to obtain a main operating license.
Gifts to obtain an operating license	Gifts expected or requested to obtain a main operating license.
Sales declared to taxes	Percentage of total annual sales that a typical firm operating in plant's sector reports for tax purposes.
Workforce declared to taxes	Percentage of total workforce that a typical firm operating in plant's sector reports for tax purposes.
Days in inspections	Total number of inspections from regulatory agencies received by the plant in LFY.
Dummy for gifts in inspections	Dummy taking value 1 if any informal gift or payment were requested during inspections from regulatory agencies.
Dummy for lawyer/consultant to help deal with permissions	Dummy taking value 1 if the plant uses/used a lawyer and/or consultant to help obtaining all the permissions and licenses needed to operate/enter the market.
Payments to obtain a contract with the government	Payments to obtain a contract with the government as a percentage of contract value.
Dummy for law-influencing firm	Dummy taking value 1 if the firm seeks to influence local or national laws.
Overdue payments to private customers	Percentage of total sales to private enterprises that involved overdue payments in LFY.
Overdue payments to state-owned enterprises (SOEs)	Percentage of total sales to government agencies or SOEs that involved overdue payments in LFY.
Weeks to resolve a case of overdue payment	Percentage of overdue payments that required the action of a court to be solved.
Overdue payments in courts	Percentage of total sales to private enterprises that involved overdue payments that were resolved in courts in LFY.
Weeks to resolve an overdue payment in courts	Weeks that it takes to resolve a typical case of overdue payment in courts
Security expenses	Security expenses as a percentage of annual total sales.
Dummy for security expenses	Dummy taking value 1 if the plant has security expenses.
Illegal payments in protection	Cost in illegal payments to avoid violence, for example to criminal organizations.
Dummy for payments in protection	Dummy taking value 1 if the plant has cost in illegal payments to avoid violence.
Cost to avoid pilferage from workers	Cost in illegal payments to reduce pilferage by workers.
Dummy for cost to avoid pilferage from workers	Dummy taking value 1 if the plant has costs to reduce pilferage by workers.
Crime losses	Crime losses as a percentage of annual total sales in LFY.
Dummy for crime losses	Dummy taking value 1 if the plant has experienced losses due to criminal attempts in LFY.
Crimes reported to police	Percentage of criminal attempts reported to the police.
Crimes solved by police	Percentage of criminal attempts solved by the police.
Days of production lost due to civil unrest	Total number of production days lost due to civil unrest during LFY.
Days of production lost due to absenteeism	Total number of production days lost due to employees absenteeism during LFY.
Dummy for tax exemption	Dummy variable that takes value 1 if the labor regulation has affected plant's employment decisions.
Dummy for lawsuit in the last 3 years	Dummy taking value 1 if the plant had any lawsuit during the last 3 years
Dummy for "gifts" for credit	Dummy if the firm had to offer a gift or an informal payment to get a credit.
Dummy for interventionist labor regulation	Dummy taking value 1 if plant's decisions on hiring and/or firing workers have been influenced by labor regulations.
Total days spent with licenses	Total number of days that were spent dealing with licenses LFY.
Dummy for accountant to accomplish taxes	Dummy if the firm uses an accountant or consultant to accomplish taxes.
Dummy for gifts to tax inspectors	Dummy if the firm had to offer a gift or an informal payment to tax inspectors.
Gifts to tax inspectors	Amount (as a percentage of total annual sales) paid to tax inspectors in gifts and/or irregular payments.
Dummy for labor conflicts	Dummy taking value 1 if the firm had any conflict with employees during LFY.
Average time to hire a skilled worker	Average days that it takes to hire a skilled production worker.
Dummy for conflicts with suppliers	Dummy taking value 1 if the firm had any conflict with suppliers during LFY.
Dummy for conflicts with clients	Dummy taking value 1 if the firm had any conflict with clients during LFY.
Cost of entry	Cost of entry to the market in terms of licenses and permissions needed.
Dummy for consultant to help deal with permissions	Dummy taking value 1 if the firm uses consultants and/or lawyers to help deal with licenses and permissions.

Source: IC data.

Table A.2.3 Definition of IC variables; finance and corporate governance

Name of the variable	Description of the variable
Dummy for trade chamber	Dummy taking value 1 if the firm belongs to a trade chamber or association.
Dummy for credit line	Dummy that takes value 1 if the firm has access to a credit line or overdraft facility.
Credit unused	Percentage of the overdraft that is not being used currently.
Dummy for loan	Dummy that takes value 1 if the firm has access to a loan line.
Dummy for loan with collateral	Dummy that takes value 1 if the firm has access to a loan line with collateral (conditional on having a loan line).
Value of the collateral	Value of the collateral as a percentage of the loan value (conditional on having a loan with collateral).
Interest rate of the loan	The interest rate applied to the last loan.
Dummy for short-term loan	Duration of the loan in years.
Borrowings in foreign currency	Percentage of firm's borrows denominated in a foreign currency.
Dummy for external auditory	Dummy that takes value 1 if the firm has its annual statements externally audited.
Owner of the lands	Percentage of the lands in which the plant operates owned by the firm.
Owner of the buildings	Percentage of the buildings in which the plant operates owned by the firm.
Dummy for owner of the buildings	Dummy taking value 1 if the almost all the buildings in which the plant operates are owned by the firm.
Dummy for owner of the buildings and lands	Dummy taking value 1 if the almost all the lands in which the plant operates are owned by the firm.
Largest shareholder	Percentage of firm's capital owned by the largest shareholder.
Working capital financed by internal funds	Percentage of firm's working capital financed by internal funds.
Working capital financed by commercial banks	Percentage of firm's working capital financed by funds from private domestic banks.
Working capital financed by foreign commercial banks	Percentage of firm's working capital financed by funds from foreign banks.
Working capital financed by leasing	Percentage of firm's working capital financed by leasing.
Working capital financed by state services	Percentage of firm's working capital financed by funds from state services (for example, Brazilian Development Bank, BNDES; Mexican labor and income generation program, PROGER; and so on).
Working capital financed by supplier or customer credit	Percentage of firm's working capital financed by trade credit (supplier or customer credit).
Working capital financed by credit cards	Percentage of firm's working capital financed by credit card.
Working capital financed by equity	Percentage of firm's working capital financed by equity, sale of stock.
Working capital financed by family/friends	Percentage of firm's working capital financed by funds from family or friends.
Working capital financed by informal sources	Percentage of firm's working capital financed by funds from informal sources (for example, money lender).
Working capital financed by other funds	Percentage of firm's working capital financed by other funds.
New investments financed by internal funds	Percentage of new investments in new lands, buildings, or machinery financed by internal funds.
New investments financed by commercial banks	Percentage of new investments in new lands, buildings, or machinery financed by funds from private domestic banks.
New investments financed by foreign commercial banks	Percentage of new investments in new lands, buildings, or machinery financed by funds from foreign banks.
New investments financed by leasing	Percentage of new investments in new lands, buildings, or machinery financed by leasing.
New investments financed by state services	Percentage of new investments in new lands, buildings, or machinery financed by funds from state services (for example, BNDES, PROGER, and so on).
New investments financed by supplier or customer credit	Percentage of new investments in new lands, buildings, or machinery financed by trade credit (supplier or customer credit).
New investments financed by credit cards	Percentage of new investments in new lands, buildings, or machinery financed by credit card.
New investments financed by equity	Percentage of new investments in new lands, buildings, or machinery financed by equity, sale of stock.
New investments financed by family/friends	Percentage of new investments in new lands, buildings, or machinery financed by funds from family or friends.
New investments financed by informal sources	Percentage of new investments in new lands, buildings, or machinery financed by funds from informal sources (for example money lender).
New investments financed by other funds	Percentage of new investments in new lands, buildings, or machinery financed by other funds.
Share of net profits reinvested	Share of net profits reinvested in the firm in the LFY.
Sales bought on credit	Percentage of establishment's inputs that were purchased on credit in LFY.
Dummy for inputs bought on credit	Days that it takes for the establishment to pay off the supply credit.
Inputs bought on credit	Percentage of establishment's total sales that were bought on credit during LFY.

Source: IC data.

Table A.2.3 Definition of IC variables; finance and corporate governance (cont.)

Name of the variable	Description of the variable
Time to pay off the credit for inputs	Average days that it takes to pay off the credits.
Inputs bought on credit with delayed payment	Share of inputs bought on credit.
Wait to clear a check	Total number of days needed on average to clear a check from the establishment's financial institution.
Charges to clear a check	Average fee charged for a check.
Wait to clear a domestic currency wire	Total number of days needed on average to clear a domestic currency wire from the establishment's financial institution.
Charges to clear a domestic currency wire	Average fee charged for a domestic currency wire.
Wait to clear a foreign currency wire	Total number of days needed on average to clear a foreign currency wire from the establishment's financial institution.
Charges to clear a foreign currency wire	Average fee charged for a foreign currency wire.
Wait to clear a letter of credit	Total number of days needed on average to clear a letter of credit from the establishment's financial institution.
Charge to clear a letter of credit	Average fee charged for a letter of credit.
Delay of payments of domestic clients	Total number of days needed on average to clear a payment from a domestic customer.
Charges to get payments from domestic clients	Average fee charged to clear a payment of a domestic customer.
Delay of payments of foreign clients	Total number of days needed on average to clear a payment from a foreign customer.
Charges to get payments from foreign clients	Average fee charged to clear a payment of a foreign customer.
Dummy for current or saving account	Dummy taking value 1 if the firm has a checking or saving account.
Dummy for foreign current or saving account	Dummy taking value 1 if the firm has a foreign checking or saving account.
Dummy for accountant	Dummy taking value 1 if the firm uses an accountant to finish annual statements.

Source: IC data.

Table A.2.4 Definition of IC variables; quality, innovation, and labor skills

Name of the variable	Description of the variable
Dummy for foreign technology	Dummy taking value 1 if the plant uses technology licensed from a foreign-owned company.
Dummy for International Organization for Standardization (ISO) quality certification	Dummy taking value 1 if the firm has any kind of quality certification.
Sales with warranty	Percentage of sales bought with warranty.
Dummy for new product	Dummy taking value 1 if the firm developed a major new product line during LFY.
Dummy for product improvement	Dummy taking value 1 if the firm improved an existing product line during LFY.
Dummy for discontinued product line	Dummy taking value 1 if the firm discontinued at least one product line during LFY.
Dummy for equipment improvement	Dummy taking value 1 if the firm improved the equipment during LFY.
Dummy for R&D	Dummy taking value 1 if the firm had expenses in R&D during LFY.
R&D expenditures	R&D expenditures as a percentage of annual total sales.
Workers engaged in design/R&D	Percentage of workers in staff engaged in R&D and design tasks.
Dummy for subcontracted R&D	Dummy taking value 1 if the firm subcontracted R&D activities during LFY.
Royalties expenditures	Total expenses in royalties as a percentage of total annual sales.
Dummy for new technology	Dummy taking value 1 if the firm introduced a new technology that substantially changed the way that the main product is produced.
Dummy for joint venture	Dummy taking value 1 if the firm agreed a new joint venture with foreign partner during LFY.
Dummy for new license agreement	Dummy taking value 1 if the firm obtained a new license agreement during LFY.
Dummy for outsourcing	Dummy taking value 1 if the firm outsourced a major production activity that was previously conducted in-house during LFY.
Dummy for in-house production	Dummy taking value 1 if the firm brought in-house a major production activity that was previously outsourced during LFY.
Dummy for new plant	Dummy taking value 1 if the firm opened a new plant during LFY.
Dummy for closed plant	Dummy taking value 1 if the firm closed an existing plant during LFY.
Staff—management	Percentage of management in staff.
Staff—professional workers	Percentage of professional production workers in staff.
Staff—skilled workers	Percentage of skilled production workers in staff.
Staff—unskilled workers	Percentage of unskilled production workers in staff.
Staff—nonproduction workers	Percentage of nonproduction workers in staff.
Staff—foreign nationals	Percentage of foreign national workers in staff.
Average education of staff	Average number of years of education of staff.
Average tenure of staff	Average number of years of experience of staff.
Average age of staff	Average age of staff.
Dummy for training	Dummy taking value 1 if the firm provides formal (either internal or external) training to its employees.
Training to skilled workers	Percentage of skilled workers receiving formal (either internal or external) training.
Training to unskilled workers	Percentage of unskilled workers receiving formal (either internal or external) training.
Training to production workers	Percentage of production workers receiving formal (either internal or external) training.
Training to nonproduction	Percentage of nonproduction workers receiving formal (either internal or external) training.
Weeks of training for skilled workers	Weeks of training received by skilled workers.
Weeks of training for unskilled workers	Weeks of training received by unskilled workers.
Workforce with computer	Percentage of workforce using a computer at job.
University staff	Percentage of staff with at least 1 year of university education.
Dummy for university staff	Percentage of staff that regularly uses computer at job.
Manager education	Dummy taking value 1 if the manager of the establishment has a bachelor degree or higher education level.
Manager's experience	Years of experience of the manager in the same industry before joining the establishment.

Source: IC data.

Table A.2.5 Definition of variables; other control variables

Name of the variable	Description of the variable
Age	Age of the firm.
Dummy for incorporated company	Dummy that takes value 1 if the firm is an incorporated company.
Dummy for limited company	Dummy that takes value 1 if the firm is a limited company.
Dummy for SOE	Dummy variable that takes value 1 if the plant is a SOE.
Dummy for foreign direct investment (FDI)	Dummy that takes value 1 if any part of firm's capital is foreign.
Dummy for holdings	Dummy variable that takes value 1 if the firm has holdings or operations in other countries.
Share of the local market	Percentage of local market that is made up by the sales of the establishment.
Share of the national market	Percentage of national market that is made up by the sales of the establishment.
Dummy for direct exports	Dummy taking value 1 if the firm exports more than 10% of the total annual sales.
Share of exports	Share of exports over total annual sales.
Exporting experience	Number of years of exporting experience.
Dummy for direct imports	Dummy taking value 1 if the firm imports more than 10% of the total purchases of intermediate materials.
Share of imports	Share of imported inputs over total purchases of intermediate materials.
Number of competitors	Total number of competitors in the local market of the establishment's main product line.
Capacity utilization	Percentage of capacity utilized.
Trade union	Percentage of workforce unionized
Dummy for privatized firm	Dummy variable that takes value 1 if the firm was previously state-owned.
Dummy for industrial zone	Dummy variable that takes value 1 if the firm is located in an industrial zone.
Days of production lost due to strikes	Total number of production days lost due to strikes.
Dummy for small firm	Dummy taking value 1 if the firm has less than 20 employees.
Dummy for medium firm	Dummy taking value 1 if the firm has in between 20 and 100 employees.
Dummy for large firm	Dummy taking value 1 if the firm has more than 100 employees.
Workers infected by HIV	Percentage of workers infected by HIV/AIDS.
Dummy for negative impact of HIV	Dummy variable that takes value 1 if the HIV/AIDS epidemic has negatively affected the firm through absenteeism of workers or high staff turnover.
Cost in HIV-prevention programs	Medical expenses for staff (HIV/AIDS related) as percentage of total sales.

Source: IC data.

Table A.3 Summary of the investment climate assessment (ICA) surveys, sorted by geographical area

		Year of survey	Years of production function (PF) variables	Total number of observations[1]	Final number of observations available for regression analysis[2]
North Africa	Algeria	2002	2000–1	952	706
	Egypt	2004	2001–3	2,931	2,629
	Morocco	2003	2000–2	2,550	2,422
Economic Community of West African States (ECOWAS)	Senegal	2003	2000–2	783	535
	Benin	2004	2001–3	591	475
	Mali	2003	2000–2	462	309
	Cape Verde[3]	2006	2005	47	47
	Mauritania*	2006	2005	80	80
	Burkina Faso*	2006	2005	51	51
	Niger*	2005	2004	64	48
	Cameroon*	2006	2005	119	118
Horn of Africa	Ethiopia**	2002	1999–2001	1,281	1,142
	Eritrea**	2002	2000–1	237	179
East African Community (EAC)	Kenya	2003	2000–2	852	577
	Uganda	2003	2001–2	900	635
	Tanzania	2003	2000–2	828	561
	Burundi[3]	2006	2005	102	101
Southern African Development Community (SADC)	Malawi	2005	2004–5	320	288
	Madagascar	2005	2002–4	870	623
	Zambia	2002	1999–2001	564	417
	Lesotho[3]	2003	2000–2	225	79
	Botswana***	2006	2005	114	112
	Namibia***	2006	2005	106	104
	Swaziland***	2006	2005	70	69
Mauritius		2005	2002–4	636	417
South Africa		2003	2001–2	1,737	1,492

Source: Authors´ calculations**;** ICA data**.**

Note:

[1] Total number of observations is equal to the total number of firms surveyed multiplied by the total number of years.

[2] The observations available for regression analysis are the total number of observations minus the observations with any PF variable missing and/ or outlier after the cleaning process.

[3] Countries for which no regression analysis was conducted.

* Countries pooled for regression analysis: Mauritania, Burkina Faso, Niger, and Cameroon.

** Countries pooled for regression analysis: Ethiopia and Eritrea.

*** Countries pooled for regression analysis: Botswana, Namibia, and Swaziland.

Table B.1 Total number of observations available for the PF variables before and after cleaning missing values and outliers

Percentage over total number of observations in parentheses

	Northern Africa			Western Africa—Economic Community of West African States (ECOWAS)								Horn of Africa	
	DZA	**EGY**	**MAR**	**SEN**	**BEN**	**MLI**	**MRT**	**BFA**	**CPV**	**NER**	**CMR**	**ETH**	**ERI**
Total number of observations	952	2,931	2,550	783	591	462	80	51	47	64	119	1281	237
a) Before cleaning													
Missing observations	605 (63.5)	1,543 (52.6)	95 (3.73)	513 (65.5)	199 (33.6)	211 (45.6)	1 (1.25)	1 (1.96)	0	49 (76.5)	2 (1.68)	150 (11.7)	171 (72.1)
of which:													
firms with one PF variable missing	419 (44.0)	1,009 (34.4)	29 (1.14)	189 (24.1)	146 (24.7)	39 (8.44)	0	1 (1.96)	0	11 (17.1)	1 (0.84)	33 (2.58)	88 (37.1)
firms with two PF variables missing	0	34 (1.16)	1 (0.04)	88 (11.2)	18 (3.05)	25 (5.41)	1 (1.25)	0	0	2 (3.13)	0 (0.00)	9 (0.70)	2 (0.84)
firms with three PF variables missing	0	319 (10.8)	2 (0.08)	57 (7.28)	8 (1.35)	18 (3.90)	0	0	0	25 (39.0)	0 (0.00)	7 (0.55)	30 (12.6)
firms with four PF variables missing	186 (19.5)	181 (6.18)	63 (2.47)	179 (22.8)	27 (4.57)	129 (27.9)	0	0	0	11 (17.1)	1 (0.84)	101 (7.88)	51 (21.5)
Outliers	62 (6.51)	131 (4.47)	103 (4.04)	29 (3.70)	42 (7.11)	10 (2.16)	0	0	0	1 (1.56)	0	83 (6.48)	10 (4.22)
of which:													
outliers in materials	24 (2.52)	78 (2.66)	69 (2.71)	23 (2.94)	31 (5.25)	5 (1.08)	0	0	0	0	0	83 (6.48)	4 (1.69)
outliers in labor cost	21 (2.21)	33 (1.13)	18 (0.71)	3 (0.38)	4 (0.68)	3 (0.65)	0	0	0	1 (1.56)	0	0	4 (1.69)
outliers in both materials and labor cost	17 (1.79)	20 (0.68)	16 (0.63)	3 (0.38)	7 (1.18)	2 (0.43)	0	0	0	0	0	0	2 (0.84)
Available observations after replacing (outliers and missing excluded)	316 (33.1)	1,317 (44.9)	2,352 (92.2)	253 (32.3)	364 (61.5)	242 (52.3)	79 (98.7)	50 (98.0)	47 (100.)	14 (21.8)	117 (98.3)	1,048 (81.8)	61 (25.7)
b) After cleaning													
Missing observations	198 (20.8)	225 (7.68)	71 (2.78)	179 (22.8)	42 (7.11)	129 (27.9)	0	0	0	11 (17.1)	1 (0.84)	101 (7.88)	51 (21.5)
of which:													
firms with one PF variable missing	12 (1.26)	9 (0.31)	8 (0.31)	0	9 (1.52)	0	0	0	0	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
firms with two PF variables missing	0	0	0	0	2 (0.34)	0	0	0	0	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
firms with three PF variables missing	0	34 (1.16)	0	0	1 (0.17)	0	0	0	0	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
firms with four PF variables missing	186 (19.5)	182 (6.21)	63 (2.47)	179 (22.8)	30 (5.08)	129 (27.9)	0	0	0	11 (17.1)	1 (0.84)	101 (7.88)	51 (21.5)
Outliers	60 (6.30)	82 (2.80)	65 (2.55)	69 (8.81)	77 (13.0)	24 (5.19)	0	0	0	5 (7.81)	0	38 (2.97)	7 (2.95)
of which:													
outliers in materials	16 (1.68)	46 (1.57)	35 (1.37)	48 (6.13)	58 (9.81)	22 (4.76)	0	0	0	4 (6.25)	0	38 (2.97)	2 (0.84)
outliers in labor cost	18 (1.89)	10 (0.34)	14 (0.55)	12 (1.53)	8 (1.35)	0 (0.00)	0	0	0	1 (1.56)	0	0 (0.00)	3 (1.27)
outliers in both materials and labor cost	26 (2.73)	26 (0.89)	16 (0.63)	9 (1.15)	11 (1.86)	2 (0.43)	0	0	0	0 (0.00)	0	0 (0.00)	2 (0.84)
Available observations after replacing (outliers and missing excluded)	706 (74.1)	2,629 (89.7)	2,422 (94.9)	535 (68.3)	475 (80.3)	309 (66.8)	80 (100)	51 (100.)	47 (100)	48 (75.0)	118 (99.1)	1,142 (89.1)	179 (75.5)

Source: IC data.

Note: The PF variables are: sales, materials, capital stock, and labor cost; the total number of hours worked per year are not included here. For the countries with panel data, the total number of observations is equal to the total number of firms surveyed, multiplied by the total number of years. For the countries with cross-sectional data the total number of observations is equal to the total number of firms surveyed. Outliers are defined as those observations with the ratio of materials to sales and/or labor cost to sales greater than 1. By useful observations we mean those observations available to run regression and to make statistical inference. Missing observations and/or outliers in sales, materials, or labor cost are therefore not initially considered useful available observations.

Table B.1 (cont.) Total number of observations available for the PF variables before and after cleaning missing values and outliers

Percentage over total number of observations in parentheses)

	Eastern Africa—East African Community (EAC excl. Burundi)			Southern Africa—Southern African Development Community (SADC incl. Burundi)								MUS	ZAF
	KEN	UGA	TZA	MWI	MDG	ZMB	BDI	BWA	LSO	NAM	SWZ		
Total number of observations	852	900	828	320	870	564	102	114	225	106	70	636	1737
							a) Before cleaning						
Missing observations	426 (50.0)	652 (72.4)	457 (55.1)	106 (33.1)	456 (52.4)	153 (27.1)	0 (0.00)	4 (3.51)	187 (83.1)	5 (4.72)	3 (4.28)	340 (53.4)	487 (28.0)
of which:													
firms with one PF variable missing	112 (13.1)	288 (32.0)	189 (22.8)	76 (23.7)	184 (21.1)	26 (4.61)	0	3 (2.63)	38 (16.8)	5 (4.72)	2 (2.85)	117 (18.4)	241 (13.8)
firms with two PF variables missing	48 (5.63)	40 (4.44)	75 (9.06)	8 (2.50)	62 (7.13)	0	0	0	7 (3.11)	0	1 (1.42)	37 (5.82)	37 (2.13)
firms with three PF variables missing	62 (7.28)	95 (10.5)	32 (3.86)	0 (0.00)	30 (3.45)	6 (1.06)	0	0	12 (5.33)	0	0	13 (2.04)	11 (0.63)
firms with four PF variables missing	204 (23.9)	229 (25.4)	161 (19.4)	22 (6.88)	180 (20.6)	121 (21.4)	0	1 (0.88)	130 (57.7)	0	0	173 (27.2)	198 (11.4)
Outliers	53 (6.22)	41 (4.56)	55 (6.64)	10 (3.13)	40 (4.60)	20 (3.55)	2 (1.96)	1 (0.88)	6 (2.67)	1 (0.94)	0	28 (4.40)	34 (1.96)
of which:													
outliers in materials	46 (5.40)	19 (2.11)	25 (3.02)	9 (2.81)	20 (2.30)	18 (3.19)	2 (1.96)	1 (0.88)	0 (0.00)	1 (0.94)	0	9 (1.42)	12 (0.69)
outliers in labor cost	4 (0.47)	16 (1.78)	19 (2.29)	1 (0.31)	17 (1.95)	2 (0.35)	0 (0.00)	0 (0.00)	5 (2.22)	0	0	14 (2.20)	14 (0.81)
outliers in both materials and labor cost	3 (0.35)	6 (0.67)	11 (1.33)	0 (0.00)	3 (0.34)	0 (0.00)	0 (0.00)	0 (0.00)	1 (0.44)	0	0	5 (0.79)	8 (0.46)
Available observations after replacing (outliers and missing excluded)	377 (44.2)	232 (25.7)	325 (39.2)	208 (65.0)	383 (44.0)	391 (69.3)	109 (106.)	100 (87.7)	37 (16.4)	100 (94.3)	67 (95.7)	271 (42.6)	1,229 (70.7)
							b) After cleaning						
Missing observations	205 (24.0)	234 (26.0)	164 (19.8)	22 (6.88)	181 (20.8)	122 (21.6)	0 (0.00)	1 (0.88)	131 (58.2)	0 (0.00)	1 (1.42)	174 (27.3)	199 (11.4)
of which:													
firms with one PF variable missing	0	5 (0.56)	3 (0.36)	0	1 (0.11)	0	0	0	1 (0.44)	0	0	1 (0.16)	1 (0.06)
firms with two PF variables missing	0	0	0	0	0	0	0	0	0	0	0	0	0
firms with three PF variables missing	1 (0.12)	0	0	0	0	0	0	0	0	0	0	0	0
firms with four PF variables missing	204 (23.9)	229 (25.4)	161 (19.4)	22 (6.88)	180 (20.6)	122 (21.6)	0 (0.00)	1 (0.88)	130 (57.7)	0 (0.00)	1 (1.42)	173 (27.2)	198 (11.4)
Outliers	70 (8.22)	35 (3.89)	106 (12.8)	10 (3.13)	66 (7.59)	25 (4.43)	1 (0.98)	1 (0.88)	16 (7.11)	2 (1.89)	0	46 (7.23)	47 (2.71)
of which:													
outliers in materials	64 (7.51)	13 (1.44)	74 (8.94)	4 (1.25)	35 (4.02)	25 (4.43)	0 (0.00)	1 (0.88)	2 (0.89)	1 (0.94)	0	28 (4.40)	18 (1.04)
outliers in labor cost	2 (0.23)	14 (1.56)	12 (1.45)	4 (1.25)	22 (2.53)	0	1 (0.98)	0	6 (2.67)	1 (0.94)	0	11 (1.73)	13 (0.75)
outliers in both materials and labor cost	4 (0.47)	8 (0.89)	20 (2.42)	2 (0.63)	9 (1.03)	0	0	0	8 (3.56)	0 (0.00)	0	7 (1.10)	16 (0.92)
Available observations after replacing (outliers and missing excluded)	577 (67.7)	635 (70.5)	561 (67.7)	288 (90.0)	623 (71.6)	417 (73.9)	101 (99.0)	112 (98.2)	79 (35.1)	104 (98.1)	69 (98.5)	417 (65.5)	1,492 (85.9)

Source: IC data.

Note: As for previous table.

Table B.2.1 Representativity of PF variables before and after cleaning missing values and outliers, by country and year

		1999		2000		2001		2002		2003		2004		2005		Total
		#Obs	Perc. available	#Obs	Perc.	#Obs	Perc.	#Obs	Perc.	#Obs	Perc.	#Obs	Perc.	#Obs	Perc.	#Obs
	Original sample			952		952										1,904
Algeria	Without replacing			552	42.0	562	41.0									1,114
	With replacing			700	26.5	712	25.2									1,412
	Original sample					197		197		197						591
Benin	Without replacing					112	43.1	123	37.6	129	34.5					364
	With replacing					143	27.4	164	16.8	168	14.7					475
	Original sample													114		114
Botswana	Without replacing													109	4.4	109
	With replacing													113	0.9	113
	Original sample													51		51
Burkina Faso	Without replacing													50	2.0	50
	With replacing													51	0.0	51
	Original sample													102		102
Burundi	Without replacing													100	2.0	100
	With replacing													101	1.0	101
	Original sample													119		119
Cameroon	Without replacing													117	1.7	117
	With replacing													118	0.8	118
	Original sample													47		47
Cape Verde	Without replacing													47	0.0	47
	With replacing													47	0.0	47
	Original sample					977		977		977						2,931
Egypt	Without replacing					631	35.4	686		0	100					1,317
	With replacing					795	18.6	902		932	4.6					2,629
	Original sample	79		79		79										237
Eritrea	Without replacing	0	100	38	51.9	23	70.9									61
	With replacing	50	36.7	62	21.5	67	15.2									179
	Original sample	427		427		427										1,281
Ethiopia	Without replacing	316	26.0	344	19.4	388	9.1									1,048
	With replacing	351	17.8	377	11.7	414	3.0									1,142
	Original sample			284		284		284								852
Kenya	Without replacing			110	61.3	119	58.1	131	53.9							360
	With replacing			185	34.9	185	34.9	215	24.3							585
	Original sample			75		75		75								225
Lesotho	Without replacing			9	88.0	12	84.0	16	78.7							37
	With replacing			20	73.3	26	65.3	33	56.0							79

Source: Authors' elaboration with IC data.

Note: Original sample includes all establishments surveyed. Without replacing includes establishments without missing values and/or outliers in PF variables. With replacing includes establishments without missing values and/or outliers in the PF variables.

Table B.2.1 (cont.) Representativity of PF variables before and after cleaning missing values and outliers, by country and year

		1999		2000		2001		2002		2003		2004		2005		Total
		#Obs	Perc. available	#Obs	Perc.	#Obs	Perc.	#Obs	Perc.	#Obs	Perc.	#Obs	Perc.	#Obs	Perc.	#Obs
Madagascar	Original sample							290		290		290				870
	Without replacing							113	61.0	134	53.8	136	53.1			383
	With replacing							183	36.9	212	26.9	228	21.4			623
Malawi	Original sample									160		160				320
	Without replacing									93	41.9	115	28.1			208
	With replacing									136	15.0	152	5.0			288
Mali	Original sample			154		154		154								462
	Without replacing			62	59.7	78	49.4	102	33.8							242
	With replacing			74	51.9	93	39.6	142	7.8							309
Mauritania	Original sample													80		80
	Without replacing													79	1.3	79
	With replacing													80	0.0	80
Mauritius	Original sample							212		212		212				636
	Without replacing							77	63.7	97	54.2	97	54.2			271
	With replacing							122	42.5	142	33.0	153	27.8			417
Morocco	Original sample			850		850		850								2,550
	Without replacing			754	11.3	794	6.6	804	5.4							2,352
	With replacing			780	8.2	813	4.4	829	2.5							2,422
Namibia	Original sample													106		106
	Without replacing													100	5.7	100
	With replacing													104	1.9	104
Niger	Original sample													64		64
	Without replacing													14	78.1	14
	With replacing													48	25.0	48
Senegal	Original sample			261		261		261								783
	Without replacing			59	77.4	84	67.8	110	57.9							253
	With replacing			135	48.3	183	29.9	217	16.9							535
South Africa	Original sample			579		579		579								1,737
	Without replacing			373	35.6	406	29.9	450	22.3							1,229
	With replacing			457	21.1	498	14.0	537	7.3							1,492
Swaziland	Original sample													70		70
	Without replacing													67	4.3	67
	With replacing													69	1.4	69
Tanzania	Original sample			276		276		276								828
	Without replacing			113	59.1	124	55.1	88	68.1							325
	With replacing			193	30.1	205	25.7	163	40.9							561
Uganda	Original sample			300		300		300								900
	Without replacing			102	66.0	112	62.7	154	48.7							368
	With replacing			169	43.7	249	17.0	277	7.7							695
Zambia	Original sample	188		188		188		0								564
	Without replacing	114	39.4	127	32.4	150	20.2	0								391
	With replacing	126	33.0	136	27.7	155	17.6	0								417

Source: Author's elaboration with IC data.

Note: As for first part of table.

Table B.2.2 Representativity of PF variables before and after cleaning missing values and outliers, by country and industry

		Food and beverages		Textiles and apparels		Chemicals, rubber, and plastics		Paper, edition, and publishing		Mach and equipment/metallic products		Wood and furniture		Nonmetallic products		Other manufacturing	
Country		**#Obs**	**Perc. available**	**#Obs**	**Perc.**	**#Obs**	**Perc.**	**#Obs**	**Perc.**	**#Obs**	**Perc.**	**#Obs**	**Perc.**	**#Obs**	**Perc.**	**#Obs**	**Perc.**
Algeria	Original sample	204		372		404		308		440				144		32	
	Without replacing	114	44.1	200	46.2	280	30.7	162	47.4	256	41.8			98	31.9	4	87.5
	With replacing	174	14.7	258	30.6	332	17.8	204	33.8	320	27.3			114	20.8	10	68.8
Benin	Original sample	120				36		135		66		189				45	
	Without replacing	71	40.8			29	19.4	75	44.4	37	43.9	125	33.9			27	40.0
	With replacing	98	18.3			34	5.6	110	18.5	48	27.3	147	22.2			39	13.3
Botswana	Original sample	12		27		0										75	
	Without replacing	12	0.0	26	3.7	0										71	5.3
	With replacing	12	0.0	27	0.0	0										74	1.3
Burkina Faso	Original sample	14						12								25	
	Without replacing	13	7.1					12	0.0							25	0.0
	With replacing	14	0.0					12	0.0							25	0.0
Burundi	Original sample	19		24												59	
	Without replacing	18	5.3	24	0.0											58	1.7
	With replacing	19	0.0	24	0.0											58	1.7
Cameroon	Original sample	31				17		19		11		18				23	
	Without replacing	31	0.0			17	0.0	18	5.3	11	0.0	18	0.0			22	4.3
	With replacing	31	0.0			17	0.0	18	5.3	11	0.0	18	0.0			23	0.0
Cape Verde	Original sample	12										16				19	
	Without replacing	12	0.0									16	0.0			19	0.0
	With replacing	12	0.0									16	0.0			19	0.0
Egypt	Original sample	468		915		453				672		174		249			
	Without replacing	225	51.9	393	57.0	219	51.7			303	54.9	67	61.5	110	55.8		
	With replacing	416	11.1	815	10.9	414	8.6			602	10.4	152	12.6	230	7.6		
Eritrea	Original sample	54		51						18						114	
	Without replacing	14	74.1	11	78.4					8	55.6					28	75.4
	With replacing	38	29.6	39	23.5					15	16.7					87	23.7
Ethiopia	Original sample	285		279						618						99	
	Without replacing	233	18.2	207	25.8					531	14.1					77	22.2
	With replacing	258	9.5	240	14.0					557	9.9					87	12.1
Kenya	Original sample	249		141		144				147						171	
	Without replacing	99	60.2	69	51.1	62	56.9			57	61.2					73	57.3
	With replacing	172	30.9	95	32.6	97	32.6			91	38.1					130	24.0
Lesotho	Original sample	54		102												69	
	Without replacing	17	68.5	8	92.2											12	82.6
	With replacing	31	42.6	24	76.5											24	65.2

Source: Author's elaboration with IC data.

Note: As for previous table.

Table B.2.2 Representativity of PF variables before and after cleaning missing values and outliers, by country and industry (cont.)

		Food and beverages		Textiles and apparels		Chemicals, rubber, and plastics		Paper, edition, and publishing		Mach and equipment/metallic products		Wood and furniture		Nonmetallic products		Other manufacturing	
Country		**#Obs**	**Perc. available**	**#Obs**	**Perc.**	**#Obs**	**Perc.**	**#Obs**	**Perc.**	**#Obs**	**Perc.**	**#Obs**	**Perc.**	**#Obs**	**Perc.**	**#Obs**	**Perc.**
Madagascar	Original sample	150		267		108		93		60		192					
	Without replacing	77	48.7	106	60.3	57	47.2	51	45.2	24	60.0	68	64.6				
	With replacing	110	26.7	175	34.5	76	29.6	76	18.3	44	26.7	142	26.0				
Malawi	Original sample	112				70				42		48				48	
	Without replacing	72	35.7			52	25.7			30	28.6	27	43.8			27	43.8
	With replacing	101	9.8			66	5.7			42	0.0	40	16.7			39	18.8
Mali	Original sample	153		30		69		33		66		54		57			
	Without replacing	82	46.4	8	73.3	29	58.0	19	42.4	47	28.8	23	57.4	34	40.4		
	With replacing	97	36.6	14	53.3	47	31.9	27	18.2	56	15.2	30	44.4	38	33.3		
Mauritania	Original sample	27								12		13				28	
	Without replacing	26	3.7							12	0.0	13	0.0			28	0.0
	With replacing	27	0.0							12	0.0	13	0.0			28	0.0
Mauritius	Original sample	117		219		72		54		93		33		18		30	
	Without replacing	53	54.7	97	55.7	32	55.6	32	40.7	29	68.8	20	39.4	8	55.6	0	100.0
	With replacing	86	26.5	139	36.5	47	34.7	50	7.4	63	32.3	23	30.3	9	50.0	0	100.0
Morocco	Original sample	216		1,722		414				147						51	
	Without replacing	196	9.3	1,584	8.0	383	7.5			140	4.8					49	3.9
	With replacing	205	5.1	1,635	5.1	390	5.8			142	3.4					50	2.0
Namibia	Original sample	18		5												83	
	Without replacing	18	0.0	5	0.0											77	7.2
	With replacing	18	0.0	5	0.0											81	2.4
Niger	Original sample	18						14								32	
	Without replacing	6	66.7					0	100.0							8	75.0
	With replacing	12	33.3					12	14.3							24	25.0
Senegal	Original sample	279		69		147		108		75		48		57			
	Without replacing	78	72.0	20	71.0	55	62.6	48	55.6	19	74.7	15	68.8	18	68.4		
	With replacing	186	33.3	46	33.3	106	27.9	73	32.4	49	34.7	29	39.6	45	21.1		
South Africa	Original sample	189		180		285		159		561		147		66		150	
	Without replacing	131	30.7	107	40.6	187	34.4	120	24.5	435	22.5	102	30.6	43	34.8	104	30.7
	With replacing	162	14.3	144	20.0	241	15.4	137	13.8	498	11.2	131	10.9	50	24.2	129	14.0
Swaziland	Original sample	14		20												36	
	Without replacing	12	14.3	19	5.0											36	0.0
	With replacing	13	7.1	20	0.0											36	0.0
Tanzania	Original sample	243		93		102		75		87		195		33			
	Without replacing	108	55.6	29	68.8	42	58.8	33	56.0	26	70.1	68	65.1	19	42.4		
	With replacing	168	30.9	58	37.6	69	32.4	55	26.7	65	25.3	117	40.0	27	18.2		
Uganda	Original sample	366		45		75		69		63		162		120			
	Without replacing	148	59.6	22	51.1	17	77.3	19	72.5	33	47.6	74	54.3	55	54.2		
	With replacing	292	20.2	37	17.8	58	22.7	44	36.2	53	15.9	120	25.9	91	24.2		
Zambia	Original sample	273		69		63				75						84	
	Without replacing	188	31.1	54	21.7	44	30.2			52	30.7					53	36.9
	With replacing	201	26.4	58	15.9	50	20.6			54	28.0					54	35.7

Source: Authors' elaboration with IC data. *Note*: As for first part of table.

Table B.3.1 Response rate of infrastructure IC variables in the final sample

	Northern Africa			Western Africa—ECOWAS								Horn of Africa		Eastern Africa—EAC			Southern Africa—SADC (incl. Burundi)								MUS	ZAF
	DZA	EGY	MAR	SEN	BEN	MLI	MRT	BFA	CPV	NER	CMR	ETH	ERI	KEN	UGA	TZA	MWI	MDG	ZMB	BDI	BWA	LSO	NAM	SWZ		
Days to clear customs to import	53.6	23.7	70.5	58.7	32.1	41.1	47.5	41.2	51.1	60.0	52.9	23.7	73.7	61.1		23.3	66.3		64.5	37.6	74.1	76.0	67.0	48.6	65.2	69.8
Longest number of days to clear customs to import	52.3	23.3	70.4	57.3	32.1	41.7	47.5	41.2	51.1	60.0	52.9	23.4	72.6	35.6		22.4			64.5	37.6	73.2	74.7	66.0	48.6	63.8	69.1
Days to clear customs to export	4.0	18.7	58.4	21.3	17.0	15.9	12.5	23.5	4.3	16.3	26.9	6.3	12.8	50.0		41.9	67.0		31.7	1.0	16.1	42.7	23.6	37.1	55.4	59.1
Longest number of days to clear customs to export	4.0	18.3	58.4	21.3	17.0	15.9	12.5	23.5	4.3	16.3	26.9	6.3	12.3	18.0		40.8	23.3		31.7	1.0	16.1	40.0	23.6	37.1	55.4	58.4
Cost to clear customs to export	99.4						12.5	19.6	2.1	7.5	26.9									1.0	13.4		18.9	37.1		
Inspections in customs			62.7				13.8	25.5	4.3	80.0	39.5									2.0	14.3		20.8	35.7		
Shipment losses in customs to export							13.8	25.5	4.3	16.3	37.8									2.0	17.0		24.5	37.1		
Dummy for profit from export facilities				47.2	21.2		16.3	25.5	4.3	17.5	39.5	5.2	12.8	57.7	18.9	26.1	23.3	28.9							65.9	
Cost of exports												99.5														
Dummy for public mechanism to cover risks in exports				47.0	21.8	20.1										43.6	24.7									
Dummy for outside clearing agent for imports				65.0	34.9	20.7	100.0	39.2	51.1	61.3	53.8			59.1	26.6	98.0		43.7	66.4	2.0	17.0		24.5	37.1	68.6	
Average number of days to clear an outgoing container through port						46.9								28.9	12.3			29.4								
Cost to clear an outgoing container through port														22.7	10.6											
Average number of days to clear an incoming container through port														55.1	22.5			42.5								
Cost to clear an incoming container through port														34.6	20.0											
Dummy for own power infrastructure (excl. generator)				99.2		99.0								98.0	100.0											
Dummy for own generator	98.1	99.9	99.6	99.2	100.0	95.5	100.0	98.0	100.0	80.0	100.0	99.7	100.0	97.4	100.0	96.6	100.0	100.0	100.0	100.0	100.0	92.0	98.1	98.6		
Electricity from own generator	98.9	99.8	99.1	89.7	84.5	97.4	100.0	96.1	100.0	75.0	95.8	12.1	97.2	93.5	99.5	0.0	93.4	99.0	98.8	100.0	98.2	82.7	96.2	97.1		
Cost of electricity from generator				37.3	24.4	39.8						93.5	89.4	60.1	32.9	76.7	18.8		95.0							
Cost of electricity from public grid				73.3		93.2						2.8	22.9	89.7	88.0							58.7				
Dummy for equipment damaged by power fluctuations				97.9	95.6	97.4									95.4	95.3		98.7								
Equipment damaged by power fluctuations				83.2		93.9									91.7	93.7		70.1								
Power outages	99.2	96.6	32.5	80.0	87.0	86.7	100.0	98.0	100.0	75.0	95.8	97.1	96.1	89.3	85.7	76.6	90.6	82.2	100.0	97.0	99.1	74.7	96.2	98.6	98.1	65.1
Average duration of power outages		86.6	32.9	86.5	84.0	85.8	100.0	96.1	95.7	71.3	86.6	92.4	96.1	88.5	81.9	75.5	92.4	98.2	100.0	97.0	100.0	84.0	96.2	97.1	97.1	64.2
Power fluctuations				64.4			88.8	74.5	97.9	62.5	90.8			81.4	84.7		84.7		95.9	93.1	90.2		89.6	88.6	95.4	
Average duration of power fluctuations				68.4										73.3	77.3				100.0							
Sales lost due to power outages	99.2	77.0	33.3		83.4	85.8								94.1	88.8	62.3		91.2	98.8			73.3				51.1
Water outages	99.2	54.6	6.8	86.7	79.6	94.2	68.8	98.0	95.7	70.0	97.5	91.3	96.1	87.2	77.2	33.5	89.2	93.4	100.0	50.5	97.3	70.7	50.0	95.7	95.0	29.9
Average duration of water outages		49.6	7.6	86.5	79.2	93.5	68.8	98.0	91.5	68.8	97.5			81.4	76.1	32.6	90.6	94.9	99.5	50.5	97.3	73.3	50.0	95.7	95.7	29.5

Sales lost due to water outages	99.4	50.7	7.1	86.7	65.1	93.5										22.2	85.8	96.8				65.3			94.2	20.2
Dummy for own well or water infrastructure	94.1			5.1	100.0	100.0						99.7	100.0	94.9	97.8	97.5		100.0	100.0	50.5	97.3		50.0	95.7		
Water from own well or water infrastructure		51.8	100.0	96.6	99.8	99.7	68.8	84.3	95.7	73.8	95.8			90.7	100.0	95.2	71.9	100.0				90.7				
Cost of water from own well				71.2	68.5	76.7								43.3	48.8	52.4		67.9								
Cost of water from public system	68.7													52.0												
Phone outages	98.9		7.4	94.3	81.3	87.7						87.2	96.1	92.1	58.4	17.5		87.2	100.0			60.0			97.4	36.4
Average duration of phone outages			8.1	93.9	77.1	87.7								92.1	57.5	15.7		87.5	100.0			62.7			96.6	36.1
Losses due to phone outages			4.0	94.5		86.1										10.7						40.0			70.3	25.9
Transport failures			7.3	29.7		33.0										13.2	89.9		100.0			62.7				30.1
Average duration of transport failures			7.6	29.5		33.0										11.3	90.6		99.3			62.7				29.7
Sales lost due to transport failures			8.1	29.1		36.2										8.4	83.7					64.0				21.8
Average duration of transport																				100.0	88.4		91.5	98.6		
Public postal service interruptions																						57.3				21.6
Average duration of public postal service interruptions																						50.7				21.2
Sales lost due to public postal service interruptions																						2.7				15.2

Source: Authors' elaboration with IC data.

Table B.3.1 Response rate of infrastructure IC variables in the final sample (cont.)

	Northern Africa			Western Africa—ECOWAS								Horn of Africa		Eastern Africa—EAC			Southern Africa—SADC (incl. Burundi)								MUS	ZAF
	DZA	EGY	MAR	SEN	BEN	MLI	MRT	BFA	CPV	NER	CMR	ETH	ERI	KEN	UGA	TZA	MWI	MDG	ZMB	BDI	BWA	LSO	NAM	SWZ		
Dummy for own roads				99.2	100.0	100.0								96.8	100.0	98.0		100.0							98.8	
Dummy for own transportation for workers				99.2	100.0	100.0								96.8	100.0	98.0		100.0							98.8	
Dummy for own waste disposal				99.2	100.0	99.7								94.3	100.0	98.0		100.0							98.8	
Dummy for contract with transportation company				96.6	96.8	99.0								92.3	81.3	78.2		100.0							98.8	
Dummy for own transportation		99.9	100.0	96.6	96.8	99.0	100.0	98.0	97.9	80.0	100.0			92.3	81.3			100.0		100.0	100.0		98.1	98.6	98.6	
Products with own transport		99.2					100.0	98.0	95.7	77.5	100.0						95.8			100.0	100.0		98.1	98.6	72.9	
Transport delay, outgoing domestic			93.1		53.4									90.3	60.9											
Transport delay, outgoing export			77.9		43.7									88.1	53.1											
Transport delay, incoming domestic			94.4		78.8									89.3	58.6											
Transport delay, incoming international			86.4		44.7									87.0	50.7											
Shipment losses, domestic	99.4	99.0	50.9		83.0		13.8	100.0	97.9	75.0	99.2	98.3	100.0	83.6	21.9	91.1	100.0	98.7	100.0	2.0	17.0	86.7	24.5	37.1	95.9	99.7
Shipment losses, exports		28.8			54.2									67.2	9.9		100.0	65.5							91.8	
Dummy for e-mail	89.3	99.9	72.5	98.9	100.0	98.4	100.0	98.0	100.0	80.0	100.0	99.9	100.0	97.4	100.0	96.2	100.0	100.0	99.3	100.0	100.0	98.7	98.1	98.6	99.5	100.0
Dummy for Web page	85.7	99.8	97.3	99.2	100.0	99.4	100.0	100.0	100.0	80.0	100.0	99.9	100.0	96.6	100.0	94.8	100.0	100.0	99.3	100.0	100.0	98.7	98.1	98.6	96.9	100.0
Wait for phone connection	31.7	13.1	97.5	55.6	53.8	73.5	30.0	21.6	19.1	23.8	16.0	34.3	33.5	61.3	42.8	23.1	48.3	23.8		6.9	24.1	36.0	38.7	20.0	43.9	39.9
Dummy for gifts to obtain a phone connection	0.0	12.6		56.8	56.1	74.1	30.0	23.5	100.0	23.8	16.8			66.0	43.1	30.8	49.3	26.0		13.9	25.9	42.7	38.7	24.3	42.7	39.9
Wait for electric supply	4.2	9.1	94.6	42.7	58.4	69.9	18.8	13.7	8.5	17.5	10.1	27.2	29.6	48.8	52.6	24.0	29.2	14.3		15.8	11.6	25.3	23.6	11.4	27.1	33.7
Dummy for gifts to obtain a electric supply		9.2		44.0	58.4	70.6	18.8	15.7	100.0	16.3	10.1			50.2	54.2	30.2	28.1	14.0		15.8	12.5	37.3	23.6	11.4	27.8	33.7
Wait for a water supply		4.8	94.2	25.5	47.9	65.4	13.8	11.8	6.4	12.5	5.9					17.7	11.5	6.3		0.0	9.8	20.0	14.2	7.1	11.8	30.1
Dummy for gifts to obtain a water supply		4.6		29.1	48.9	65.7	15.0	11.8	100.0	11.3	6.7					25.8	11.8	5.8		0.0	9.8	32.0	14.2	7.1	12.0	30.1
Wait for an import license		9.4		26.9	27.3	49.5	10.0	21.6	19.1	37.5	33.6			46.6	17.8	14.0	22.9	3.2	15.6	19.8	14.3	25.3	21.7	15.7	8.4	26.9
Dummy for gifts to obtain an import license		9.1		32.2	29.4	54.7	11.3	21.6	100.0	38.8	31.1			47.4	18.7	25.2	22.6	5.8		19.8	16.1	33.3	20.8	18.6	8.4	27.1
Low quality supplies	99.4	100.0	98.6	89.7	96.4	97.4						99.1		96.6	94.6	96.8	95.5	99.2	99.5			86.7			97.6	99.7
Sales lost due to delivery delays, domestic	99.4		98.7	75.4	21.8	92.6						99.1		92.9	94.3	85.9	27.8	86.7	99.3			88.0			92.1	99.3
Sales lost due to delivery delays, imports				42.7	93.7	40.1								76.3	30.2		2.1	48.5								
Transport delays in domestic sales			63.6	75.2	86.8	92.9								90.7	94.0	86.0		86.7								
Transport delays in international sales			62.4	43.0	93.7	41.1								75.1	30.2			48.8								
Illegal payments to obtain public utilities				62.7	86.8									62.3	31.7											
Days of inventory of main supply		98.4		89.1		98.4	100.0	94.1	97.9	72.5	98.3	85.3		78.1	91.5	81.9	94.1		99.3	100.0	99.1	86.7	97.2	94.3	95.2	99.0
Days of inventory of finished goods				85.1		98.4						99.6	39.7	96.6	85.7	82.8		90.7							96.4	0.0

Source: Authors' elaboration with IC data.

Table B.3.2 Response rate of red tape, corruption, and crime IC variables in the final sample

	Northern Africa			Western Africa—ECOWAS								Horn of Africa		Eastern Africa—EAC			Southern Africa—SADC (incl. Burundi)								MUS	ZAF
	DZA	EGY	MAR	SEN	BEN	MLI	MRT	BFA	CPV	NER	CMR	ETH	ERI	KEN	UGA	TZA	MWI	MDG	ZMB	BDI	BWA	LSO	NAM	SWZ		
Manager's time spent in bureaucratic issues	99.4		99.6	83.8	92.9	94.2	100.0	90.2	100.0	76.3	99.2	97.8	98.3	93.1	97.6	96.6	97.9	95.8	100.0	100.0	98.2	92.0	98.1	97.1	98.1	99.1
Payments to deal with bureaucratic issues	20.6	97.6		78.3	88.2	93.9	75.0	45.1	89.4	43.8	68.1			78.5	55.3	90.0	53.5	96.6	77.0	92.1	84.8	72.0	74.5	91.4	80.8	93.3
Illegal payments to obtain licenses				63.2										62.1	31.2											
Illegal payments to tax administrators				60.4																						
Days spent with regulation agencies												98.7	23.5													
Cost dealing with regulation agencies												76.2	11.7													
Wait for a construction permit	99.4		89.4	20.4	30.0	54.4	12.5	7.8	19.1	11.3	4.2			24.3	24.1	9.7	21.5			1.0	6.3	18.7	17.0	4.3		16.1
Dummy for gifts to obtain a construction permit				21.1	32.4	57.9	12.5	7.8	100.0	11.3	5.9			27.5	25.2	2.0	22.6			1.0	8.0	32.0	18.9	7.1		16.4
Wait for an operating license		24.0	96.8	20.6	38.7	56.6	5.0	9.8	12.8	25.0	83.2			81.6	98.6	67.3	34.0			9.9	67.0	37.3	36.8	30.0		25.9
Gifts to obtain a operating license		24.3		24.8	40.3	63.4	5.0	9.8	100.0	25.0	78.2			79.4	97.0	100.0	36.5			9.9	68.8	53.3	35.8	31.4		26.1
Sales declared to taxes	33.2	98.4	98.1	78.1	90.1	96.4	95.0	94.1	87.2	76.3	96.6			87.9	69.9	90.9	83.7	98.9	77.5	100.0	98.2	64.0	95.3	98.6	86.3	94.3
Workforce declared to taxes	99.4	98.5			91.6		96.3	94.1	91.5	71.3	95.8						81.3	100.0		100.0	100.0		94.3	94.3	86.1	
Days in inspections		99.9	99.6	91.8	100.0	90.9	87.5	98.0	100.0	73.8	95.8	99.6	98.3	96.0	99.7	95.3	99.3	75.3	100.0	84.2	49.1	82.7	24.5	70.0	95.9	97.8
Dummy for gifts in inspections		100.0		88.2	89.1	66.7	87.5	82.4	74.5	57.5	61.3			60.5	40.6	90.5	92.7	35.2	96.6	84.2	49.1	12.0	27.4	75.7	6.5	79.1
Dummy for lawyer/consultant to help deal with permissions					38.7														62.8							
Payments to obtain a contract with the government	99.4	98.6		47.6	89.9	89.0	73.8	68.6	93.6	40.0	49.6			49.6	53.5	64.8	94.8	98.7		91.1	87.5	41.3	72.6	92.9	88.7	74.1
Dummy for law-influencing firm				99.8	100.0	100.0								96.2	99.8	95.9		0.0	100.0			93.3				99.8
Overdue payments to private customers	99.4	91.9	98.9	91.0	96.6	96.8	100.0	100.0	100.0	80.0	100.0			90.7	98.3	43.8	96.9	96.1	99.3	100.0	100.0	68.0	98.1	98.6	95.4	98.6
Overdue payments to SOEs		99.4	44.8											51.0	26.9	31.8		0.0	43.4	100.0	100.0	50.7	98.1	98.6		69.4
Weeks to resolve a case of overdue payment		49.9	57.6	70.9	77.9	85.4								82.2	49.8	47.8		87.6	79.1			50.7			85.1	93.4
Overdue payments in courts	99.4	97.1	60.6	92.6	64.1	96.8	100.0	100.0	100.0	80.0	100.0			29.8	18.3	10.0		96.1	47.0			41.3			88.0	93.8
Weeks to resolve an overdue payment in courts		11.8	14.4	17.1	19.7	51.8								26.7	9.6	10.4	18.4	4.3	23.3			14.7			14.4	36.4
Security expenses	68.7		97.2	88.2	81.9	96.1	38.8	82.4	87.2	26.3	82.4	97.4	48.6	93.7	96.4	93.0	94.8	97.3	99.8	39.6	30.4	64.0	28.3	42.9	77.0	99.5
Dummy for security expenses	68.1		97.2	88.2	81.9	96.1	100.0	100.0	97.9	80.0	100.0	97.4	48.6	93.7	96.4	93.0	94.8	97.3	99.8	100.0	100.0	64.0	98.1	98.6	77.0	99.5
Illegal payments in protection			85.6	65.9	83.2	93.2								90.1	91.7	86.6	96.2	98.7	100.0						72.2	99.5
Dummy for payments in protection			85.6	65.9	83.2	93.2								90.1	91.7	86.6	94.8	98.7	100.0						70.3	99.5
Cost to avoid pilferage from workers														90.7												
Dummy for cost to avoid pilferage from workers														90.7												
Crime losses	99.4		96.4	89.9	97.7	94.8	98.8	100.0	100.0	76.3	97.5	95.2		92.1	99.7	36.5	87.2	97.0	99.8	99.0	98.2	44.0	97.2	98.6	95.7	98.6
Dummy for crime losses	99.4		96.4	89.9	97.7	94.8	100.0	100.0	100.0	80.0	100.0	95.2		92.1	99.7	36.5	73.6	97.0	99.8	100.0	100.0	58.7	98.1	98.6	79.9	98.6
Crimes reported to police			6.5	44.2	17.6	75.4						11.6		27.5	33.9	64.6	70.8	22.5	82.7			61.3			15.6	81.9
Crimes solved by police			6.0	18.1	13.2	40.5						8.9		27.7	28.0	34.9	66.0	22.0	65.9			38.7			16.1	69.4
Days of production lost due to civil unrest	99.4			96.4	87.4	94.2						98.4				1.3	99.0	99.5	100.0			58.7			95.0	92.4

Days of production lost due to absenteeism	97.6		93.7	88.0	96.1						98.0					97.6	98.1	98.8			68.0			87.3	88.6
Dummy for tax exemption											97.5	97.2					98.1							95.4	
Dummy for lawsuit in the last 3 years				91.8													99.8								
Dummy for "gifts" for credit				96.2																					
Dummy for interventionist labor regulation	100.0			70.8		7.5	96.1	97.9	80.0	96.6	100.0					99.0			0.0	7.1		3.8	12.9		
Total days spent with licenses																		86.8							
Dummy for accountant to accomplish taxes																		100.0							
Dummy for gifts to tax inspectors																		100.0							
Gifts to tax inspectors																		100.0							
Dummy for labor conflicts											92.5														
Average time to hire a skilled worker											91.5														
Dummy for conflicts with suppliers											98.9														
Dummy for conflicts with clients											99.6														
Cost of entry											20.8	10.6													
Dummy for consultant to help deal with permissions											20.9														

Source: Authors' elaboration with IC data.

Table B.3.3 Response rate of finance and corporate governance IC variables in the final sample

	Northern Africa			Western Africa—ECOWAS								Horn of Africa		Eastern Africa—EAC			Southern Africa—SADC (incl. Burundi)								MUS	ZAF
	DZA	EGY	MAR	SEN	BEN	MLI	MRT	BFA	CPV	NER	CMR	ETH	ERI	KEN	UGA	TZA	MWI	MDG	ZMB	BDI	BWA	LSO	NAM	SWZ		
Dummy for trade chamber	99.2	99.9	99.1	99.4	100.0	100.0	100.0	100.0	100.0	80.0	100.0	100.0		82.2	100.0	98.4	99.7	99.5	100.0			93.3			99.5	99.7
Dummy for credit line	97.9	99.6	100.0	99.4	96.8	99.7						98.9	98.9	95.3	100.0	94.3	100.0	100.0	100.0	100.0	100.0	89.3	98.1	98.6	98.1	100.0
Credit unused	97.9	99.9	100.0	91.2	21.2	95.5	100.0	100.0	100.0	80.0	100.0					94.3	64.6	97.6	99.3			82.7			83.0	74.6
Dummy for loan	100.0	100.0	100.0	96.6	96.2	98.7	100.0	100.0	100.0	80.0	100.0	99.5	98.9	91.3	100.0	44.0	100.0	100.0	100.0	100.0	100.0	100.0	98.1	98.6	97.6	100.0
Dummy for loan with collateral	100.0	97.7	53.2	99.8	27.1	100.0	17.5	13.7	31.9	11.3	16.8	94.7	45.3	44.7	21.6	44.0	99.3	100.0	100.0	100.0	100.0	14.7	98.1	98.6	97.4	88.9
Value of the collateral	87.0	99.7	44.8	79.6	20.8	90.0	100.0	98.0	100.0	75.0	98.3	62.2	38.0	34.4	18.6	41.7	89.9	97.8	100.0	12.9	21.4		27.4	14.3	87.3	89.3
Interest rate of the loan	94.7		44.5	93.3	26.3	95.8	100.0	100.0	100.0	80.0	100.0			40.1	20.0	44.0	97.2	97.9	98.8	100.0	97.3	13.3	97.2	98.6	94.7	66.0
Dummy for short-term loan		21.0	52.4	62.9	75.6	67.0									5.2	18.6	97.6	81.5		100.0	100.0	100.0	98.1	98.6	98.3	100.0
Borrows in foreign currency	99.4	99.9		93.9	40.1	93.2	100.0	100.0	100.0	80.0	100.0			77.7	95.4	85.0	96.5	22.0	70.3	100.0	100.0	18.7	98.1	98.6		95.5
Dummy for external auditory			99.1	99.8	97.7	100.0	100.0	98.0	100.0	77.5	95.8	99.4	98.9	98.0	100.0	99.8	100.0	100.0	100.0			88.0			98.8	98.6
Owner of the lands				97.1	90.3	80.6								94.5	96.9	88.2	100.0		100.0	100.0	100.0	92.0	98.1	98.6	91.6	99.8
Owner of the buildings				98.7	100.0	91.3								95.5	99.4	95.7	100.0		100.0			90.7			95.2	99.5
Dummy for owner of the buildings												99.7														
Dummy for owner of the buildings and lands		98.0					100.0	92.2	95.7	80.0	94.1	99.2														
Largest shareholder	99.4	100.0	96.6	93.7	95.0	96.1	100.0	100.0	97.9	80.0	99.2	16.1	16.8	95.7	95.1	95.7	97.2	95.8	99.3	100.0	100.0	98.7	94.3	95.7	96.6	98.6
Working capital financed by internal funds	89.9	100.0	99.8	95.8	96.6	93.9	100.0	100.0	97.9	80.0	99.2	98.5	98.9	93.1	99.8	94.3	98.6	99.5	99.0	100.0	99.1	80.0	98.1	98.6	93.3	99.0
Working capital financed by commercial banks	89.9	100.0	99.8	95.8	96.6	93.9						98.5	98.9	93.1	99.8	94.3	98.6	99.5	99.0	100.0	99.1	80.0	98.1	98.6	93.3	99.0
Working capital fin. by foreign commercial banks		100.0	99.8	95.8	96.6	93.9						98.5		93.1	99.8	94.3	98.6	99.5	99.0			80.0			93.3	99.0
Working capital financed by leasing		100.0	99.8	95.8	96.6	93.9						98.5		93.1	99.8	94.3	98.6	99.5	99.0			80.0			93.3	99.0
Working capital financed by state services		100.0	99.8	95.8	96.6	93.9	100.0	100.0	97.9	80.0	99.2	98.5	98.9	93.1	99.8	94.3	98.6		99.0			80.0				99.0
Working capital fin. by supplier or customer credit	89.9	100.0	99.8	95.8	96.6	93.9						98.5	98.9	93.1	99.8	94.3	98.6	99.5	99.0	100.0	99.1	80.0	98.1	98.6	93.3	99.0
Working capital financed by credit cards		100.0	99.8	95.8	96.6	93.9								93.1	99.8	94.3	98.6	99.5	99.0			80.0			93.3	99.0
Working capital financed by equity		100.0	99.8	95.8	96.6	93.9	100.0	100.0	97.9	80.0	99.2	98.5	98.9	93.1	99.8	94.3	98.6	99.5	99.0			80.0			93.3	99.0
Working capital financed by family/friends	89.9	100.0	99.8	95.8	96.6	93.9	100.0	100.0	97.9	80.0	99.2	98.5	98.9	93.1	99.8	94.3	98.6	99.5	99.0	100.0	99.1	80.0	98.1	98.6	93.3	99.0
Working capital financed by informal sources		100.0	99.8	95.8	96.6	93.9						98.5	98.9	93.1	99.8	94.3	98.6	99.5	99.0	100.0	99.1	80.0	98.1	98.6	93.3	99.0
Working capital financed by other funds		73.3	99.8	95.8	96.6	93.9	35.0	96.1	93.6	77.5	80.7	98.5		93.1	99.8	94.3	98.6	70.9	99.0			80.0			93.3	99.0
New investments financed by internal funds	62.8	73.3	90.5	90.3	85.3	77.7	35.0	96.1	93.6	77.5	80.7	42.6	77.1	70.6	66.6	60.8	68.8	70.9	71.2	51.5	48.2	56.0	56.6	62.9	79.9	89.6
New investments financed bcommercial banks	62.8	73.3	90.5	90.3	85.3	77.7						42.6	77.1	70.6	66.6	60.8	68.8	70.9	71.2	51.5	48.2	56.0	56.6	62.9	79.9	89.6
New investments fin. by foreign commercial banks		73.3	90.5	90.3	85.3	77.7						42.6		70.6	66.6	60.8	68.8	70.9	71.2			56.0			79.9	89.6

New investments financed by leasing		73.3	90.5	90.3	85.3	77.7						42.6		70.6	66.6	60.8	68.8	70.9	71.2			56.0			79.9	89.6
New investments financed by state services		73.3	90.5	90.3	85.3	77.7	35.0	96.1	93.6	77.5	80.7	42.6		70.6	66.6	60.8	68.8		71.2			56.0				89.6
New investments fin. by supplier or customer credit	62.8	73.3	90.5	90.3	85.3	77.7						42.6		70.6	66.6	60.8	68.8	70.9	71.2	51.5	48.2	56.0	56.6	62.9	79.9	89.6
New investments financed by credit cards		73.3	90.5	90.3	85.3	77.7								70.6	66.6	60.8	68.8	70.9	71.2			56.0			79.9	89.6
New investments financed by equity		73.3	90.5	90.3	85.3	77.7	35.0	96.1	93.6	77.5	80.7	42.6	77.1	70.6	66.6	60.8	68.8	70.9	71.2			56.0			79.9	89.6
New investments financed by family/friends	62.8	73.3	90.5	90.3	85.3	77.7	35.0	96.1	93.6	77.5	80.7	42.6	77.1	70.6	66.6	60.8	68.8	70.9	71.2	51.5	47.3	56.0	56.6	62.9	79.9	89.6
New investments financed by informal sources		73.3	90.5	90.3	85.3	77.7						42.6	77.1	70.6	66.6	60.8	68.8	70.9	71.2	51.5	48.2	56.0	57.5	62.9	79.9	89.6
New investments financed by other funds		75.5	90.5	90.3	85.3	77.7						42.6		70.6	66.6	60.8	68.8	70.9	71.2			56.0			79.9	89.6
Share of net profits reinvested		99.8	96.9	83.0	69.1	92.9	100.0	100.0	100.0	80.0	100.0			90.1	79.1	90.5	76.7	78.0	68.1			82.7			95.0	98.3
Sales bought on credit				97.9		95.1						100.0		94.5	99.5	97.3	100.0	98.4	100.0	100.0	100.0	84.0	98.1	98.6		
Dummy for inputs bought on credit		99.4		99.8	99.4	99.7	100.0	100.0	100.0	78.8	99.2			95.3	99.8	98.4		99.5	98.6							
Inputs bought on credit				96.2	57.4	97.7								85.6	99.8	97.9	99.3	99.0		100.0	98.2		98.1	97.1		
Time to pay off the credit for inputs														87.4	62.0	61.7		99.0								
Inputs bought on credit with delayed payment				68.8																						
Wait to clear a check			99.6		88.0							77.2	97.2			81.2		95.3	99.0			61.3			96.9	90.8
Charges to clear a check												52.0	86.0			37.9			74.6			14.7				11.4
Wait to clear a domestic currency wire			99.2		87.2							43.2	21.8			52.1		88.6	76.0			32.0			80.6	86.4
Charges to clear a domestic currency wire												28.1	19.0			36.0			58.0			13.3				9.7
Wait to clear a foreign currency wire			96.1		63.9							11.8	10.1			34.5		61.3	82.3			48.0			83.0	68.9
Charges to clear a foreign currency wire												6.0	10.1			6.3			64.7			6.7				10.9
Wait to clear a letter of credit																			25.7							47.3
Charge to clear a letter of credit																			18.9							8.1
Delay of payments of domestic clients				91.8		96.4								83.6	95.3											
Charges to get payments from domestic clients				70.3		84.8								48.2	71.0											
Delay of payments of foreign clients				51.2		31.7								48.6	22.2											
Charges to get payments from foreign clients				36.6		20.4	100.0	98.0	97.9	80.0	100.0			28.5	16.7											
Dummy for current or saving account		99.6												93.7	99.2	96.1		99.8		100.0	100.0		98.1	98.6		
Dummy for foreign current or saving account															99.2		100.0									
Dummy for accountant				99.8	95.0	99.4												100.0								

Source: Authors' elaboration with IC data.

Table B.3.4 Response rate of quality, innovation, and labor skills IC variables in the final sample

	Northern Africa			Western Africa—ECOWAS								Horn of Africa		Eastern Africa—EAC			Southern Africa—SADC (incl. Burundi)								MUS	ZAF
	DZA	EGY	MAR	SEN	BEN	MLI	MRT	BFA	CPV	NER	CMR	ETH	ERI	KEN	UGA	TZA	MWI	MDG	ZMB	BDI	BWA	LSO	NAM	SWZ		
Dummy for foreign technology		99.0	99.3	97.3	100.0	99.0	100.0	100.0	100.0	53.8	99.2			91.3		98.4	100.0	100.0	98.6	100.0	100.0	94.7	98.1	98.6	99.5	100.0
Dummy for ISO quality certification		100.0	99.1	99.6	99.4	100.0	98.8	98.0	100.0	56.3	100.0	100.0	97.2	98.0		97.9	98.6	100.0	100.0	100.0	100.0	93.3	98.1	98.6	99.5	99.3
Sales with warranty												99.9														
Dummy for new product	99.4	100.0	100.0		100.0		100.0	100.0	100.0	55.0	99.2	98.5		94.7		98.4	100.0	100.0	100.0	100.0	100.0	89.3	97.2	98.6	98.8	99.7
Dummy for product improvement		99.8	100.0		99.4		100.0	100.0	100.0	80.0	99.2			94.7		98.4	100.0	100.0	100.0	100.0	100.0	90.7	97.2	98.6	99.5	99.7
Dummy for discontinued product line																										99.7
Dummy for equipment improvement																	100.0									
Dummy for R&D		99.8		87.0	98.1	89.3	100.0	100.0	100.0	80.0	100.0	98.7	78.8	89.9		99.1	100.0									
R&D expenditures		100.0		69.5	81.9	80.9	100.0	100.0	97.9	80.0	96.6	95.4	40.2	74.3		91.1										
Workers engaged in design/R&D												97.3														
Dummy for subcontracted R&D				97.5								100.0														
Royalties expenditures												98.4														
Dummy for new technology	37.4		100.0	99.8	100.0									96.4	100.0	98.4		100.0	100.0			90.7			99.5	99.7
Dummy for joint venture		99.9	99.8													98.4			100.0			86.7				99.7
Dummy for new license agreement		100.0	99.4													98.4	100.0		100.0			88.0				99.7
Dummy for outsourcing		100.0	99.5													98.4	100.0		100.0	100.0	100.0	77.3	98.1	98.6		99.7
Dummy for in-house production			99.5													97.9	100.0		100.0			76.0				99.5
Dummy for new plant																										99.7
Dummy for closed plant																										99.7
Staff—management			100.0	97.1	98.7	99.4						88.1	100.0	91.3	97.6	93.0	97.6	97.1	100.0			78.7				100.0
Staff—professional workers	68.7	99.5	100.0	95.6	98.9	98.4						88.2	98.3	91.5	97.6	62.3	98.3		99.3			57.3			82.0	100.0
Staff—skilled workers	68.7	99.5	100.0	95.4	98.7	98.7	100.0	100.0	100.0	72.5	97.5	87.9	100.0	91.5	97.6	84.3	97.6	97.1	100.0	100.0	99.1	65.3	97.2	98.6	80.3	100.0
Staff—unskilled workers	68.7	99.5	100.0	95.4	98.7	98.7	100.0	100.0	100.0	72.5	97.5	87.6	98.3	90.9	98.1	66.7	96.9	96.6	100.0	100.0	99.1	64.0	97.2	98.6	77.5	100.0
Staff—nonproduction workers	68.7	99.5	100.0	95.4	98.9	98.4						87.0	100.0	91.5	98.1	70.3	96.9	96.6	100.0			57.3			76.7	100.0
Staff—foreign nationals			96.4	0.0										85.2	98.3		95.8		99.3			78.7				99.3
Average education of staff												95.2														
Average tenure of staff												94.3														
Average age of staff												95.4														
Dummy for training	97.1	99.8	99.5	99.4	92.6	100.0	100.0	100.0	100.0	80.0	100.0	96.8	93.3	93.3	100.0	85.9	99.3	100.0	100.0	100.0	100.0	92.0	98.1	98.6	96.6	100.0
Training to skilled workers	89.1	99.2	98.4	25.3	55.7	23.9	100.0	100.0	100.0	78.8	100.0	98.7	8.9	87.7	91.5	93.0	97.2	98.2	100.0			89.3			81.8	97.6
Training to unskilled workers	89.1	98.9	97.5	23.0	23.1	17.2	100.0	100.0	100.0	78.8	100.0	98.6	14.5	76.9	79.2	89.4	97.9	99.2	99.0			84.0			74.3	95.5
Training to production workers		99.2	96.7																	100.0	100.0		98.1	98.6		
Training to nonproduction		98.7	96.7																	100.0	100.0		96.2	98.6		
Weeks of training for skilled workers				16.0	22.1	19.1	100.0	80.4	70.2	72.5	79.8			75.9	91.0	69.8	88.2	97.8	94.2			86.7			87.5	96.0
Weeks of training for unskilled workers				4.4	8.6	3.2	100.0	80.4	70.2	70.0	79.8			56.7	79.2	55.6	69.1	98.1	78.9			82.7			83.0	91.2
Workforce with computer	99.4		98.7	98.1	100.0	99.4								83.4	99.5	97.5			98.6						84.2	84.8
University staff		97.1	100.0	80.8	90.8	79.3						96.8	91.6	79.1	97.0	89.4		74.3	96.9			77.3				99.8
Dummy for university staff							100.0	96.1	91.5	62.5	99.2									100.0	99.1		96.2	97.1		
Manager's education		99.9	99.2	96.4	32.1	99.4	100.0	100.0	100.0	80.0	100.0	99.1	100.0	98.0	99.1		99.3	26.3	99.3	100.0	100.0	92.0	98.1	98.6	100.0	99.8
Manager's experience		99.8	98.8	83.0	98.1	98.4	100.0	98.0	100.0	76.3	99.2	99.7	100.0	66.4	79.7	80.1	96.5		98.8	100.0	100.0	77.3	98.1	98.6	74.6	99.8

Source: Authors' elaboration with IC data.

Table B.3.5 Response rate of other control C variables in the final sample

	Northern Africa			Western Africa—ECOWAS							Horn of Africa		Eastern Africa—EAC			Southern Africa—SADC (incl. Burundi)								MUS	ZAF
	DZA	EGY	MAR	SEN	BEN	MLI	MRT	BFA	CPV	NER	CMR	ETH	ERI	KEN	UGA	TZA	MWI	MDG	ZMB	BDI	BWA	LSO	NAM	SWZ		
Age	99.4	99.7	100.0	99.2	99.6	99.7	100.0	100.0	100.0	33.8	100.0	100.0	100.0	97.4	99.4	98.9	97.9	99.7	99.3	100.0	100.0	100.0	98.1	95.7	98.8	99.8
Dummy for incorporated company	98.5	99.4	99.9	99.8	100.0	99.7	100.0	100.0	100.0	80.0	100.0	100.0	100.0	98.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.1	98.6	96.4	100.0
Dummy for limited company	99.2	99.4	99.9	99.8	100.0	99.7	100.0	100.0	100.0	80.0	100.0	100.0		98.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.1	98.6	100.0	100.0
Dummy for SOE	98.7	100.0	100.0	99.2	100.0	99.7	100.0	100.0	100.0	80.0	100.0	100.0	100.0	98.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	94.7	98.1	98.6	100.0	100.0
Dummy for FDI	98.7	100.0	100.0	99.2	100.0	99.7	100.0	100.0	100.0	80.0	100.0	98.6	100.0	98.0	100.0	97.9	100.0	100.0	100.0	100.0	100.0	94.7	98.1	98.6	100.0	100.0
Dummy for holdings				99.8	100.0	99.7						100.0		94.1	100.0	20.0	100.0	100.0	100.0			98.7			98.8	100.0
Share of the local market	41.0		64.1	53.5	82.4	83.8	93.8	62.7	83.0	61.3	92.4			78.1		96.6				96.0	98.2	57.3	90.6	81.4		92.9
Share of the national market	36.6		63.2	52.4	83.0	79.3	90.0	66.7	72.3	60.0	84.0			94.5	62.2	86.9	82.3	49.4	75.8	98.0	96.4	40.0	88.7	81.4	57.3	93.3
Dummy for direct exports	97.5	99.7	99.9	96.4	95.8	87.4	100.0	100.0	100.0	80.0	100.0	98.9	100.0	65.0	99.5	98.9	99.3	99.7	100.0	100.0	100.0	86.7	98.1	98.6	97.6	99.5
Share of exports	97.5	99.7	99.9	96.4	95.8	87.4	100.0	100.0	100.0	80.0	100.0	98.9	100.0	94.5	99.5	98.9	99.3	99.7	100.0	100.0	100.0	86.7	98.1	98.6	97.6	
Exporting experience	97.7	97.4	99.9	94.9	92.4	91.9	100.0	100.0	100.0	72.5	100.0	98.0		92.1	17.8	27.2	99.3	31.9	98.8	100.0	99.1	89.3	97.2	94.3	97.6	
Dummy for direct imports		97.9	100.0	90.1	97.9	92.9	68.8	100.0	100.0	80.0	100.0	98.9	100.0	92.9	92.6	93.6	99.0	97.6	92.6	78.2	100.0	100.0	98.1	98.6	93.8	97.4
Share of imports		99.9	100.0	90.1	97.9	92.9	100.0	100.0	100.0	80.0	100.0	97.9	100.0	94.1	93.7	94.8	99.3	99.0	100.0	100.0	100.0	93.3	98.1	98.6	93.8	
Number of competitors	94.3	83.2			95.0							65.4		94.5	98.6	96.1	99.3	72.4	63.5			86.7			72.7	99.5
Capacity utilization	95.2	99.7	99.2	82.7	97.7	98.4	100.0	96.1	91.5	53.8	99.2	87.0	100.0	90.1	94.8	99.5	97.6	95.2	99.8	100.0	98.2	89.3	95.3	94.3	97.1	98.1
Trade union	99.4	95.9	97.3	95.6	97.5	98.7	98.8	100.0	100.0	76.3	100.0	98.6	93.9	91.3	99.4	91.6	100.0	94.7	86.3	100.0	100.0	64.0	97.2	95.7	97.1	99.1
Dummy for privatized firm	97.7	98.1		97.3	98.5	98.7						95.6	100.0	89.5		93.7	97.6	96.8	100.0			98.7			96.9	100.0
Dummy for industrial zone		99.9	100.0	99.2	99.2	99.7	100.0	96.1	100.0	80.0	98.3	97.6	100.0	98.0	100.0	98.6	99.3			100.0	100.0		98.1	98.6	99.5	0.0
Days of production lost due to strikes	99.4	97.4		96.4	87.4	95.5						98.8		89.7	98.9	1.6	65.3	99.5	100.0			65.3			95.0	92.6
Workers infected by HIV				68.2	65.1	57.0								56.3			98.3	70.0							80.8	35.4
Dummy for negative impact of HIV				64.4	76.7	90.6	100.0	98.0	97.9	80.0	99.2	99.9		84.0	94.3	94.8	91.0	64.8		98.0	100.0	62.7	93.4	98.6	88.2	100.0
Cost in HIV-prevention programs				67.0	75.8	75.1						99.9		19.0	66.3	48.7	87.5	78.5							90.9	44.7

Source: Authors' elaboration with IC data.

Table B.4 Classification of the main infrastructure variables (INFs)

	Name of the variable	Description of the variable
Customs clearance	Days to clear customs to import	Average number of days to clear customs when importing (logs)
	Days to clear customs to export	Average number of days to clear customs when exporting directly (logs)
	Wait for an import license	Number of days waiting for an import license (logs)
Energy/ Electricity	Dummy for own power infrastructure	Dummy taking value 1 if the firm provides its own power infrastructure, excluding generators
	Dummy for own generator	Dummy variable taking value 1 if the firm has its own power generator
	Electricity from own generator	Percentage of the electricity used by the plant provided by the own generator
	Cost of electricity from generator	Estimated annual cost of generator fuel as percentage of annual sales
	Cost of electricity from public grid	Average cost per kilowatt-hour (Kw/H) when using power from the public grid (logs)
	Dummy for equipment damaged by power fluctuations / Equipment damaged by power fluctuations	Dummy taking value 1 if any machine or equipment was damaged by power fluctuations / Value of the losses of machinery and equipment damaged by power fluctuations as a percentage of the net book value of machinery and equipment (NBVC)
	Power outages / Average duration of power outages / Sales lost due to same	Total number of (logs) / Average duration of (logs) / Percentage of sales loss due to power outages suffered by the plant in the last fiscal year (LFY) (conditional on the plant reports having power outages)
	Power fluctuations / Average duration of power fluctuations	Total number of (logs) / Average duration of (logs) power fluctuations suffered in hours (conditional on the plant reports having power fluctuations)
	Wait for electric supply	Number of days waiting to obtain an electricity supply (logs)
Water	Water outages / Average duration of water outages /Losses due to same	Total number of (logs) / Average duration of (logs) / Percentage of sales lost due to water outages suffered by the plant in LFY (conditional on the plant reports having water outages)
	Dummy for own well or water infrastructure	Dummy taking value 1 if the plant has its own or shared borehole or well or builds its own water infrastructure
	Water from own well or water infrastructure	Percentage of firm's water supply from its own or shared well
	Cost of water from own well	Total annual cost of self-provided water as a percentage of total annual sales
	Cost of water from public system	Unit cost of using water from the public water system (logs)
	Wait for a water supply	Number of days waiting for a water supply (logs)
Telecom. and ICT	Phone outages / Average duration of phone outages / Losses due to same	Total number of (logs) / Average duration of (logs) / Percentage of sales lost due to phone outages suffered by the plant in LFY (conditional on the plant reports having phone outages)
	Wait for phone connection	Number of days waiting to obtain a phone connection (logs)
	Dummy for e-mail	Dummy variable taking value 1 if the plant mainly uses e-mail to communicate with clients and suppliers
	Dummy for web page	Dummy variable taking value 1 if the plant uses its own Web page to communicate with clients and suppliers
Transport	Transport failures / Average duration of transport failures / Sales lost due to same	Total number (logs) of / Average duration of (logs)/ Percentage of sales lost due to transport failures suffered by the plant in LFY (conditional on the plant reporting on transport failures)
	Dummy for own roads	Dummy taking value 1 if the firm provides its own roads
	Dummy for own transportation for workers	Dummy taking value 1 if the firm provides its own transportation for workers
	Dummy for contract with transportation company	Dummy taking value 1 if the firm arranges transport services for the delivery of finished products or raw materials by directly contracting with the transportation company
	Dummy for own transportation	Dummy taking value 1 if the firm arranges transport services for the delivery of finished products or raw materials with its own transportation
	Products with own transport	Percentage of products delivered with firm's own transport
	Transport delay	Percentage of times that transport services are late in picking up sales for domestic (or international) markets at the plant for delivery
	Shipment losses	Percentage of the consignment value of the products shipped for domestic (or international) transportation lost while in transit because of theft, breakage, or spoilage
	Sales lost due to delivery delays	Percentage of domestic (or international) sales lost due to delivery delays from suppliers in LFY
	Low quality supplies	Percentage of domestic inputs/supplies that are of lower than agreed-upon quality

Source: ICS data.

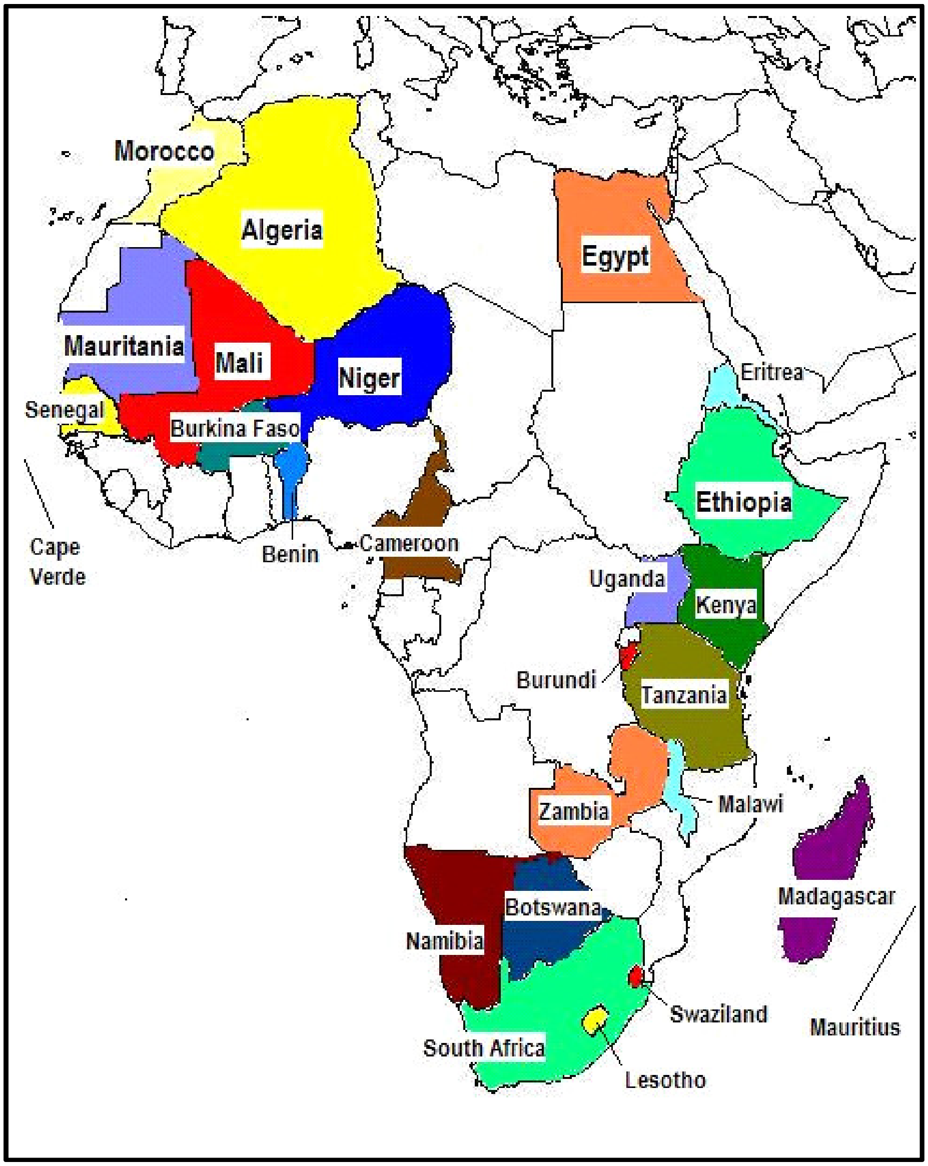
Table C.1 Summary of cross-country comparisons based on alternative rankings of economic performance

Ranking based on per capita GDP	DBR 2007 (rank)	Ranking ACR'07 (rank within sample)	*Demeaned* Aggregate Productivity (rank)	Firms' perceptions: infras. As an obstacle	Perc. Contributions of infrastructure to log-productivity		Perc. contributions of infrastructure to productivity via simulations	
				% abs. contribution (rank)	Average log-productivity	Allocative efficiency in logs	Average productivity	Allocative efficiency
					% abs. contribution (rank)	% abs. contribution (rank)	% abs. contribution (rank)	% abs. contribution (rank)
MUS (1)	32 (2)	2.0 (2)	4.2 (1)	13.9 (2)	26.6 (19)	17.1 (18)	21.8 (6)	12.4 (4)
SWZ (2)	76 (5)	1.4 (7)	n.a	22.4 (10)	25.6 (21)	14.3 (20)	27.4 (10)	17.6 (10)
ZAF (3)	29 (1)	2.3 (1)	4 (3)	16.2 (5)	28.6 (18)	19.7 (17)	17.4 (4)	11.0 (2)
BWA (4)	48 (4)	1.7 (3)	3.4 (6)	15.6 (4)	17.5 (22)	7.41 (23)	23.2 (8)	8.8 (1)
DZA (5)	116 (12)	1.5 (4)	2.9 (7)	18.3 (7)	48.6 (7)	31.1 (4)	34.9 (18)	26.4 (17)
NAM (6)	42 (3)	1.5 (6)	4.2 (2)	18.3 (6)	16.5 (23)	32.9 (3)	22.7 (7)	36.7 (20)
EGY (7)	165 (22)	1.5 (5)	3.7 (4)	14.0 (3)	26.0 (20)	23.8 (12)	19.9 (5)	16.1 (8)
MAR (8)	115 (11)	1.1 (9)	3.6 (5)	9.9 (1)	31.3 (15)	16.6 (19)	16.2 (3)	14.8 (6)
CMR (9)	152 (18)	0.8 (16)	1.9 (18)	27.5 (23)	41.6 (10)	25.4 (11)	31.2 (13)	23.2 (13)
MRT (10)	148 (16)	0.6 (19)	2.1 (15)	25.3 (17)	35.4 (11)	21.1 (15)	28.3 (12)	16.2 (9)
SEN (11)	146 (15)	0.9 (12)	n.a	22.7 (11)	58.5 (3)	40.9 (2)	52.1 (21)	42.2 (22)
BEN (12)	137 (13)	0.6 (20)	2.1 (11)	25.6 (18)	59.9 (2)	12.4 (21)	33.3 (17)	23.3 (14)
KEN (13)	83 (6)	1.0 (11)	2.8 (8)	25.6 (19)	30.3 (17)	19.9 (16)	26.1 (9)	23.2 (12)
MLI (14)	155 (19)	0.9 (14)	2.1 (14)	21.6 (9)	42.7 (9)	26.8 (9)	42.5 (19)	33.5 (19)
UGA (15)	107 (9)	0.6 (21)	2 (17)	23.3 (12)	58.4 (4)	29.8 (5)	45.4 (20)	42.0 (21)
BFA (16)	163 (21)	0.8 (15)	2.1 (12)	26.9 (22)	35.3 (12)	27.0 (8)	27.6 (11)	12.0 (3)
ZMB (17)	102 (8)	0.7 (18)	n.a	24.0 (14)	50.6 (6)	26.8 (10)	15.4 (2)	15.1 (7)
TZA (18)	142 (14)	0.2 (23)	2.7 (9)	24.3 (15)	34.1 (14)	28.3 (6)	32.3 (15)	29.1 (18)
NER (19)	160 (20)	0.8 (17)	n.a	26.2 (20)	34.7 (13)	11.1 (22)	31.6 (14)	22.1 (11)
MWI (20)	110 (10)	0.4 (22)	2.1 (13)	24.5 (16)	65.9 (1)	45.8 (1)	53.7 (22)	55.2 (23)
MDG (21)	149 (17)	1.4 (8)	2 (16)	23.5 (13)	30.6 (16)	27.9 (7)	11.1 (1)	14.3 (5)
ETH (22)	97 (7)	1.0 (10)	2.3 (10)	26.7 (21)	52.6 (5)	21.9 (14)	33.2 (16)	25.0 (15)
ERI (23)	170 (23)	0.9 (13)	n.a	20.7 (8)	46.1 (8)	22.5 (13)	54.7 (23)	25.3 (16)

Source: Authors' calculations using ICA data, DBR (2007), ACR (2007), and Penn World Table.
Note: n.a = not available.

Figures from section 1

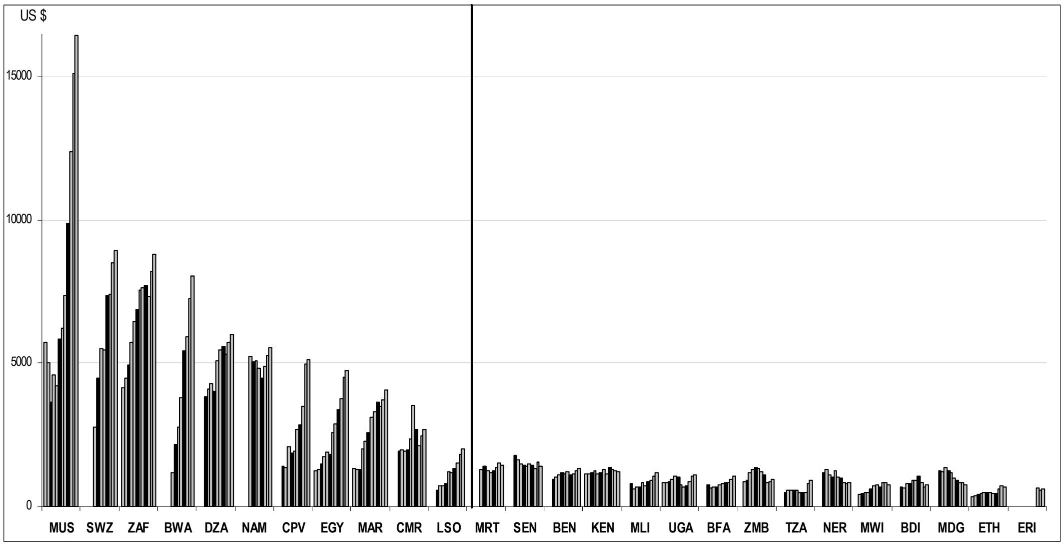
Figure 1.1 Geographical locations of the 26 countries considered in the investment climate assessment (ICA)



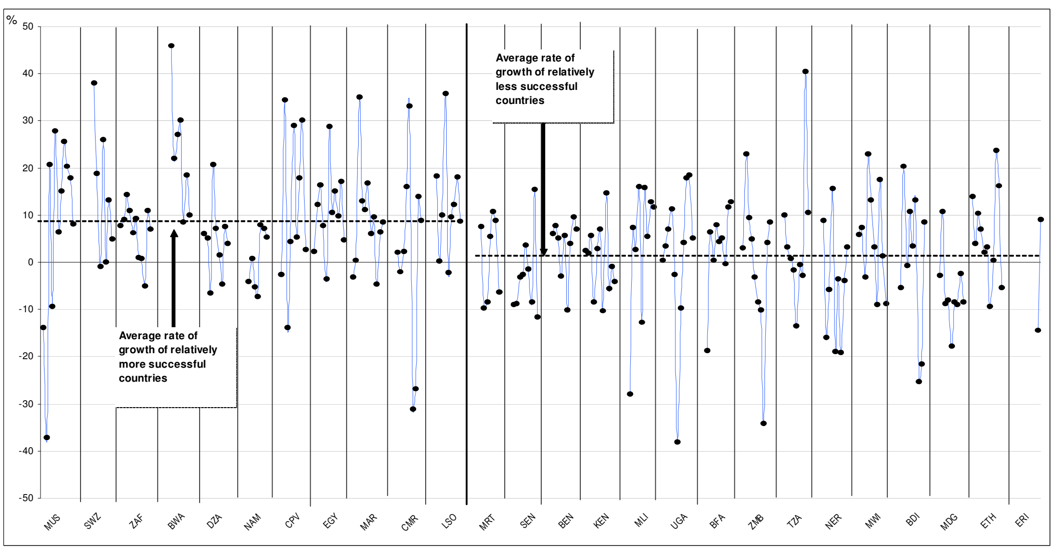
Source: Authors' elaboration.

Figure 1.2 The evolution of gross domestic product (GDP) per capita and ranking based on the ease of doing business in African countries

A. Country by country evolution of GDP per capita, 1950 - 2003



B. Country by country five-year rate of growth of GDP per capita, 1950 - 2003

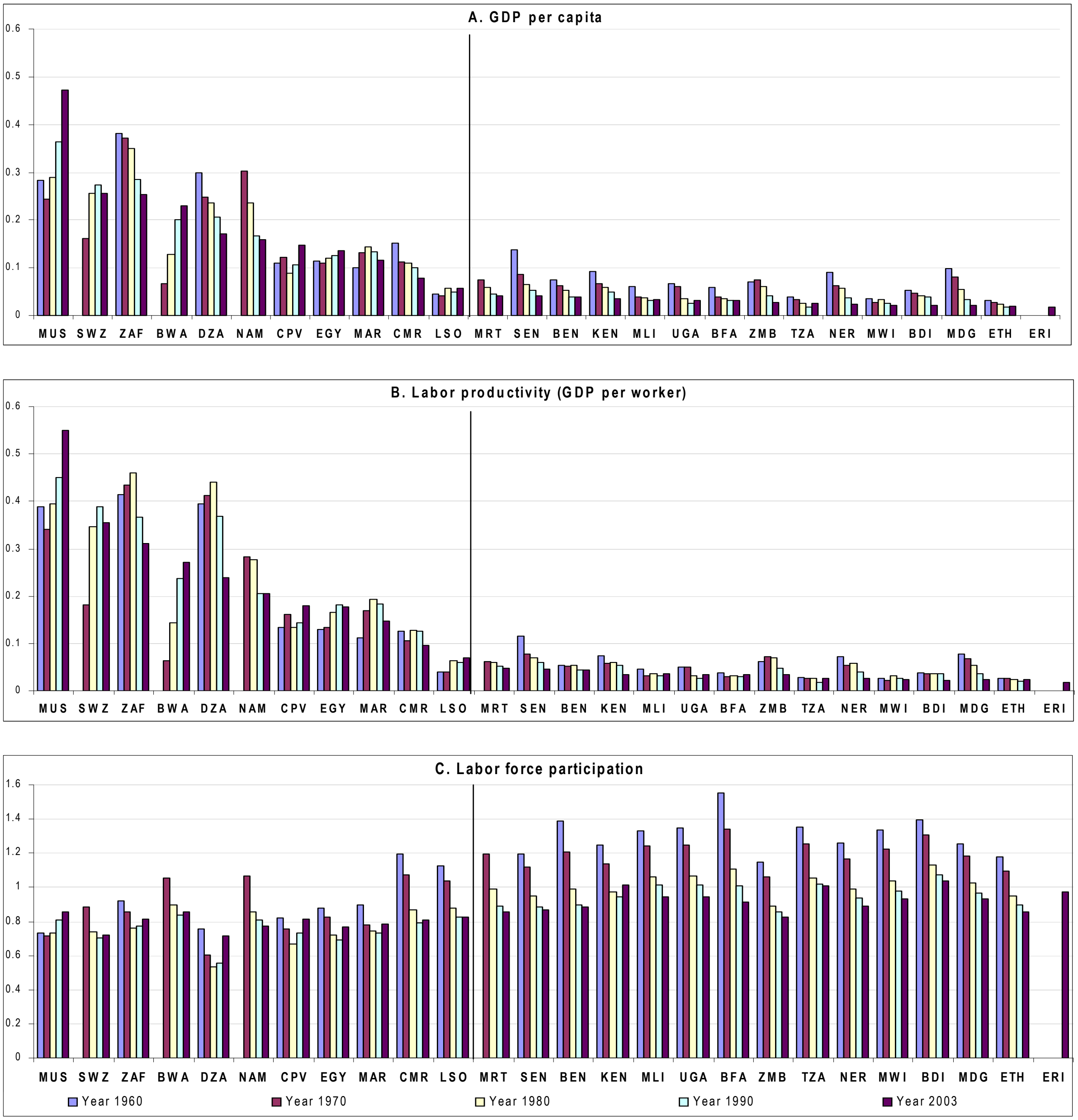


C. Ranking of countries on the ease of doing business, World Bank's Doing Business Report

	MUS	SWZ	ZAF	BWA	DZA	NAM	CPV	EGY	MAR	CMR	LSO	MRT	SEN	BEN	KEN	MLI	UGA	BFA	ZMB	TZA	NER	MWI	BDI	MDG	ETH	ERI
2007	32	76	29	48	116	42	125	165	115	152	114	148	146	137	83	155	107	163	102	142	160	110	166	149	97	170
2006	32	67	28	44	123	39	125	165	117	147	116	146	152	139	80	166	103	171	92	150	170	106	160	148	96	168

Source: Source: Penn World Table, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, 2006 and World Bank's Doing Business Report 2007.

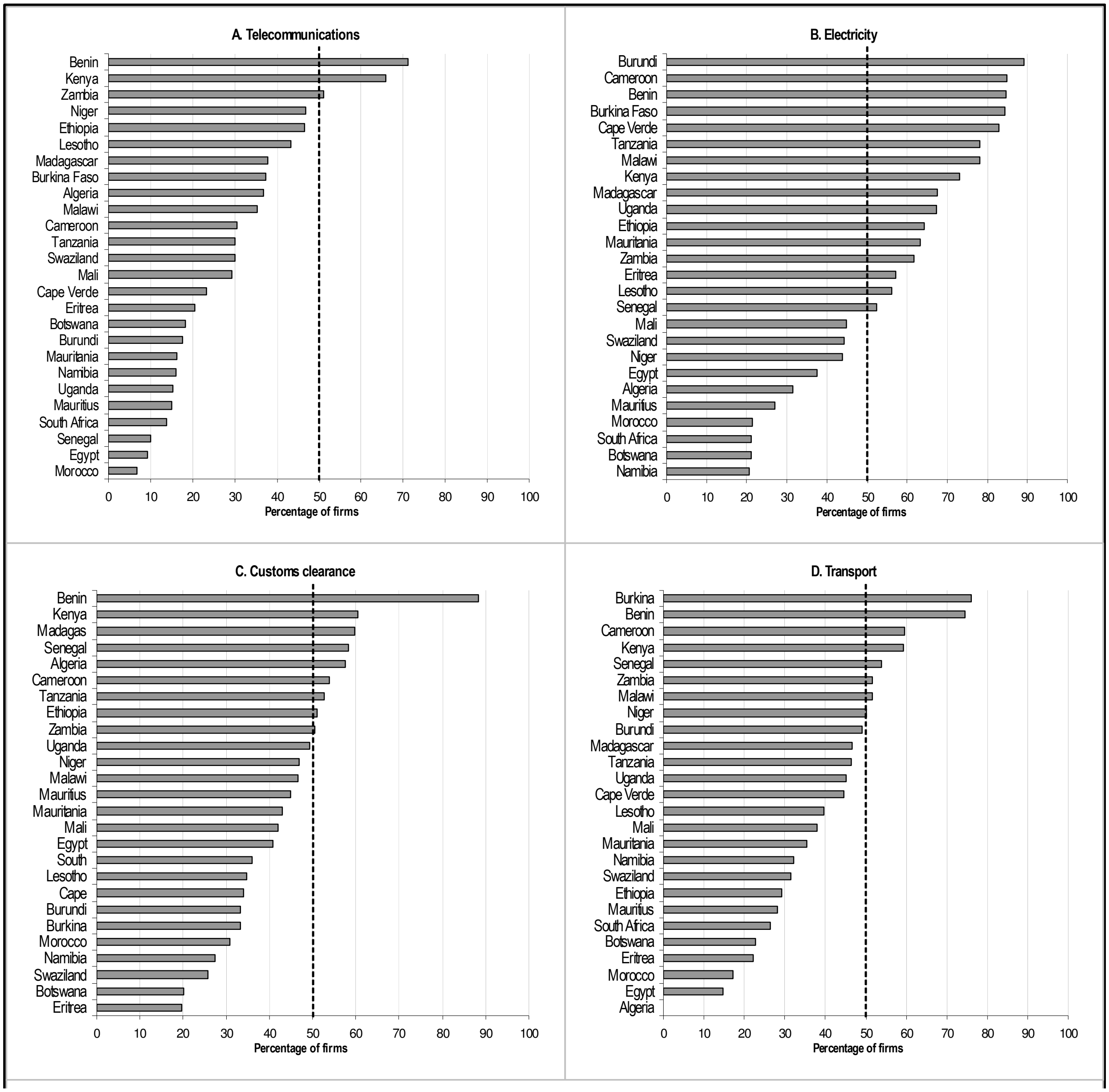
Figure 1.3 Evolution of per capita income in Africa relative to the United States, 1960–2003



Per capita income of country J (Y^J/P^J) is decomposed into the product of labor productivity (Y^J/L^J) and the labor participation rate (L^J/P^J) by: $(Y^J/P^J) = (Y^J/L^J)*(L^J/P^J)$. Therefore, per capita income relative to the United States becomes: $[(Y^J/P^J)/(Y^{US}/P^{US})] = [(Y^J/L^J)/(Y^{US}/L^{US})]*[(L^J/P^J)/(L^{US}/P^{US})]$.

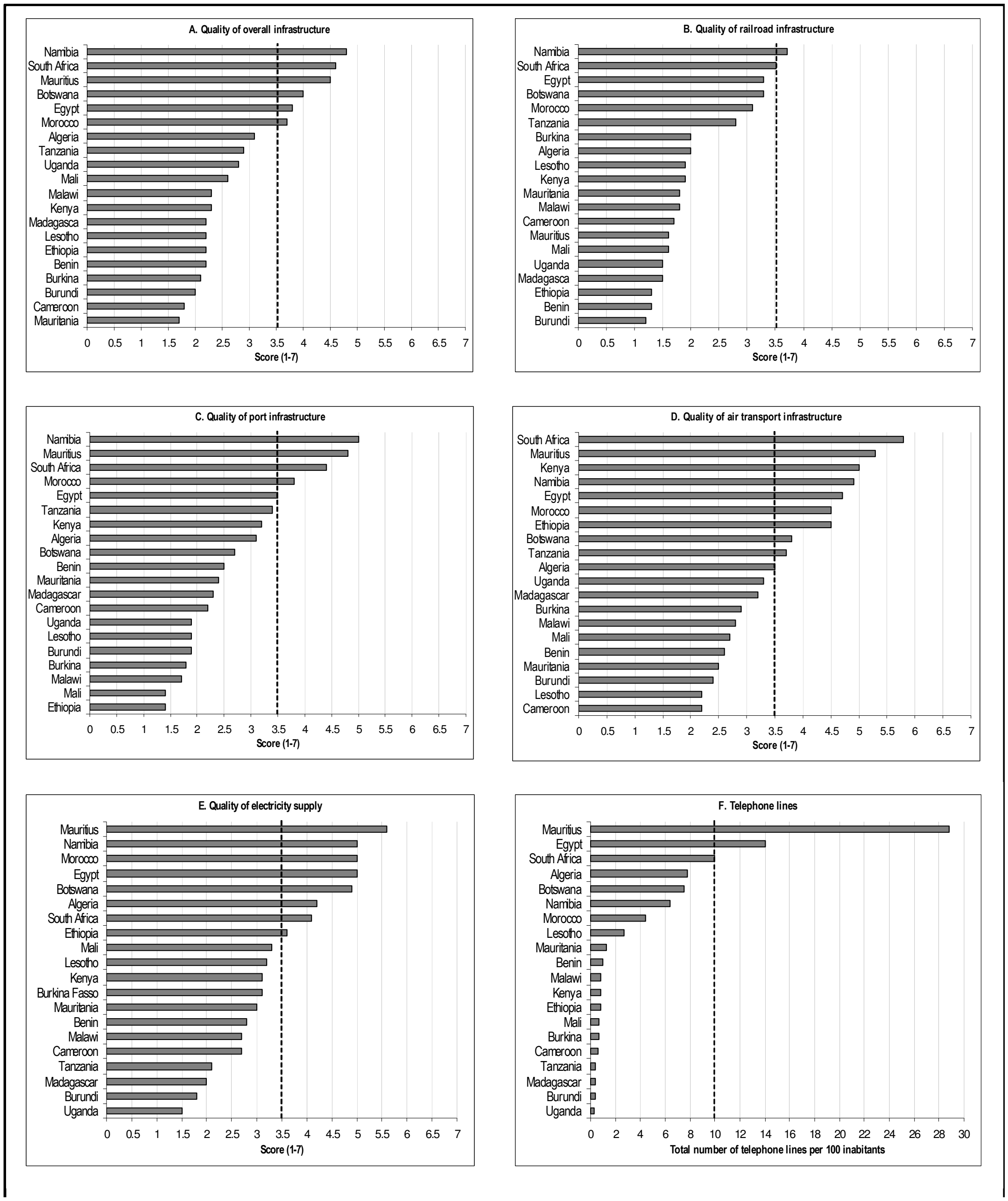
Source: Penn World Table, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, 2006.

Figure 1.4 Percentage of firms that consider telecommunications, electricity, customs, and transport as severe or very severe constraints on economic performance (by country)



Source: Authors' calculations based on IC data.
Note: No data are available for perceptions of transport in Algeria.

Figure 1.5 The state of infrastructure in Africa, at first glance

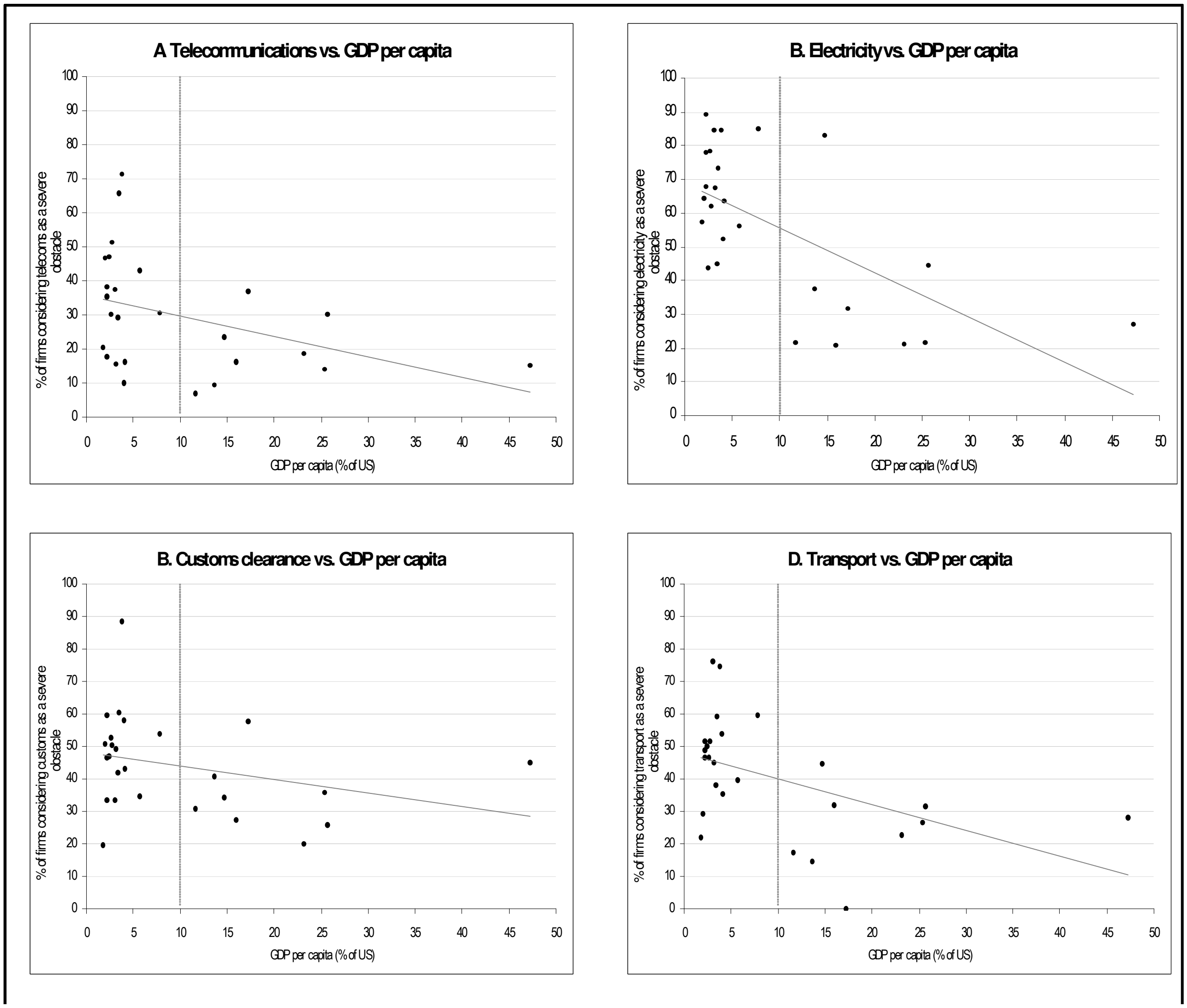


Source: Africa Competitiveness Report (2007), World Bank, Washington, DC.

Note: No data are available for Cape Verde, Eritrea, Niger, Senegal, Swaziland, or Zambia.

Figure 1.6 A simple illustration (cross-plots) of the relation between per capita GDP and infrastructure perceptions of severe or very severe obstacles to growth in Africa

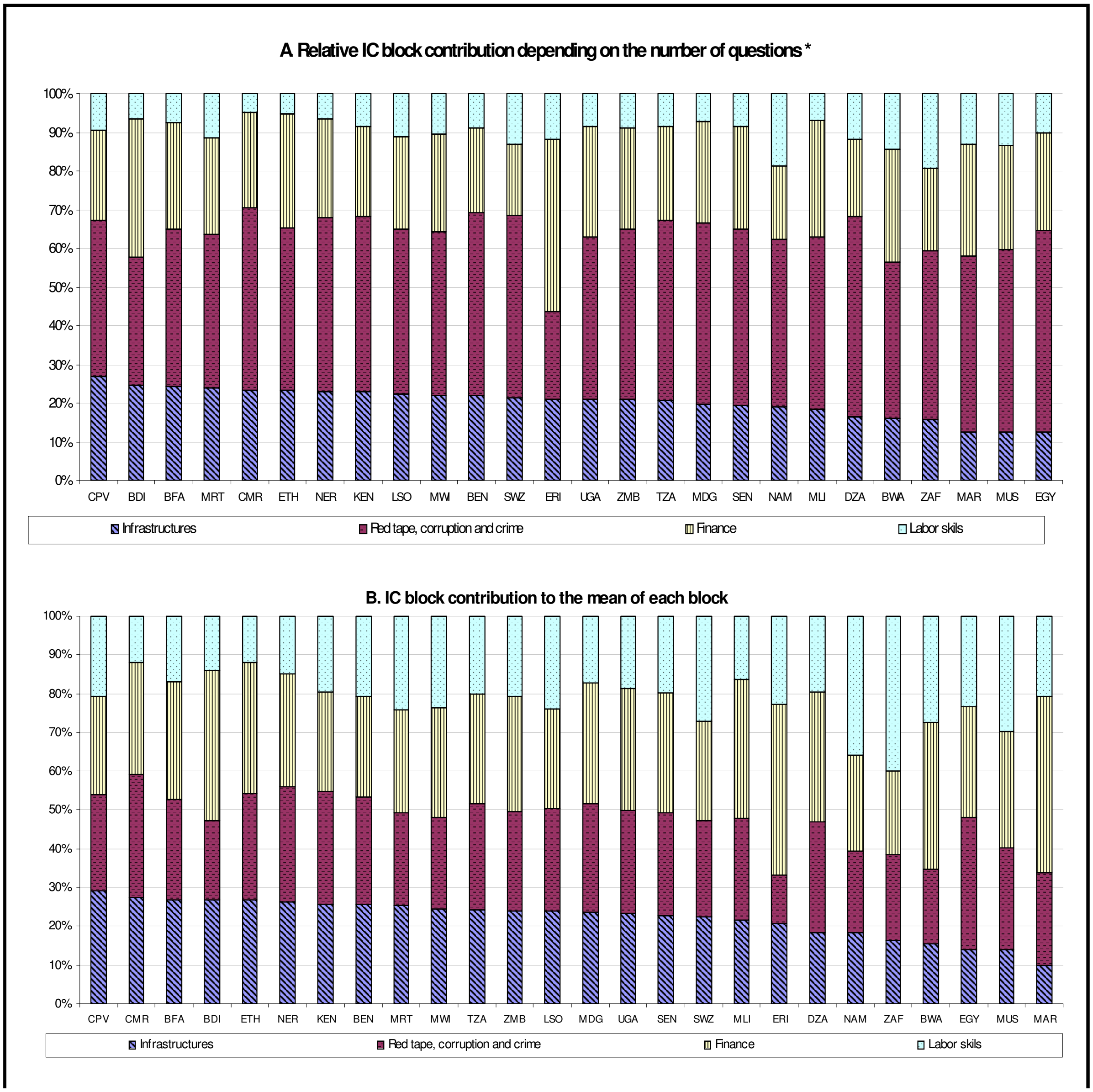
GDP per capita relative to United States



Source: Authors' calculations based on IC data.
Note: No data are available on perceptions of transport in Algeria.

Figures from section 5

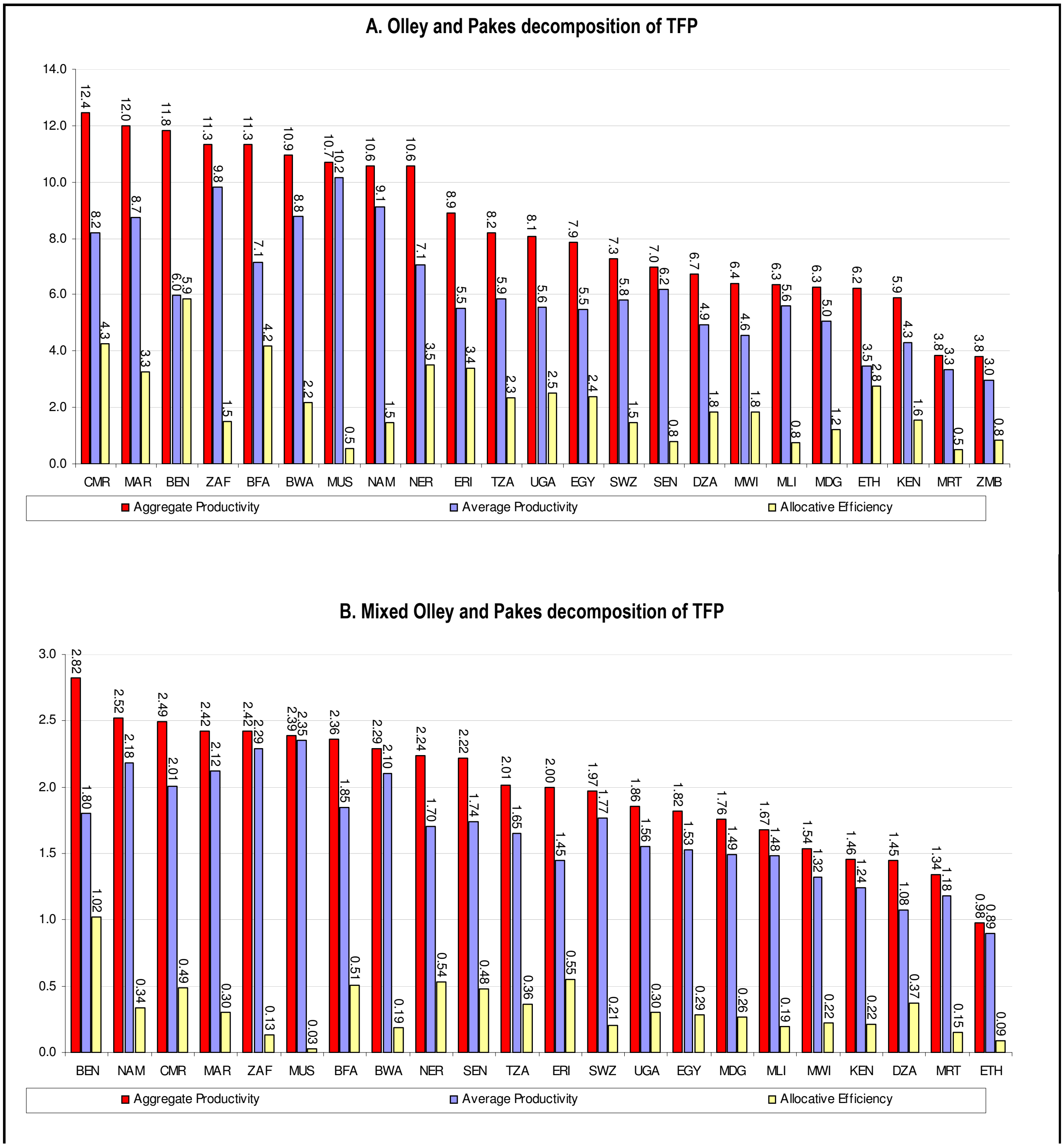
Figure 5.1 Rankings of firms' perceptions of severe and very severe obstacles to growth



Source: Authors' calculations from IC data.

Note: * = Number of questions on perceptions by blocks of IC variables: Infrastructure, 4 questions; red tape, corruption, and crime, 9 questions; finance and corporate governance, 2 questions; labor skills, 2 questions.

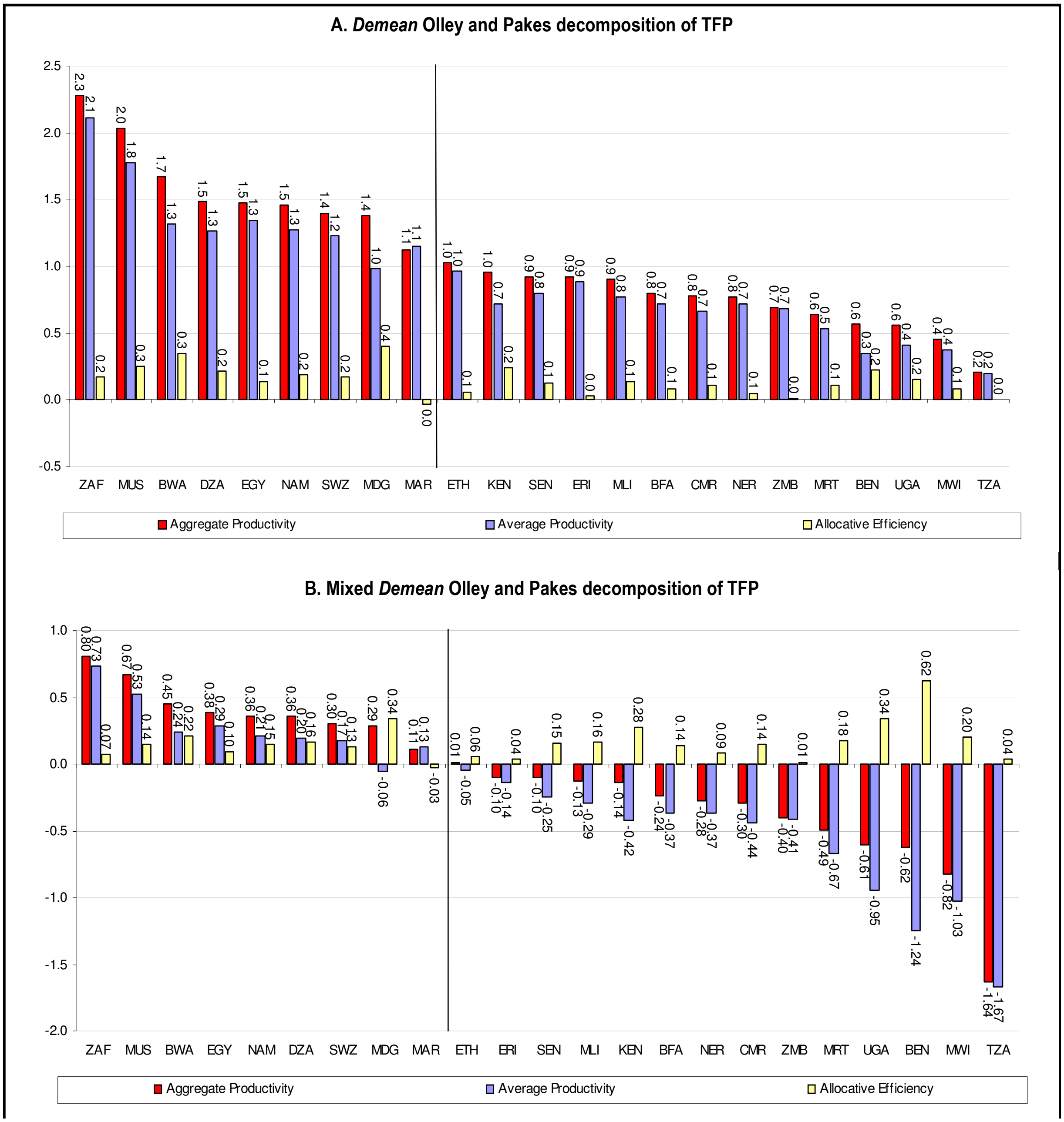
Figure 5.2 Olley and Pakes (O&P) decompositions of total factor productivity (TFP)



Source: Authors' calculations from IC data.

Notes: The Olley and Pakes (O&P) decomposition of TFP in levels is obtained from equation 4.4a of section 4. The mixed O&P decomposition is obtained from equation 4.4b. Sales in levels are used to compute the share of sales in both O&P decompositions.

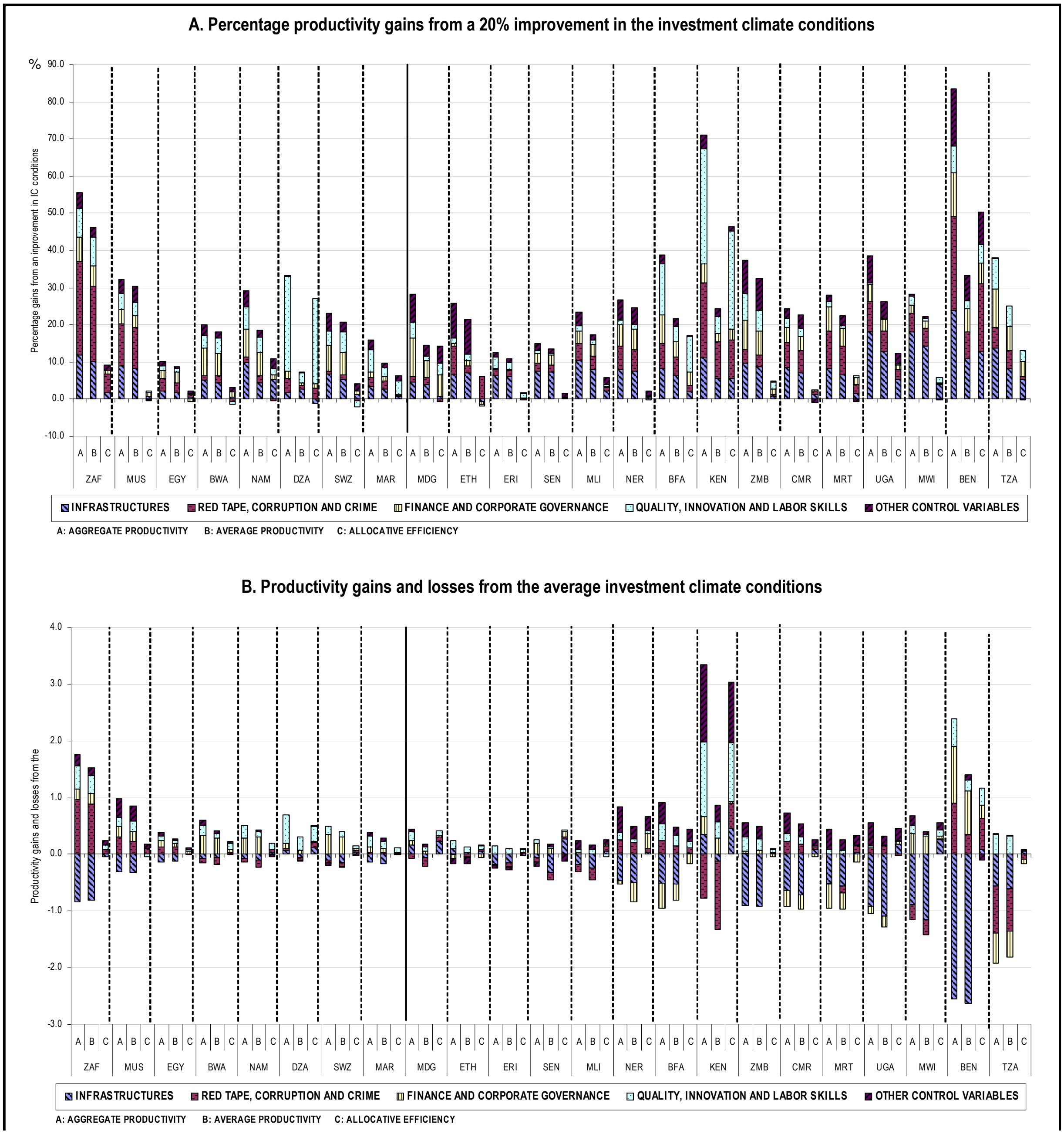
Figure 5.3 Demeaned O&P decompositions of TFP



Source: Authors' calculations from IC data.

Notes: The demeaned Olley and Pakes (O&P) decomposition of TFP in levels is given by equation 4.7. It is derived from equation 4.4a, using as the productivity measure the demeaned counterpart of the restricted Solow residual (see equation 4.3b) in levels. The demeaned mixed O&P decomposition comes from equation 4.4b, with the demeaned log-TFP of equation 4.3b in logs. Sales in levels are used to compute the share of sales in both O&P decompositions.

Figure 5.4 Demeaned productivity by groups of IC variables: simulations and average contributions

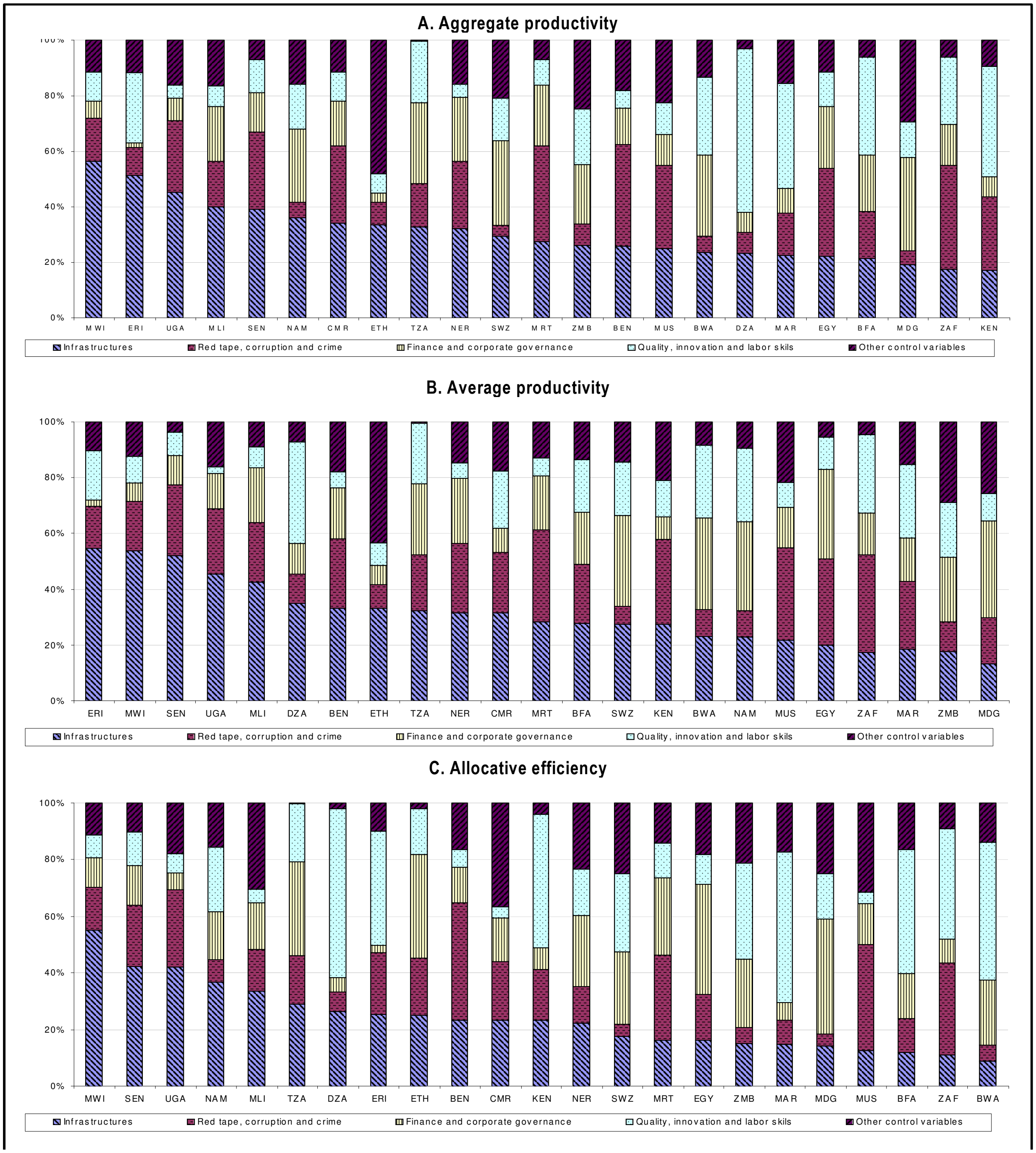


Source: Authors' calculations from IC data.

Note: The simulations are done variable by variable. The total percentage productivity gain from each group of variables (infrastructure; red tape, etc.) is computed as the sum of the individual productivity gains caused by the improvement in the IC variables of that group (one by one). Therefore, the final productivity gain should be interpreted in *ceteris paribus* terms: how much does productivity increase when the corresponding variable improves by 20 percent, holding everything else constant?

The productivity gains and losses from the average investment climate come from the decomposition of the demeaned Olley & Pakes decomposition in logs by groups of variables (4.8). The productivity gain or loss from the infrastructure group for each country is computed as the sum of the percentage contributions to average log-TFP caused by the average individual infrastructure variables. The same holds for the rest of the groups of IC and C variables.

Figure 5.5 Simulation of infrastructure absolute effects on productivity (20 percent improvement)



Source: Authors' calculations from IC data.

Note: The percentage contribution of the infrastructure group is computed as the sum of the absolute values of the percentage contributions of the individual infrastructure variables, divided by the cumulative sum in absolute terms of the percentage contributions of all the IC and C variables, including infrastructure. The holds for the rest of the IC blocks of variables.

Figure 5.6 Infrastructure absolute effects on productivity: Mixed *demeaned* O&P decomposition

Source: Authors' calculations from IC data.

Note: The percentage contribution of the infrastructure group is computed as the sum of the absolute values of the percentage contributions of the individual infrastructure variables, divided by the cumulative sum in absolute terms of the percentage contributions of all the IC and C variables, including infrastructure. The holds for the rest of the IC blocks of variables.

Figure 5.7. Cross-plot between demeaned aggregate productivity and GDP per capita (% of US)

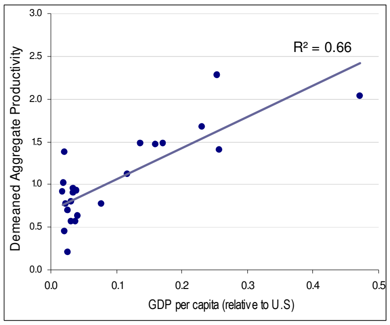


Figure 5.8. Cross-plot between demeaned aggregate productivity and ranking on the ease of doing business*

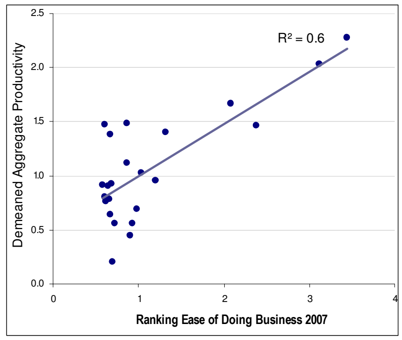
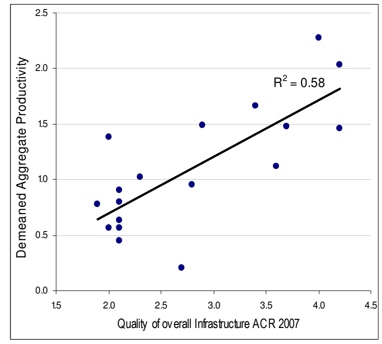


Figure 5.9. Cross-plot between demeaned aggregate productivity and quality of overall infrastructure from ACR 2007*



*Rank is computed as: (total number of firms in DBR-Rank)/ total number of firms in DBR

Source: Authors´ calculations with IC data, Doing Business Report (2007) and Penn World Table.

Figure 5.10. Cross-plot between demeaned aggregrate productivity and firms´ perceptions on infrastructure as an obstacle

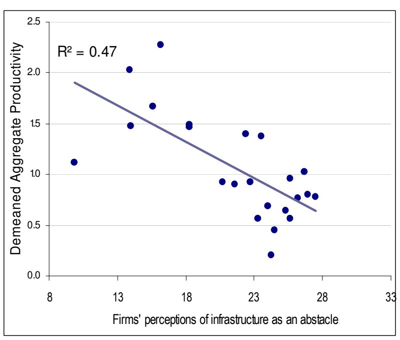


Figure 5.11. Cross-plot between demeaned aggregate productivity and percentage absolute contribution of infrastructure to average log-productivity

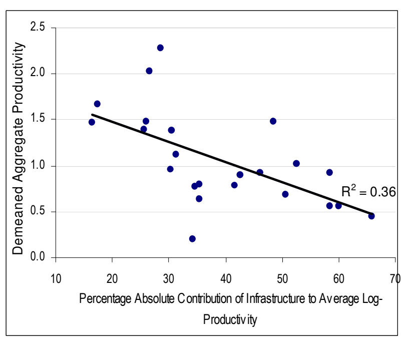
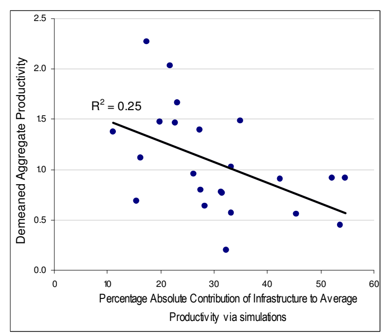


Figure 5.12. Cross-plot between demeaned aggregate productivity and percentage absolute contribution of infrastructure to average productivity via simulations



Source: Authors´ calculations with IC data.

Figure 5.13. Cross-plot between demeaned aggregate productivity and percentage absolute contributions of infrastructure to allocative efficiency (TFP in logs)

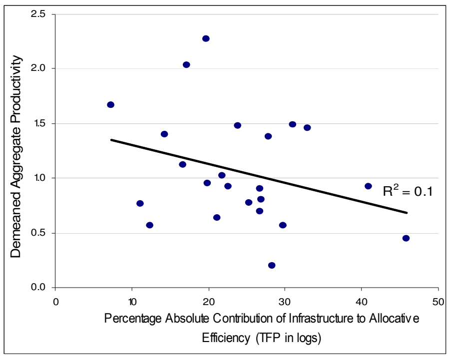


Figure 5.14. Cross-plot between demeaned aggregate productivity and percentage absolute contributions of infrastructure to allocative efficiency via simulations

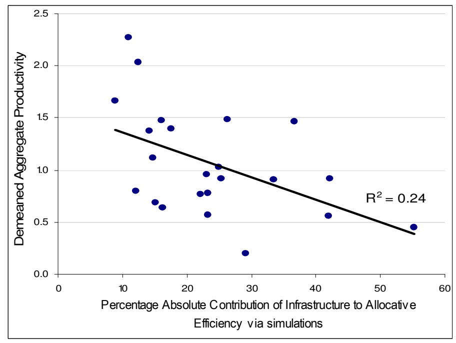


Figure 5.15. Cross-plot between percentage absolute contribution to average log–productivity and contributions via simulations

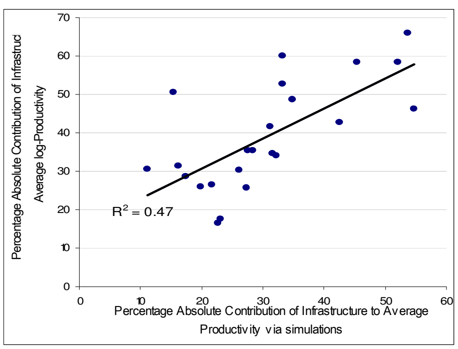
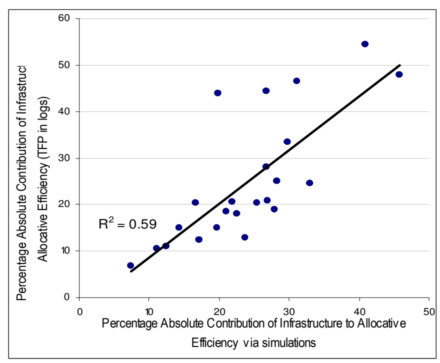


Figure 5.16. Cross-plot between percentage absolute contribution to allocative efficiency (with TFP in logs) and contributions via simulations



Source: Authors' calculations with IC data.

Figures from section 6

Figure 6.1 Impact of infrastructure on productivity in Algeria

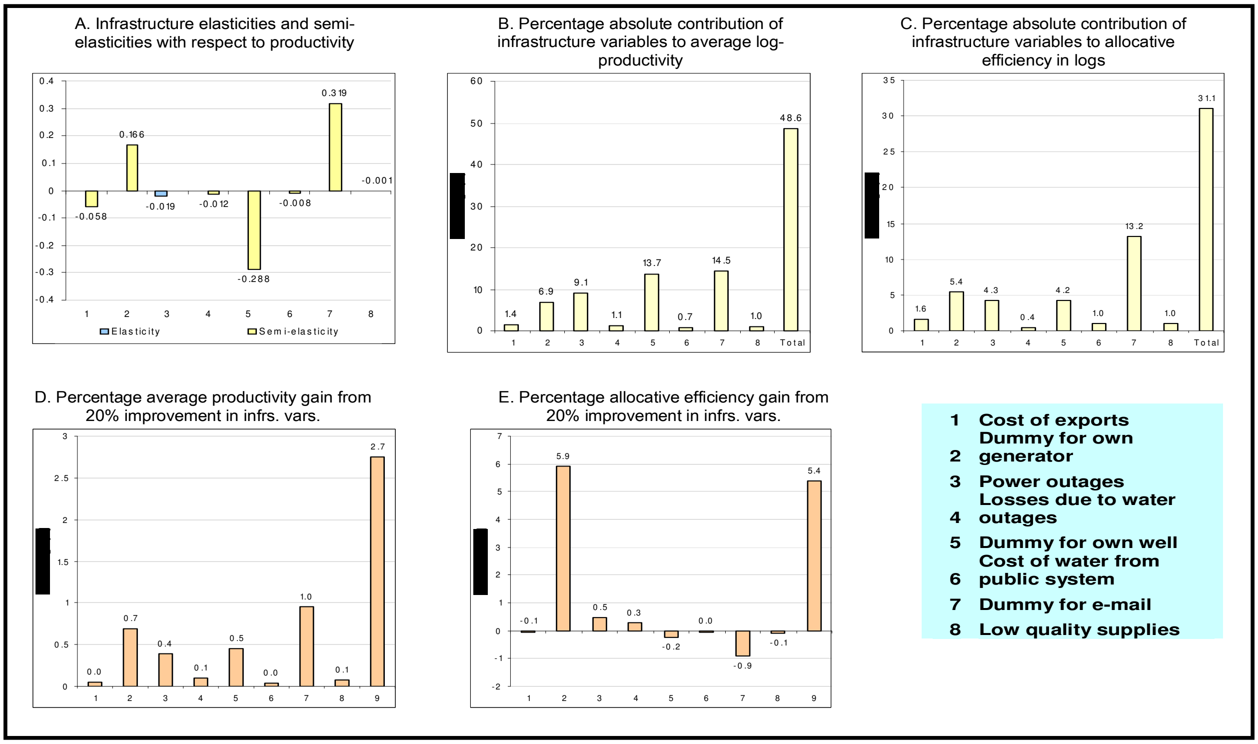


Figure 6.2 Impact of infrastructure on productivity in Benin

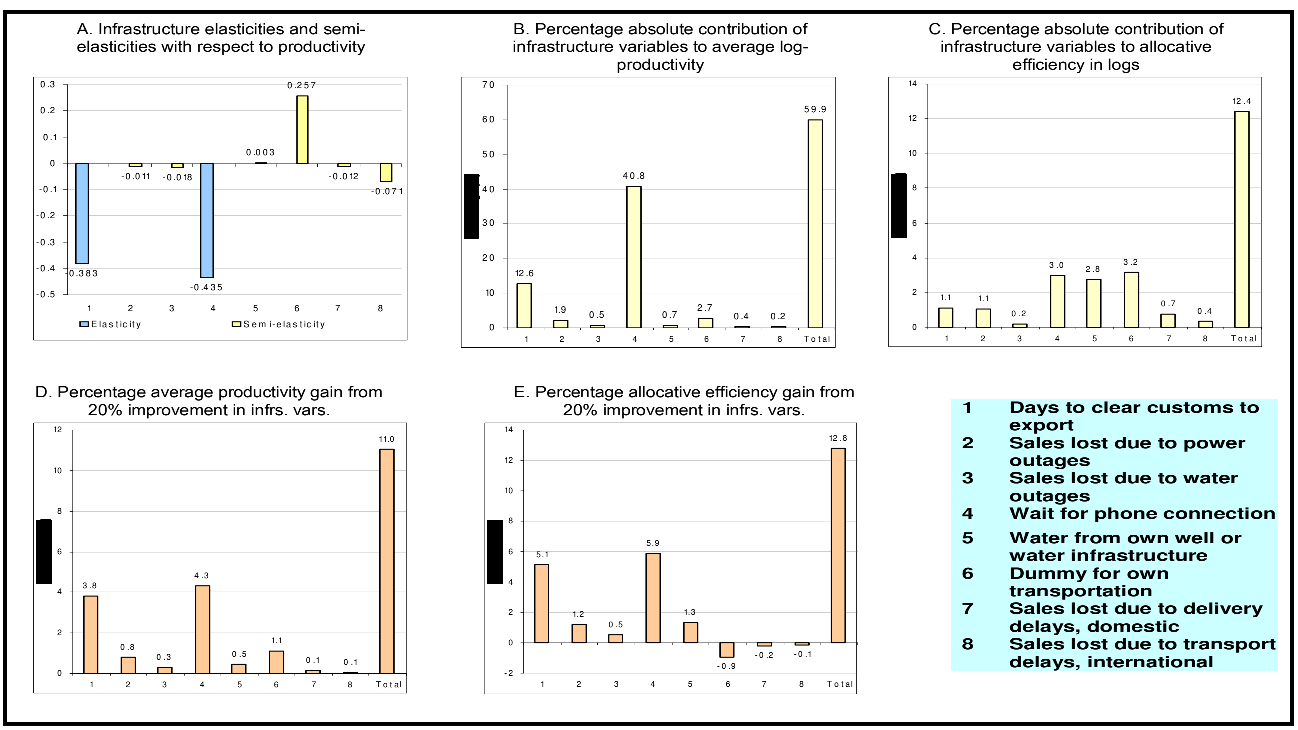


Figure 6.3 Impact of infrastructure on productivity in Botswana

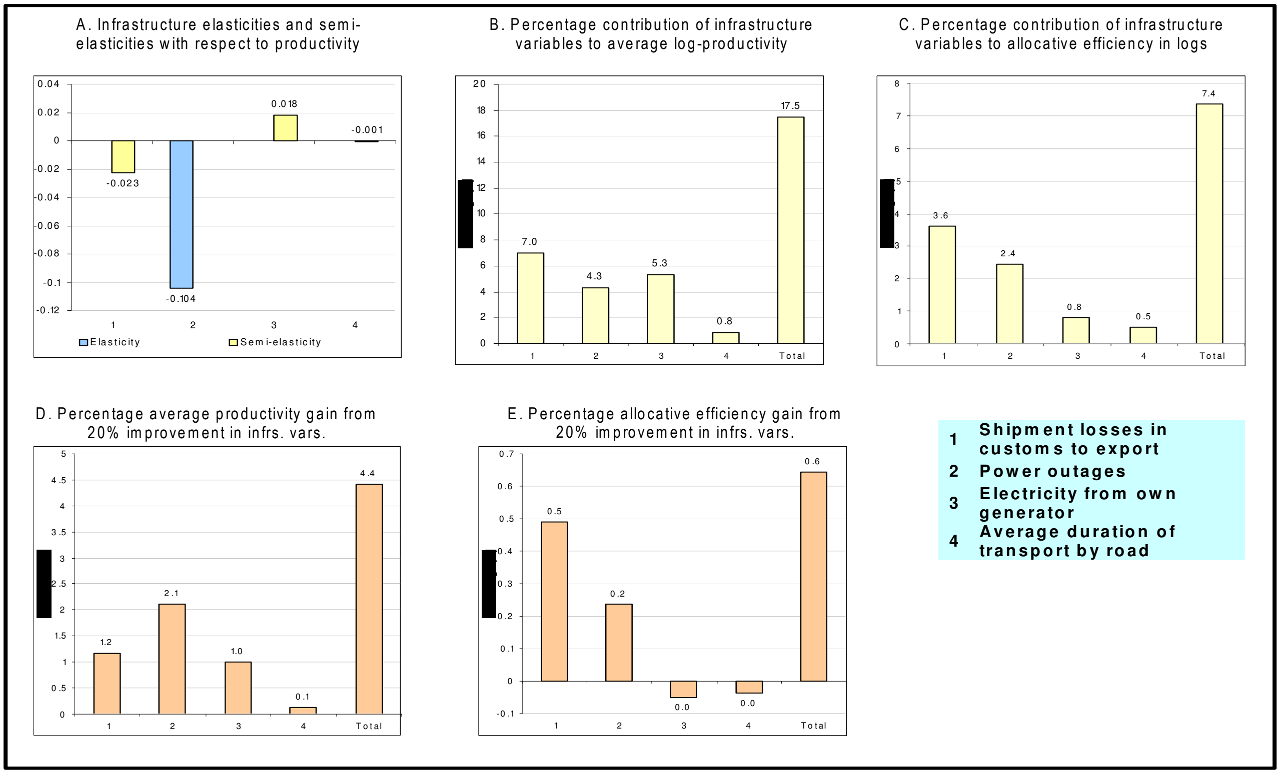


Figure 6.4 Impact of infrastructure on productivity in Burkina Faso

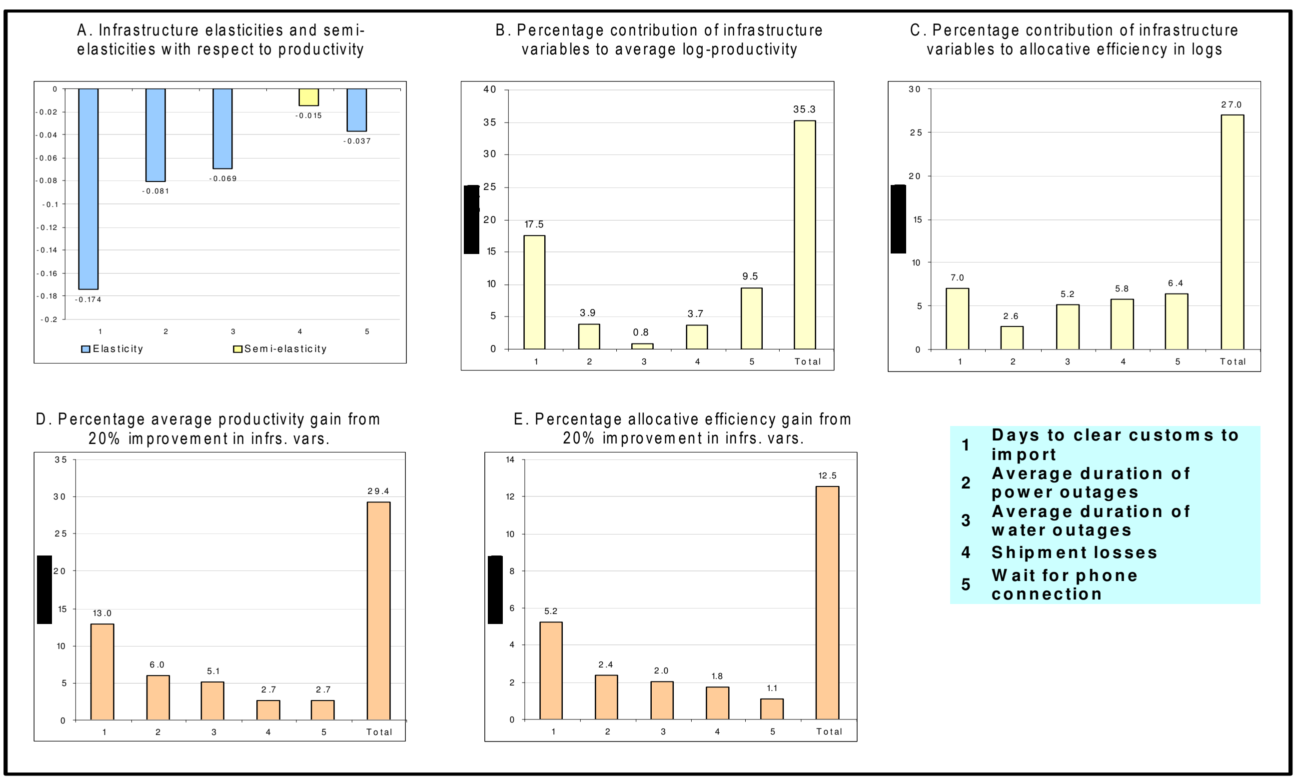


Figure 6.5 Impact of infrastructure on productivity in Cameroon

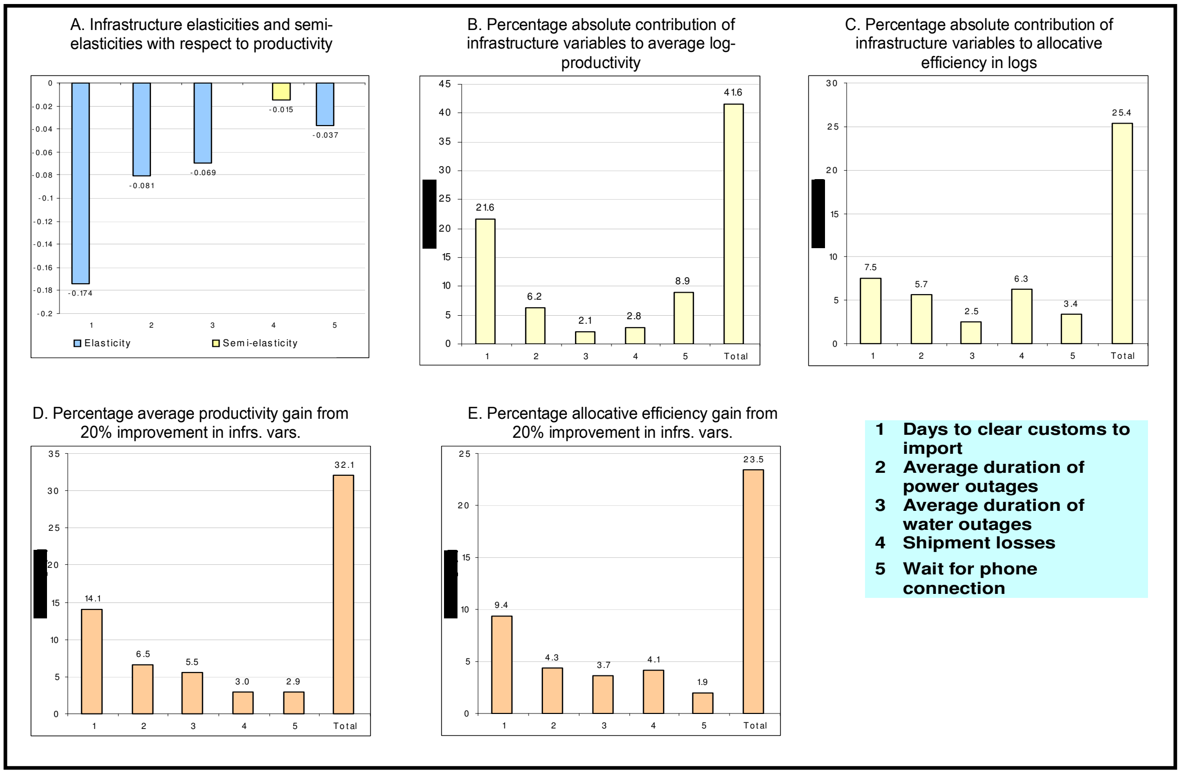


Figure 6.6 Impact of infrastructure on productivity in Egypt

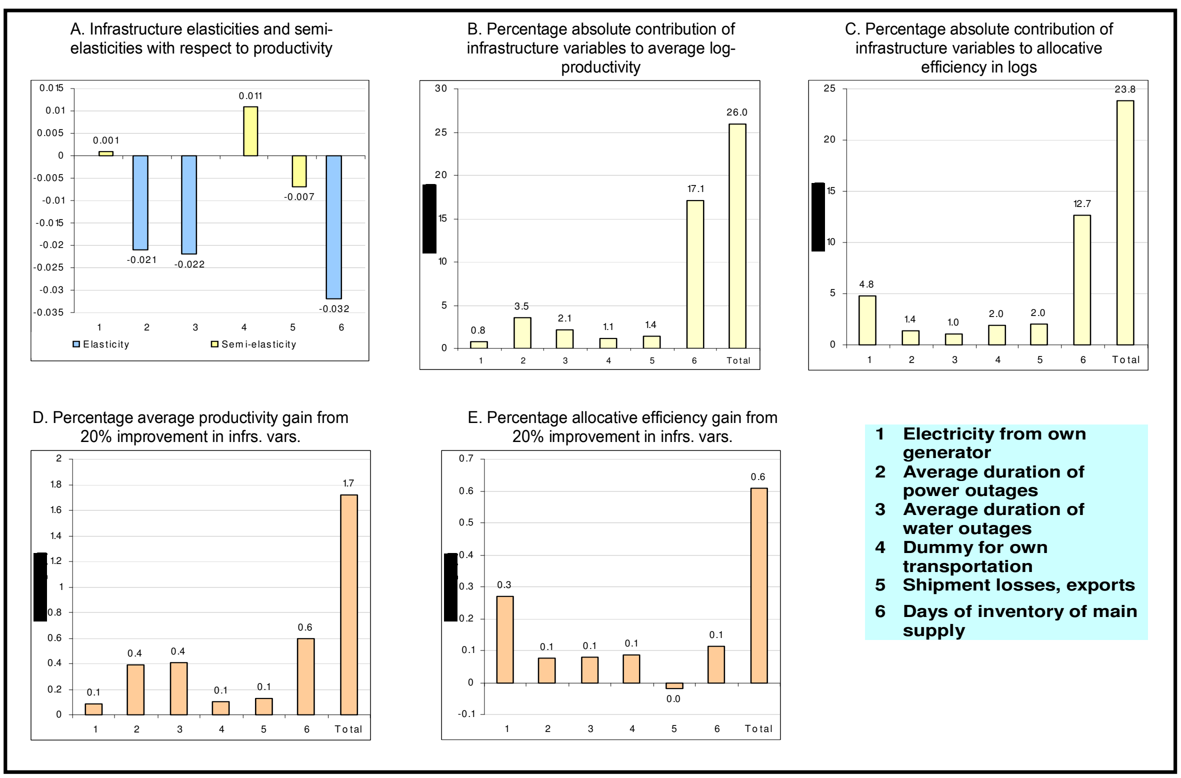


Figure 6.7 Impact of infrastructure on productivity in Eritrea

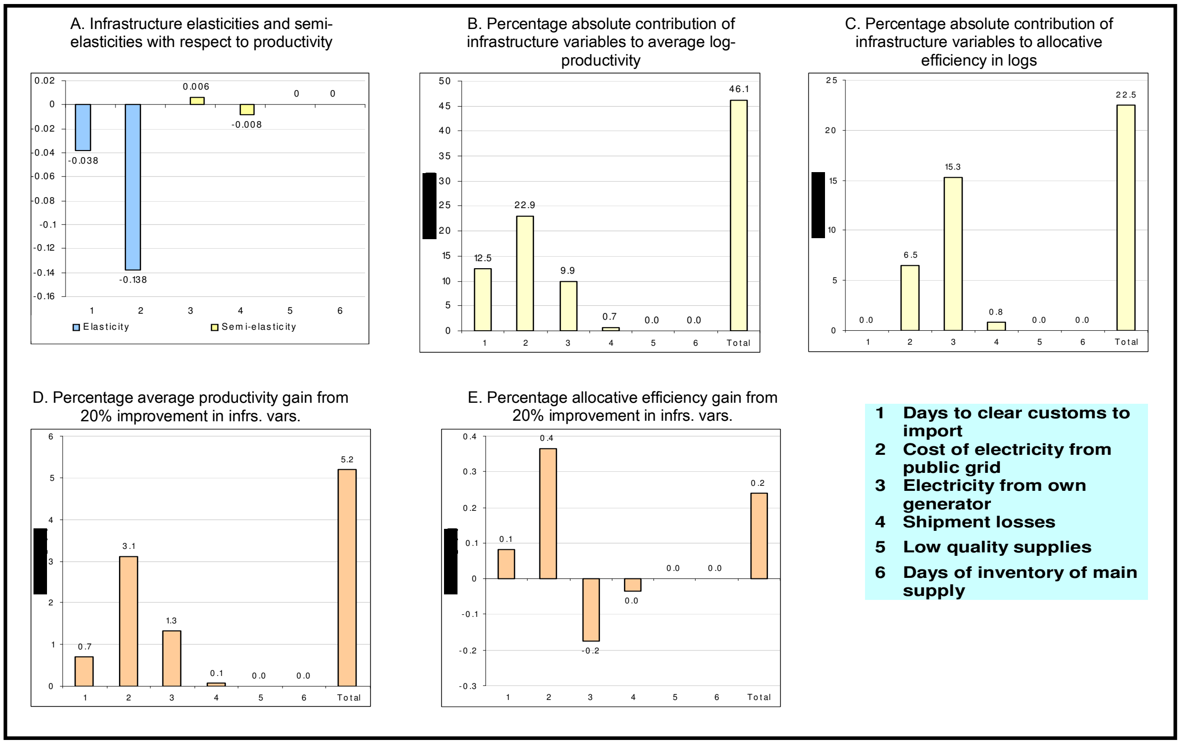


Figure 6.8 Impact of infrastructure on productivity in Ethiopia

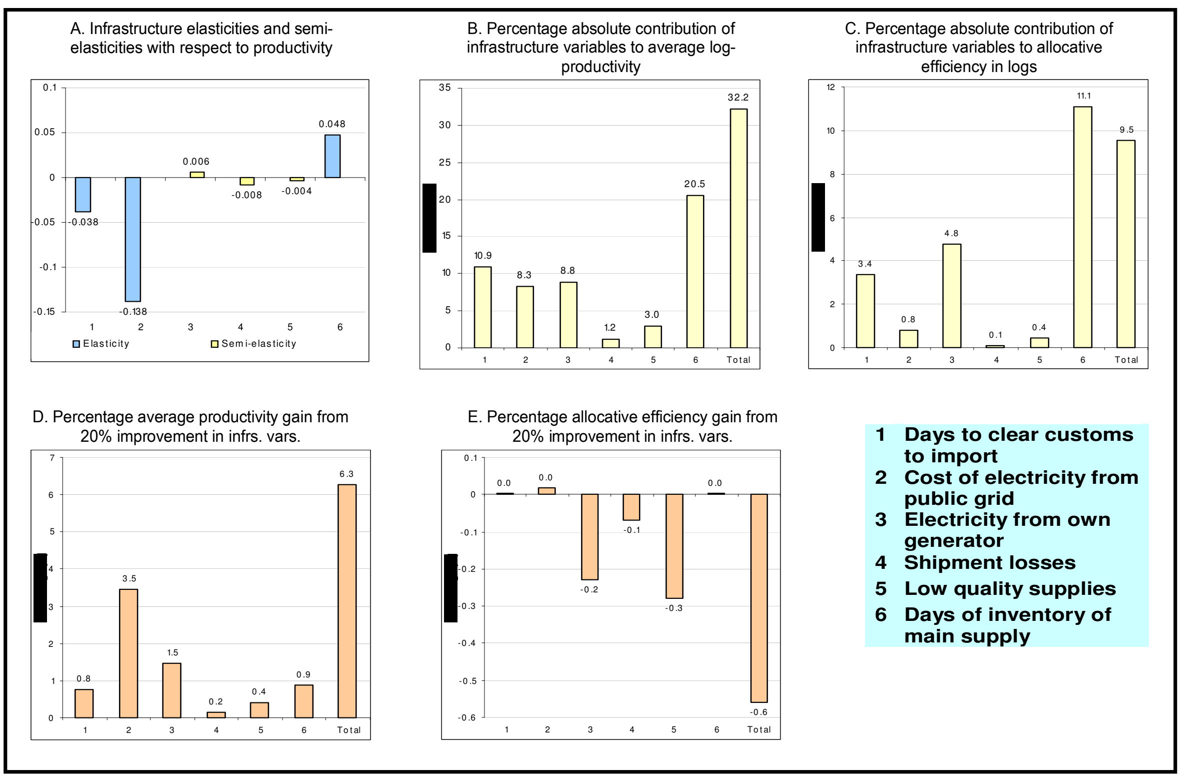


Figure 6.9 Impact of infrastructure on productivity in Kenya

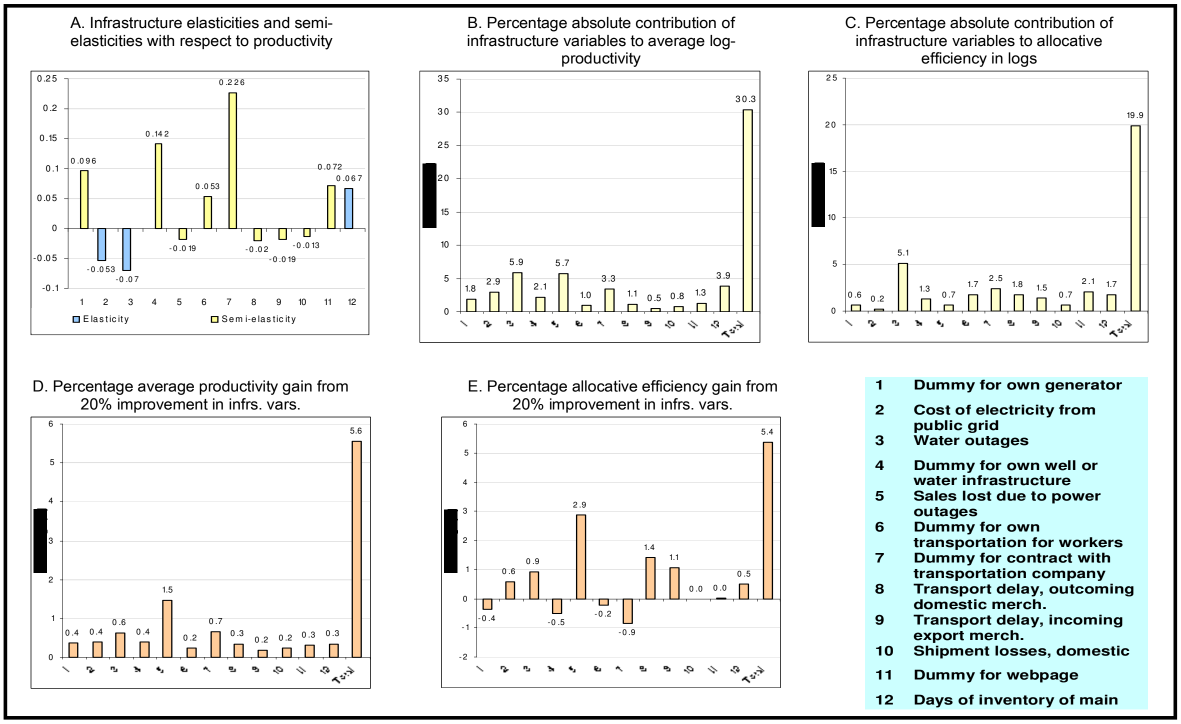


Figure 6.10 Impact of infrastructure on productivity in Madagascar

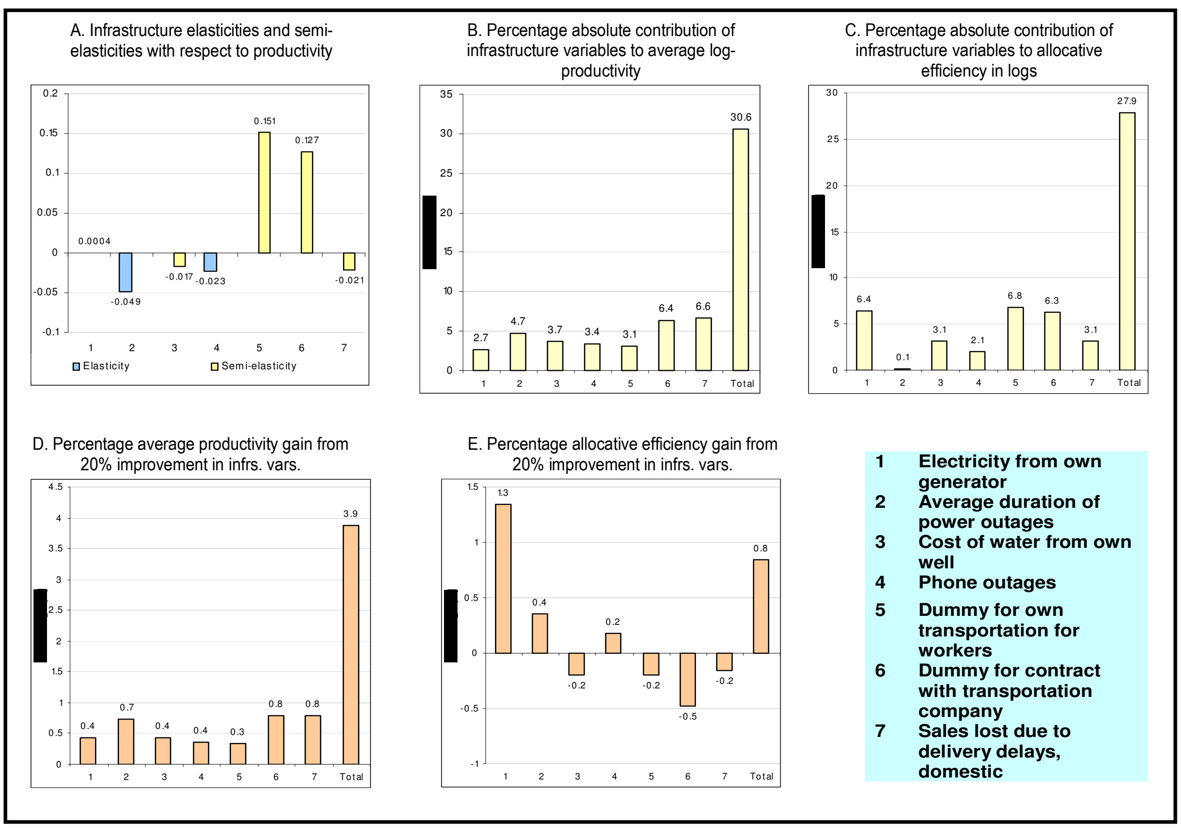


Figure 6.11 Impact of infrastructure on productivity in Malawi

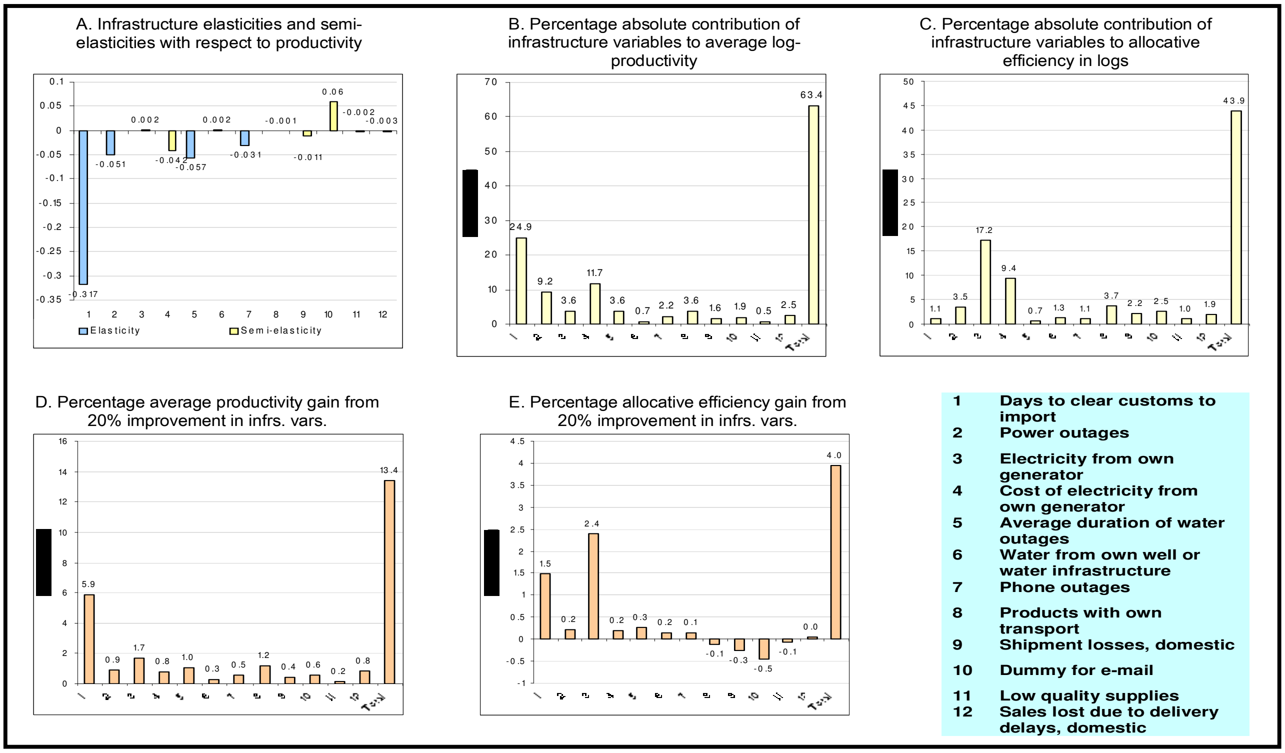


Figure 6.12 Impact of infrastructure on productivity in Mali

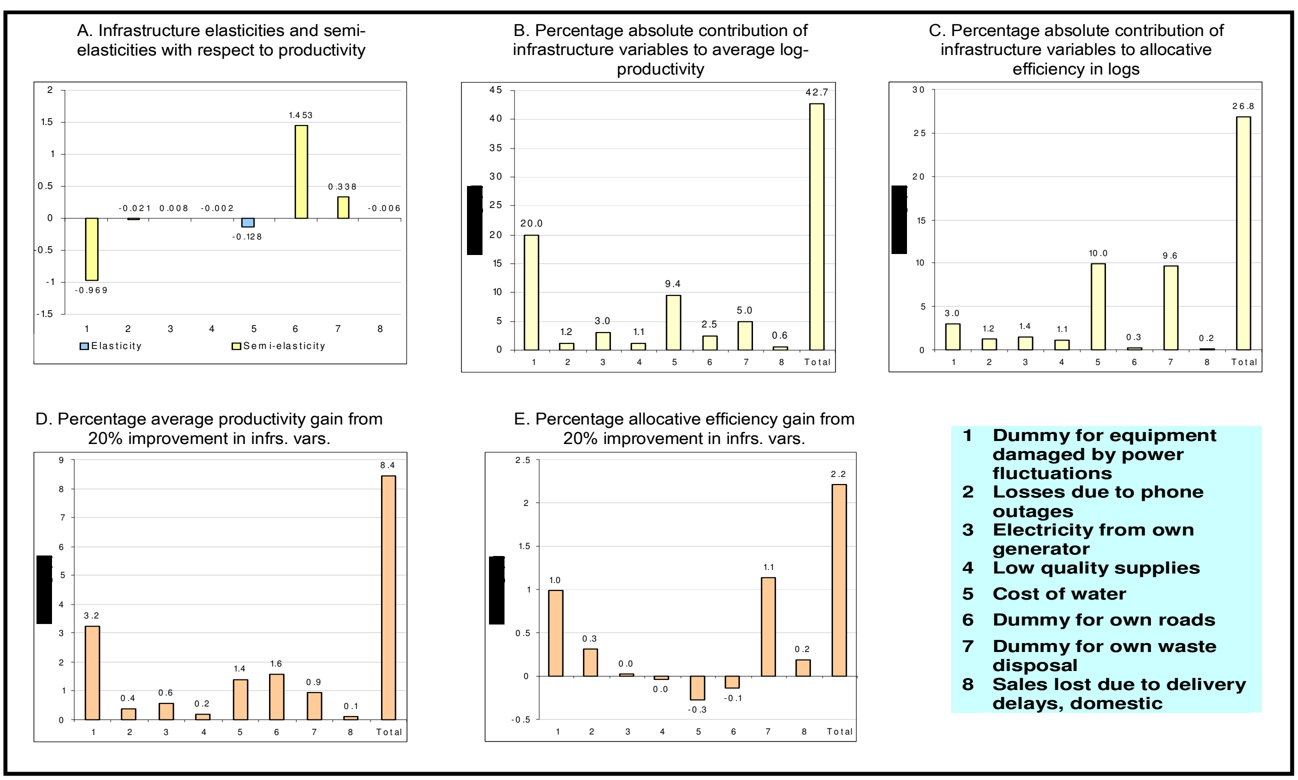


Figure 6.13 Impact of infrastructure on productivity in Mauritania

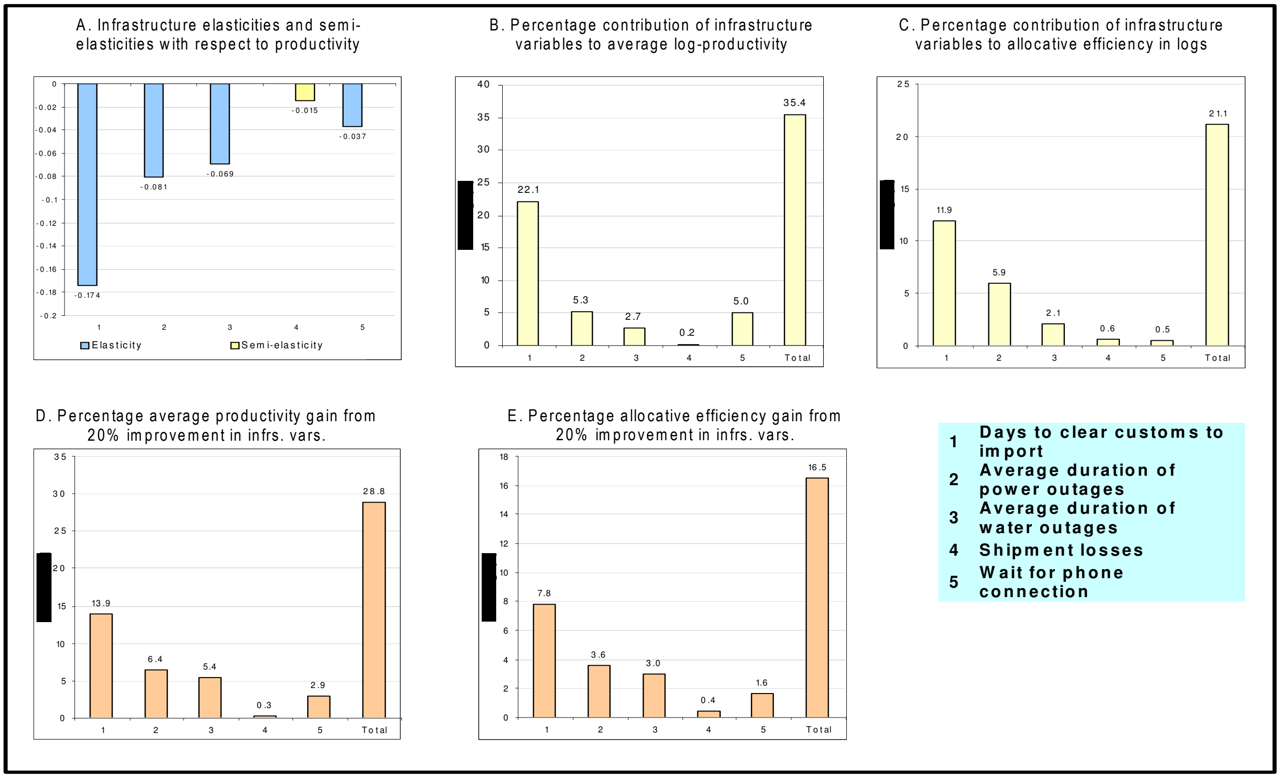


Figure 6.14 Impact of infrastructure on productivity in Mauritius

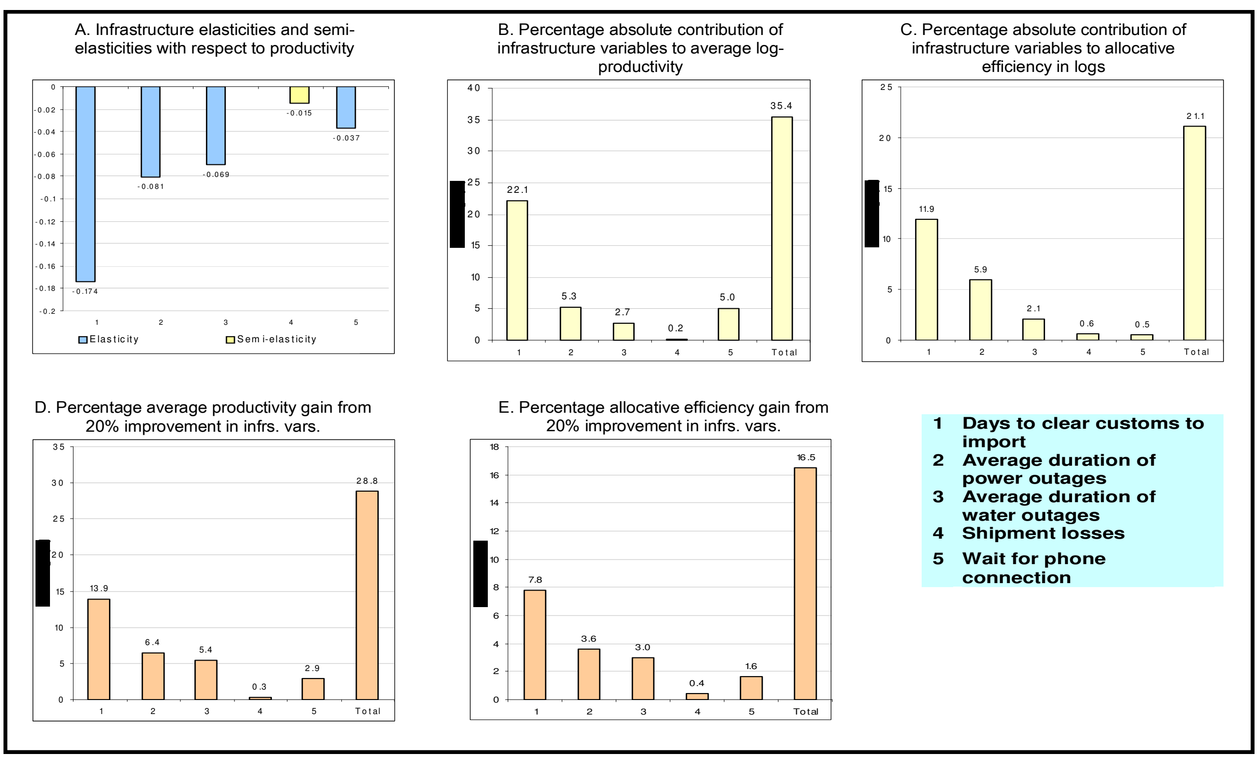


Figure 6.15 Impact of infrastructure on productivity in Morocco

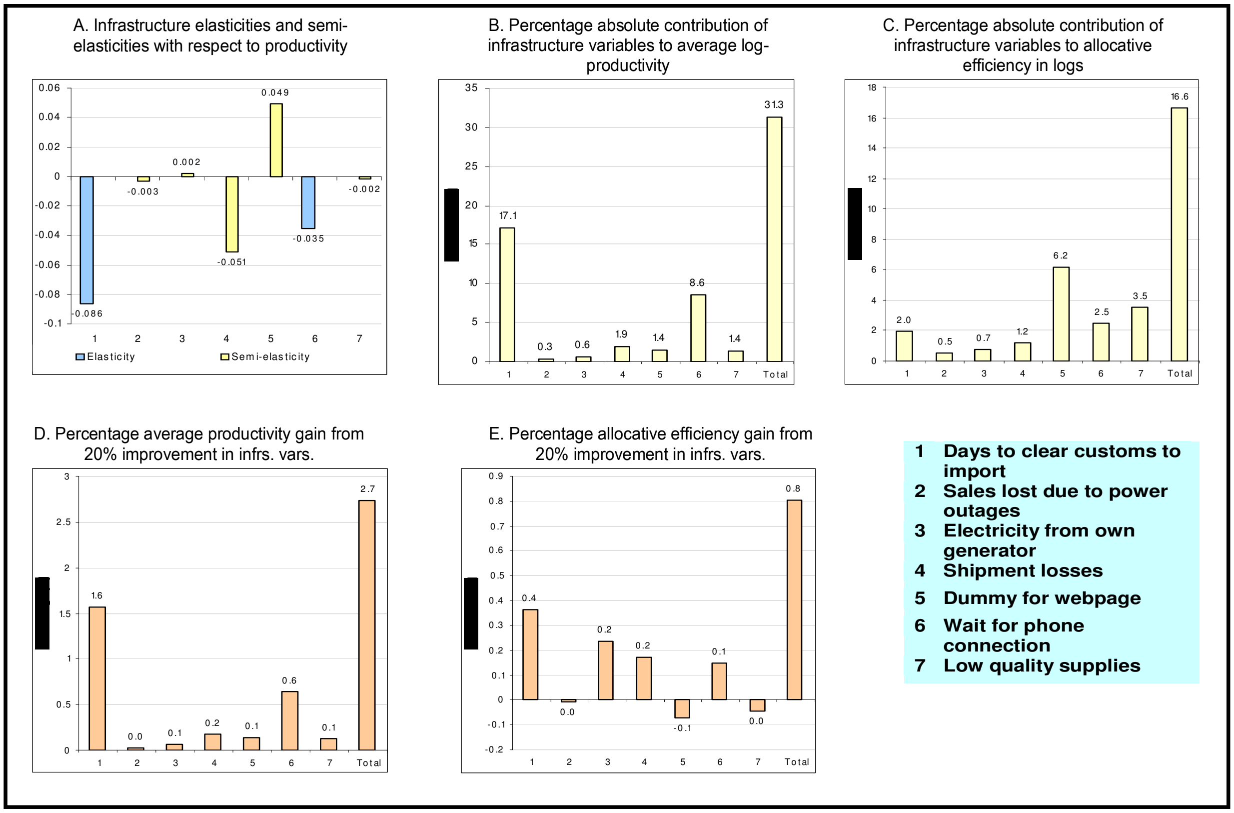


Figure 6.16 Impact of infrastructure on productivity in Namibia

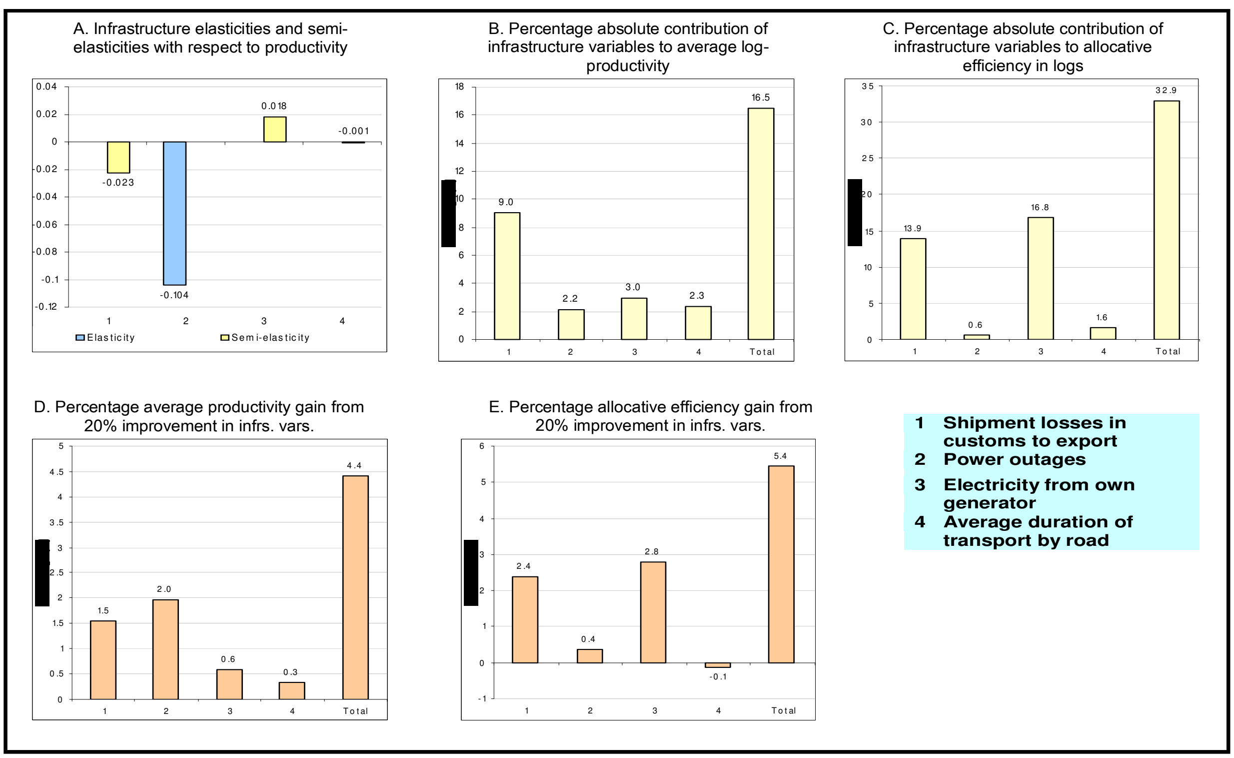


Figure 6.17 Impact of infrastructure on productivity in Niger

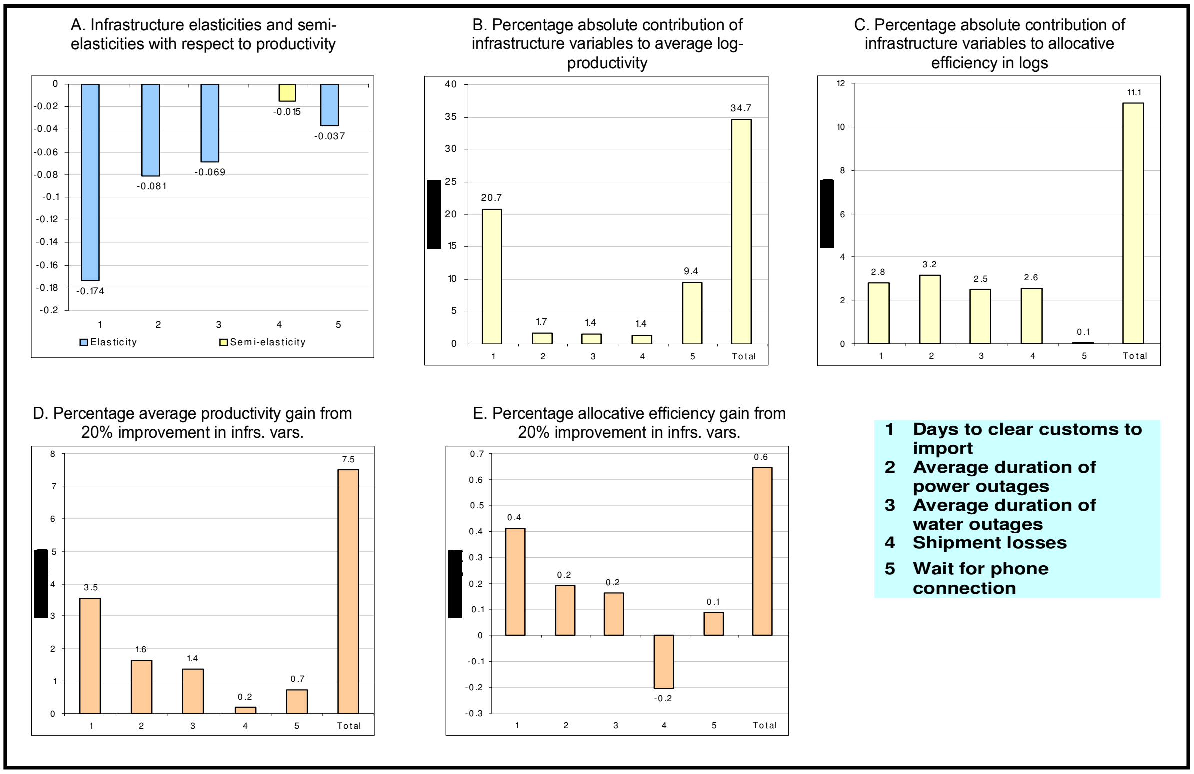


Figure 6.18 Impact of infrastructure on productivity in Senegal

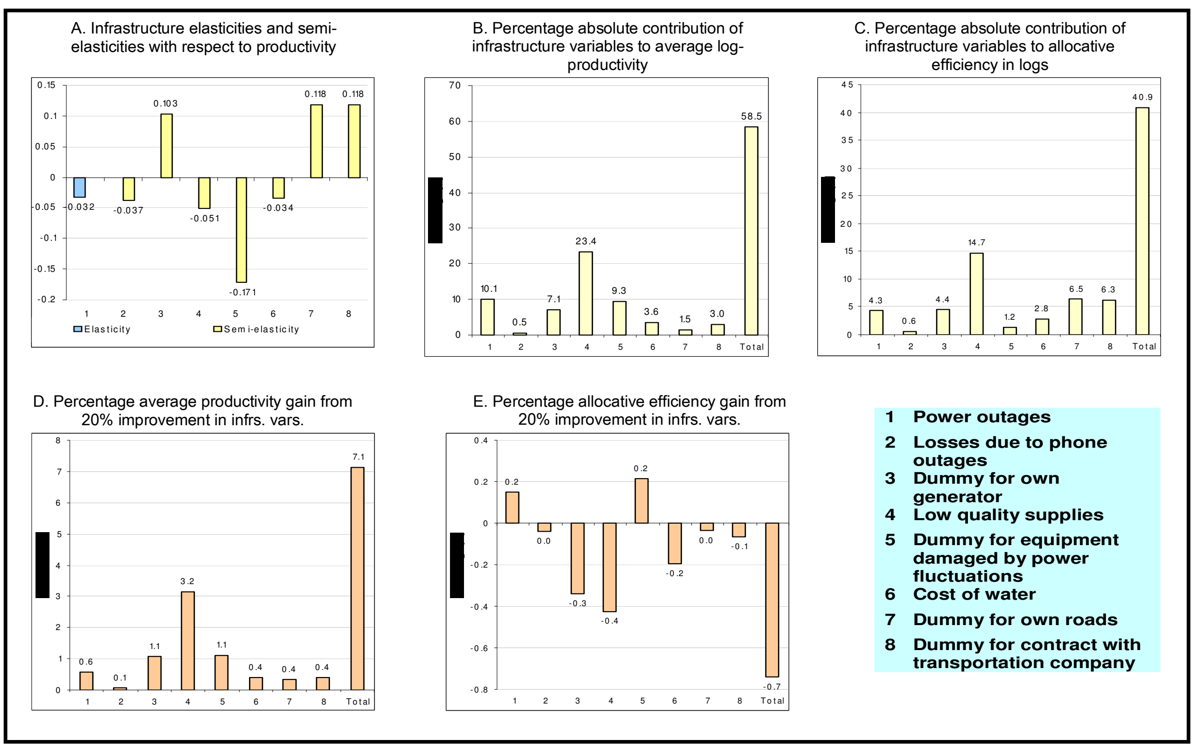


Figure 6.19 Impact of infrastructure on productivity in South Africa

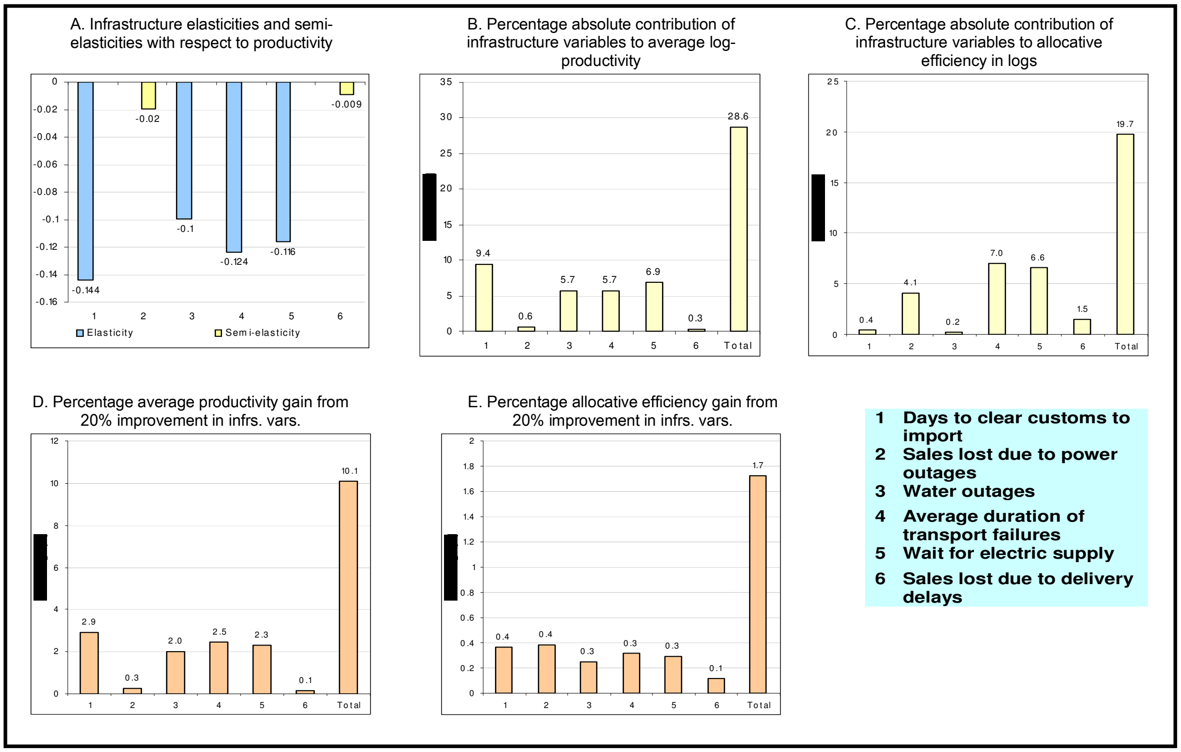


Figure 6.20 Impact of infrastructure on productivity in Swaziland

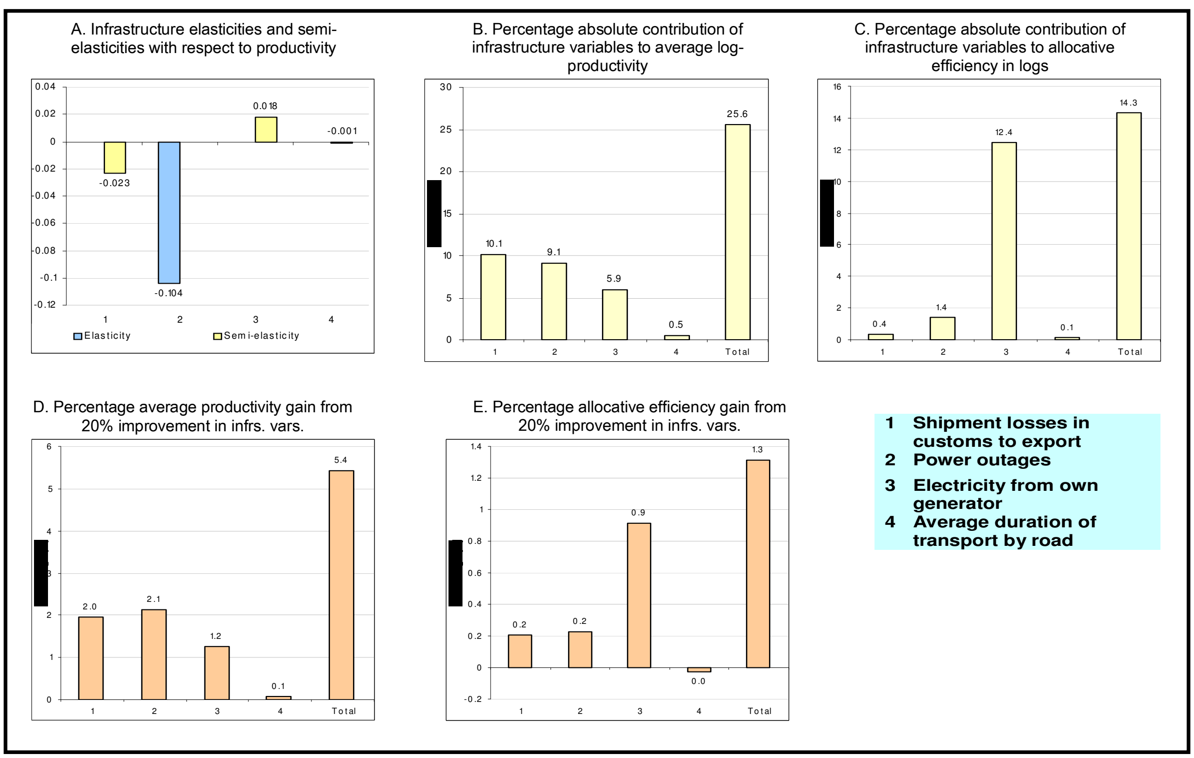


Figure 6.21 Impact of infrastructure on productivity in Tanzania

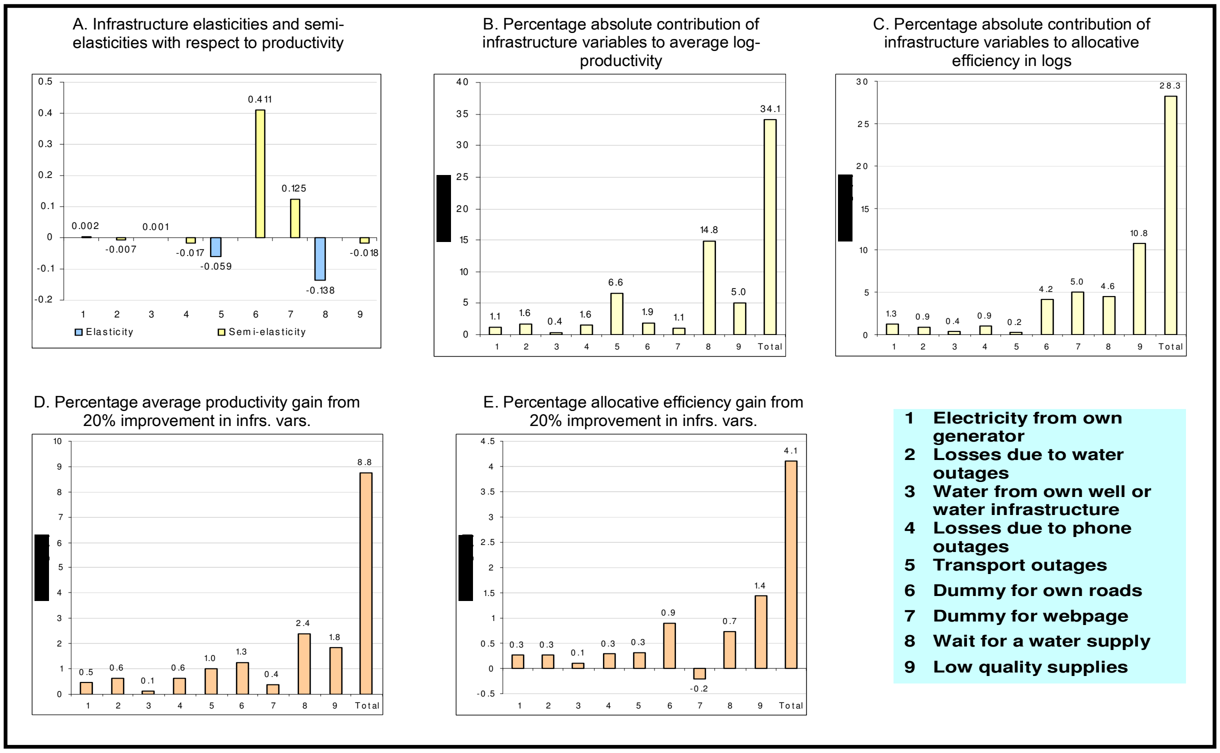


Figure 6.22 Impact of infrastructure on productivity in Uganda

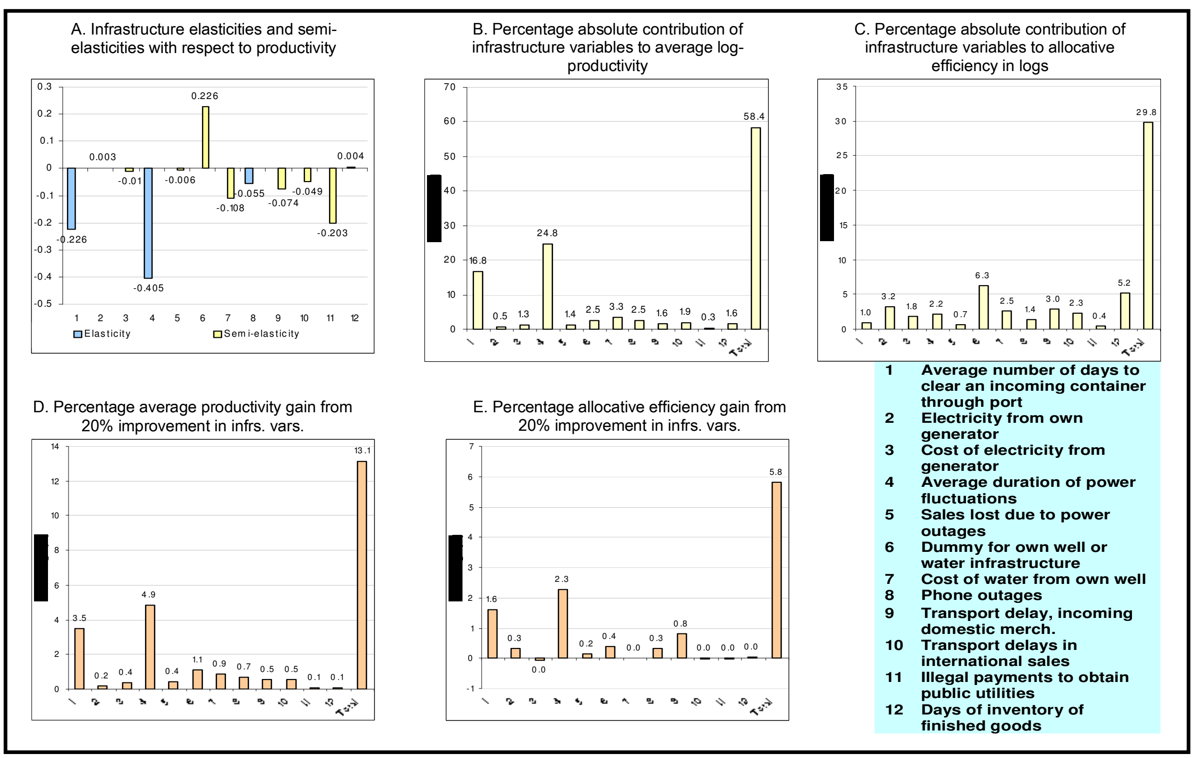


Figure 6.23 Impact of infrastructure on productivity in Zambia

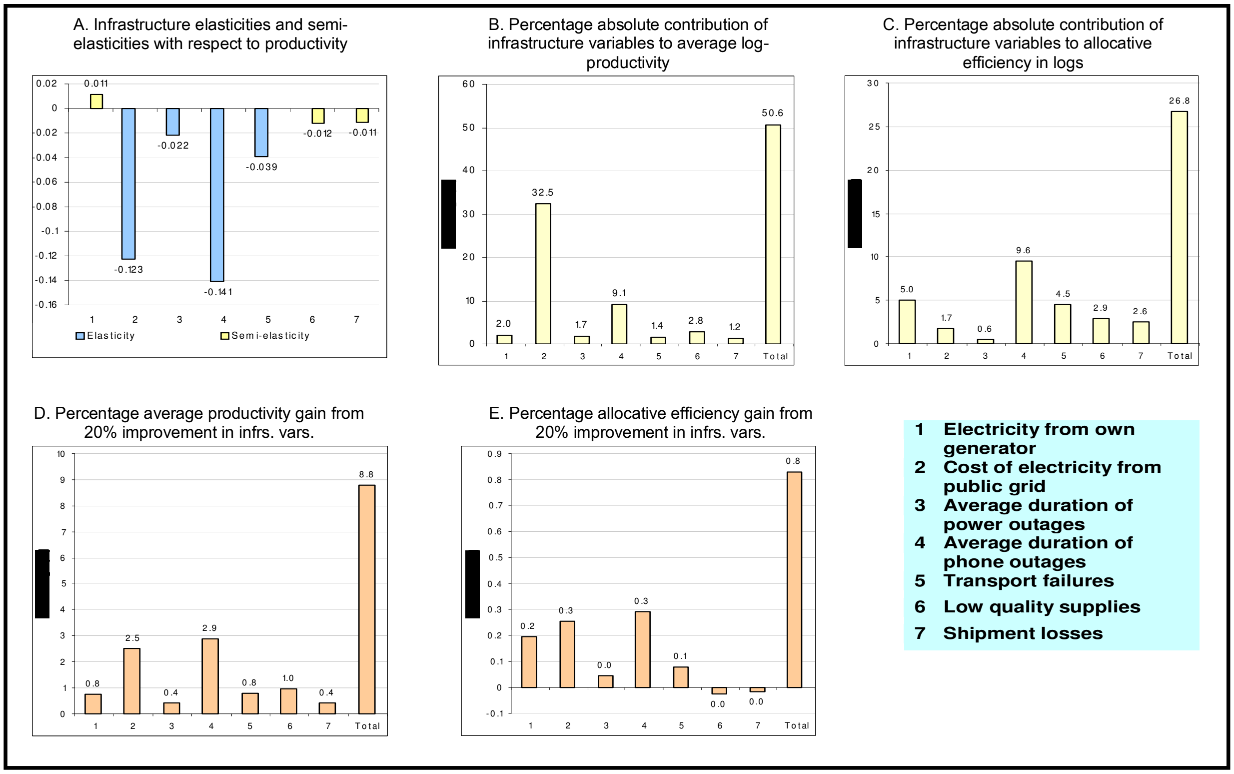
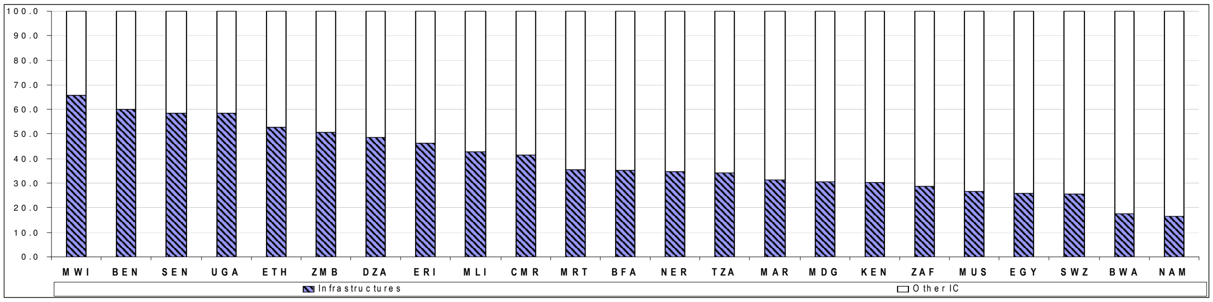
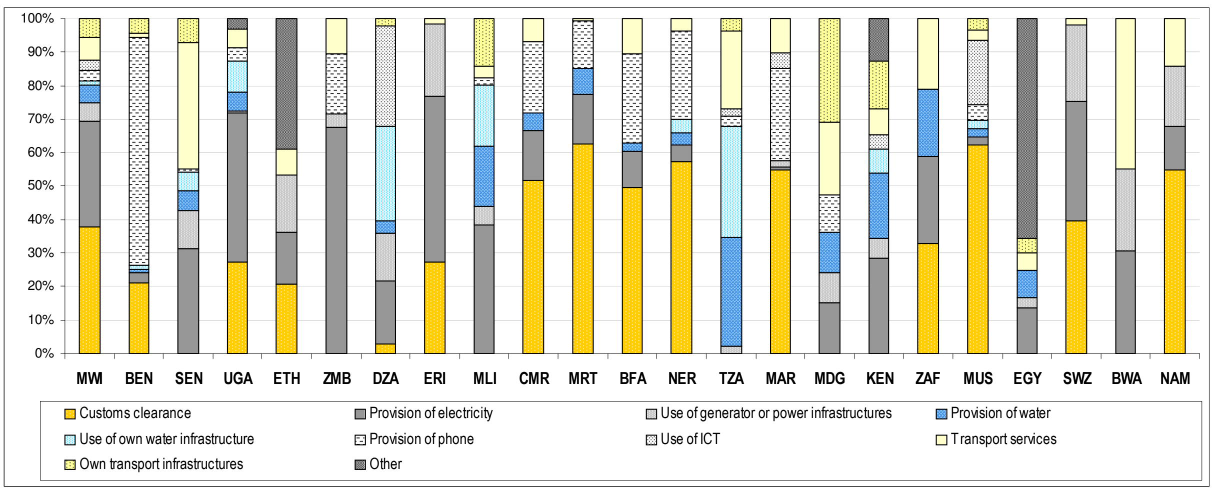


Figure 6.24 Infrastructure's impact on average log productivity by key factors (I)

A. Percentage absolute contribution of infrastructure to average log-productivity



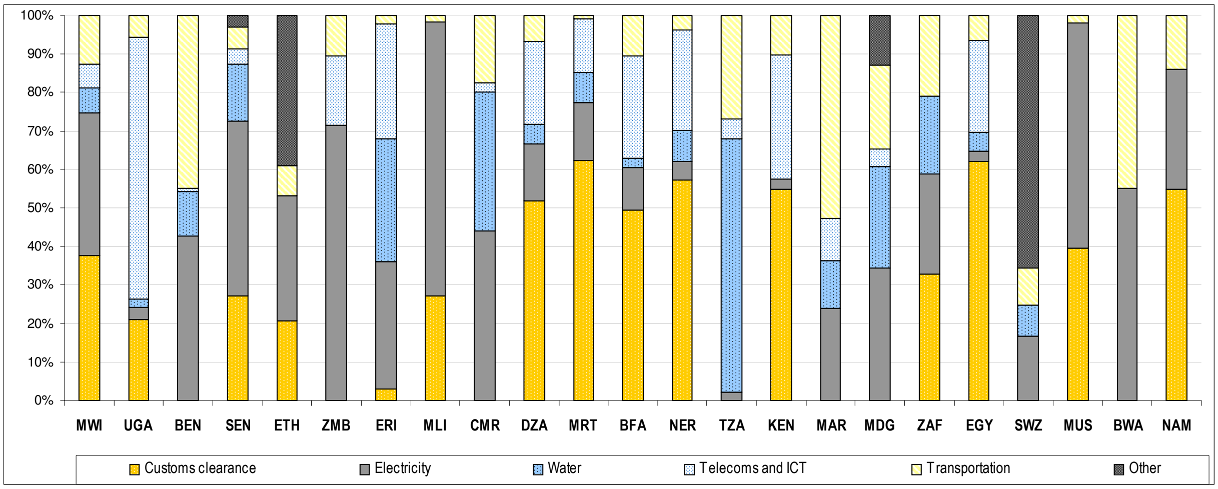
B. Percentage absolute contribution of infrastructure to average log-productivity by key factors



Source: Author's calculations with IC data.

Note: Customs clearance includes: days to clear customs to export and import; shipment losses in customs; inspections in customs; wait for an import license. Provision of electricity includes: power outages; avg. duration of power outages; losses due to power outages, wait for an electricity supply; power fluctuations; avg. duration of power fluctuations; cost of electricity from the public grid; cost of electricity from private system. Use of power infrastructures includes: dummy for own generator; electricity from own generator; dummy for own power infrastructures (excl. generators). Provision of water includes: water outages; avg. duration of water outages; losses due to water outages, wait for a water supply; cost of water from the public grid; cost of water from private system. Use of water infrastructures includes: dummy for own water infrastructures; water from own well. Provision of phone includes: phone outages; avg. duration of phone outages; losses due to phone outages, wait for a phone connection. Use of ICT includes: dummy for e-mail; dummy for webpage. Transport services includes: sales lost due to transport delays; sales lost due to delivery delays; shipment losses; low quality supplies; transport delays. Own transport infrastructures include: dummy for own roads; dummy for own transportation for workers; products with own transport. Other: inventories, illegal payments to obtain public utilities.

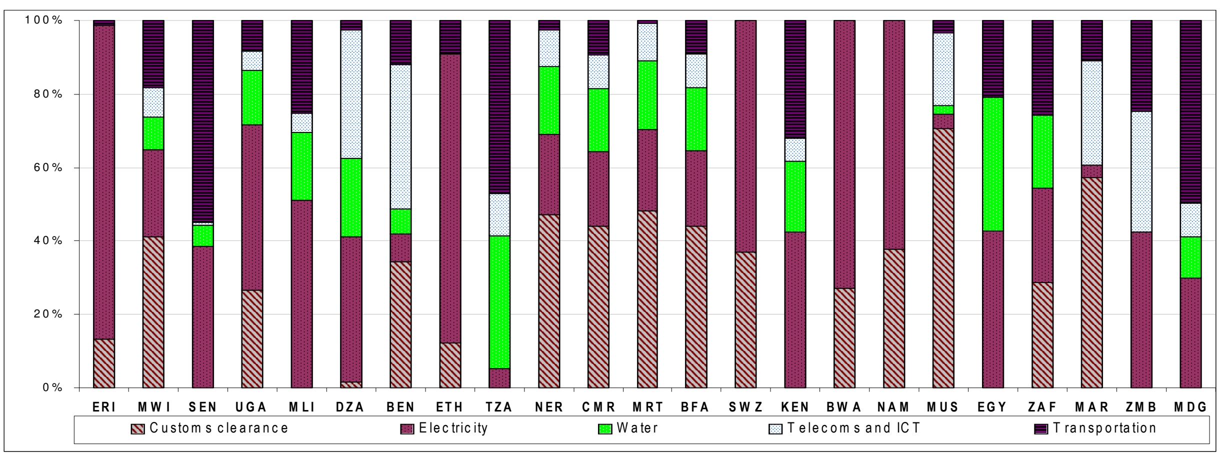
Figure 6.25 Infrastructure impact on average log productivity by key factors (II)



Source: Author's calculations with IC data.

Note: For a description of the variables contained in each group see footnote in Figure 6.24

Figure 6.26 Infrastructure's impact on average productivity by key factors via simulations

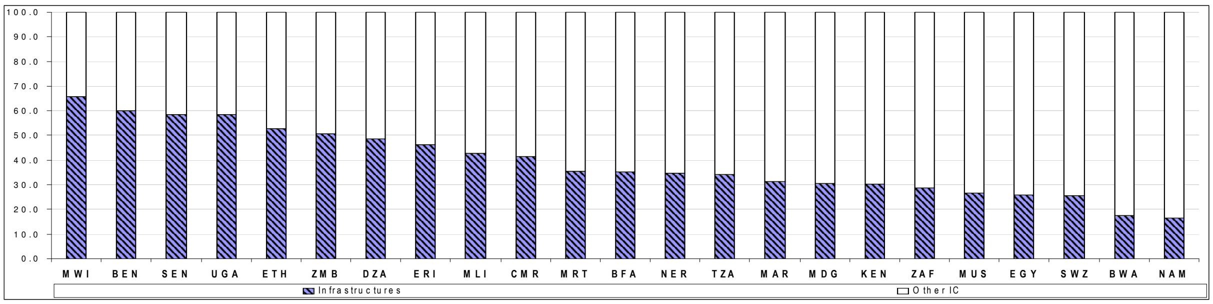


Source: Author's calculations with IC data.

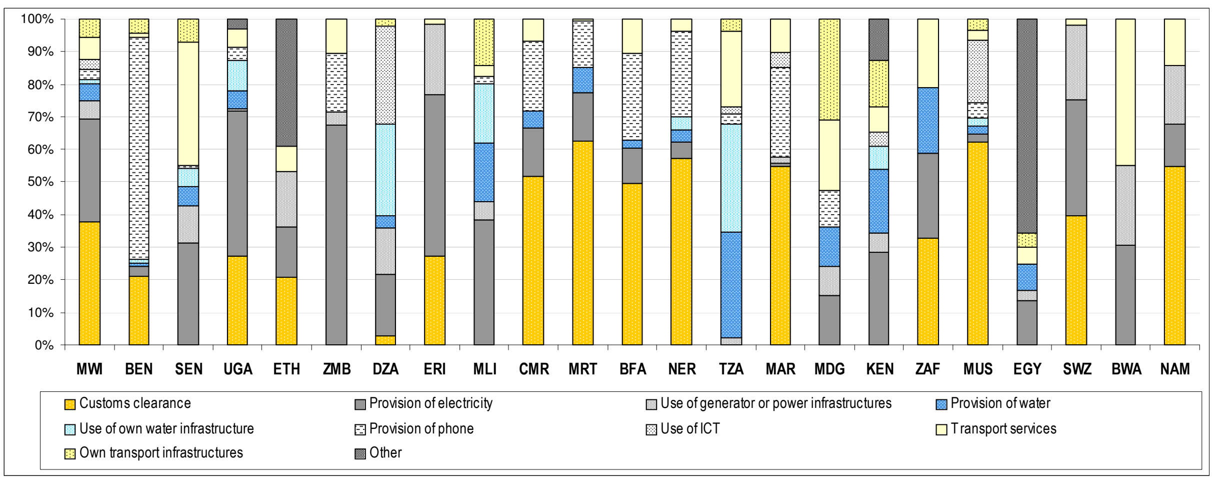
Note: For a description of the variables contained in each group see footnote in Figure 6.24

Figure 6.27 Infrastructure impact on allocative efficiency in logs by key factors (I)

A. Percentage absolute contribution of infrastructure to allocative efficiency in logs



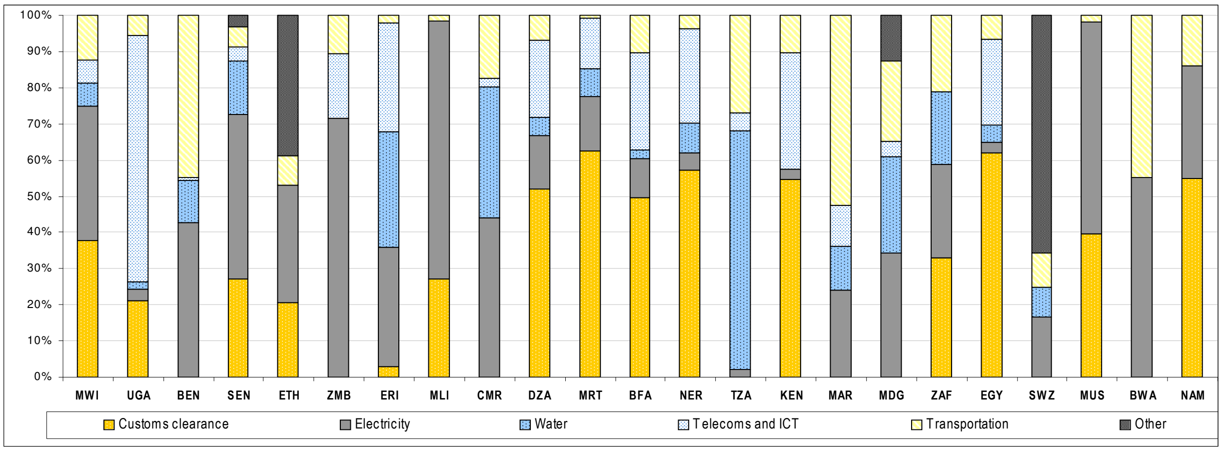
B. Percentage absolute contribution of infrastructure to allocative efficiency in logs by key factors



Source: Author's calculations with IC data.

Note: For a description of the variables contained in each group see footnote in Figure 6.24

Figure 6.28 Infrastructure's impact on allocative efficiency in logs by key factors (II)



Source: Author's calculations with IC data.

Note: For a description of the variables contained in each group see footnote in Figure 6.24

CHAPTER III

EMPIRICAL ECONOMETRIC EVALUATION OF ALTERNATIVE METHODS OF DEALING WITH MISSING VALUES IN INVESTMENT CLIMATE SURVEYS

Empirical Econometric Evaluation of Alternative Methods of Dealing with Missing Values in Investment Climate Surveys*

by

Alvaro Escribano† and Jorge Pena‡

May, 2009

Abstract

The Investment Climate surveys (ICSs) are valuable instruments improving our understanding of the economic, social, political and institutional factors determining economic growth, particularly in emerging and transition economies. However at the same time they have to overcome some difficult issues related with the quality of the information provided; measurement errors, outlier observations and missing data are frequently found in this datasets. In this paper we discuss the applicability of recent procedures to deal with missing observations. In particular we present a simple replacement mechanism—for application in models with a large number of explanatory variables—, which we call *ICA method*, which in turn is a proxy of two methods: multiple imputation and EM algorithm. We evaluate the performance of this ICA method in the context of TFP estimation in extended production functions using ICSs from four countries: India, South Africa, Tanzania and Turkey. We find that *ICA method* is very robust and perform reasonably well even under different assumptions on the nature of the mechanism generating missing data.

Keywords: Investment Climate surveys, missing observations, incomplete data, random sampling, sample selection, EM-algorithm and bootstrap.

* We have benefited from suggestions of Daniel Peña, Ariel Pakes, Rodolfo Stucchi and Eric Verhoogen.

† Telefonica-UC3M Chair of Economics of Telecommunications, Department of Economics, Universidad Carlos III de Madrid; alvaroe@eco.uc3m.es.

‡ Department of Economics, Universidad Carlos III de Madrid; jpizquie@eco.uc3m.es.

1. Introduction

The Investment Climate (IC) surveys (or Enterprise Surveys) have been created as part of a new strategy of the World Bank on putting more emphasis on intangible assets of developing countries such as knowledge, institutions and culture.[1] This new set of information that is becoming available for both scholars and policy makers is intended to be a valuable instrument to help improving our understanding of the economic, social, political and institutional factors determining economic growth, particularly in emerging and transition economies. However, at other level IC surveys are also a source of troubles for researchers. In general, economic data are far of being perfect and when one is doing econometric or statistical analysis with a typical dataset too often have to deal with the problem of missing values.[2] IC datasets are not an exception in this issue. Their imperfections make our job difficult and often even impossible (Griliches, 1986).

Incomplete data is a ubiquitous problem that standard econometric and statistical methods have nothing whatsoever to say about it or how to solve it. The simplest solution to this problem is to exclude from the analysis any cross-sectional observation with any missing value in it. This strategy is commonly known as *casewise deletion, listwise deletion* or *complete case analysis*. The advantage of this method lies obviously in its simplicity. The disadvantage is also rather evident to anyone who has used it: in many applications, *casewise deletion* excludes from the analysis a large fraction of the original sample. In the context of IC surveys this is quite a large cost in terms of information lost, apart from the monetary cost coming from losing a large proportion of, in the other hand, very expensive interviews.

The debate we want to introduce is whether the researcher should apply some treatment on missing values when using investment climate surveys (ICSs) or rather it is preferable to operate with the complete case only. One of the main characteristics of the ICSs is the wide set of information they provide. Concretely, the surveys have been designed to perform a variety of economic and statistical analyses, among which especially interesting are those linking investment climate variables and several measures of firms' economic performance such as productivity, labor demand, sales, exporting activity, FDI propensity, etc. This means having matrices of data with a remarkably large number of rows and therefore the possibility of using econometric models with a wide set of right hand side variables. Unfortunately, in many cases

[1] Key determinants of the investment climate, which are included and properly measured in the Investment Climate (IC) series of surveys, include physical and institutional infrastructure, economic and political stability, rule of law, infrastructure, approaches to regulations and taxes, functioning of labor and finance markets, and broader features of governance, such as corruption. The World Bank group has long been a supporter of investment climate reform, recognizing the importance of shaping a business environment conducive to the successful start-up and operation of firms of all sizes in all sectors.

[2] Information is missing for some reasons. A sizeable fraction of the respondents refuse, forget or overlook to answer some questions. In other cases, even well trained interviewers may neglect to ask some questions. Sometimes respondents just say they do not have the information available to them or they do not know the question. Some questions are simply not applicable to some respondents (see Allison, 2001). All of these cases may be applicable to IC data.

the problem of missing data is so serious that prevent us to use those kinds of models. In some of these cases the missingness problem reduces the cross sectional observations available in the complete case to even 0% of the original sampling frame.[3] Should the researcher therefore constraint himself to use models with a reduced number of independent variables under the risk of introducing a more serious omitted variables problem? Or is it preferable to impute missing data in order to be able to use structural models with a wide set of explanatory variables? If we assume the later as a reasonable solution, the question that arises then is: should we input missing cells in both LHS (independent) and RHS (dependent) variables, or as opposite should we satisfy ourselves by replacing missing data in only those explanatory variables of the model?

Over the last years statisticians have proposed many alternative methods to handle incomplete datasets that offer substantial improvements over *casewise deletion*. These approaches may be grouped into two families of methods, maximum likelihood and multiple imputation, see Allison (2001), Meng (2000) and Little and Rubin (1987) for a review. However, these methods depend on easily violated assumptions that, to make things worse, are difficult or even impossible to test. In this paper we discuss the applicability of these methods to four IC surveys with very different patterns of missing data among them: India, Turkey, South Africa and Tanzania. In particular, we propose a simple imputation mechanism (which we call *ICA method*) that in part departs from the EM-algorithm, and that has been widely applied in various empirical works (Escribano et al, 2008a, b; Escribano, Guasch and Pena 2009 and Escribano et al. 2009). We compare the performance of this method with several alternative approaches to deal with incomplete data and we discuss the different assumptions we need to hold for the different imputation mechanisms to work well. We evaluate the validity of the different methods in the context of the extended production function of Escribano and Guasch (2005 and 2008).[4] The extended production function framework used here fits very well with the objective of the paper as the RHS of the equation is compounded by a broad set of explanatory variables.[5] On the other hand, although we concentrate in PF variables, the results of the analysis can be easily extended to any variable with missing information included in the ICSs.

We demonstrate that, besides the imputation method used, a careful knowledge of the missingness mechanism in the context of ICSs is a requisite. The missing data problem is at the core of statistical and econometric analysis done with ICSs and therefore a proper treatment of the missing data mechanism is inevitably. We also show that the so called *ICA method* proposed performs reasonably well, even under very different patterns of missing data. The differences of

[3] The number of observations available in the complete case decreases as we consider more and more investment climate variables. If we consider all the variables included in the survey the complete case due to missing cells is 0% in most of the cases. However, If we construct models using only those investment climate variables with a response rate higher than 80% the complete case increases to 20% to 30% of the original sampling frame.

[4] Although it is straightforward to apply this method to any kind of model, especially those involving a large number of RHS variables or structural system of equations.

[5] The underlying philosophy of the Escribano and Guasch (2005 and 2008) extended production function is to incorporate in a Cobb-Douglas (or Translog) function a large set of investment climate variables to correct for observable fixed effects.

the *ICA method* with respect to other more sophisticated imputation mechanisms, such as EM algorithms, multiple imputation, bootstrap methods or Heckman models, are not remarkably significant, so we propose it as a benchmark, homogeneous, simple and easy to implement method for models with large numbers of covariates in ICSs and more importantly for very complex and unbalanced patterns of missing data.

The structure of the paper is as follows. In section 2 we review the patterns of missing data observed in the four IC surveys considered. We compare the original sampling frame with the complete case and we see that in most cases the representativity of the original sample is modified and the total number of observations available for regression analysis is considerably reduced. We compare these numbers with the observations available after the replacement mechanism we propose. Section 3 presents the *ICA method* and other imputation mechanism used as comparators. We also comment the different assumptions underlying the different methods proposed. We discuss to what extent the missing data mechanism (MDM) presented in the four surveys analyzed may be considered as missing completely at random (MCAR), missing at random (MAR), or non-ignorable. Section 4 shows the regression results for the extended production function under the different replacement methods. Finally, section 5 concludes. All tables and figures are included in a large appendix at the end of the paper.

2. Missing data and investment climate surveys

We introduce the problem at hand with Table 1.1 (see appendix on tables and figures) which shows the total number of observations, the observations available in the complete case and the final number of observations we have after the replacement process we propose—which we discus later on—in 43 different ICSs. All the surveys shares similar characteristics in the sampling procedure applied and, more importantly, in the information provided. The number of observation lost varies among all the surveys considered. The replacement process considerably increases the sample size in all the cases (the method is described in section 3).[6] The problem of incomplete data is common to all the IC surveys considered, although it is more persistent in countries like Thailand, Niger, Paraguay, Tanzania and Turkey, in which the percentage of observations available in the complete case is below 30%.[7] In Table 1.1 we only consider missing values in production function variables. When we consider all variables likely to be used in regression analysis (all investment climate variables), the complete case reduces to even 0% in some cases.[8]

---

[6] The sample with replacement fills missing values of all variables of the survey (both production function and IC variables).

[7] By means of simplification we understand by *complete case* the sample with replacement only in IC variables.

[8] As said, the problem of missing data is, in lower or greater extent, common to almost all the variables presented in the IC surveys. We here consider the missingness and its treatment in production function variables (sales, materials, capital and employment), although all we say about imputing missing information in production function variables can be easily extended to any other IC variable.

[TABLE 1.1 ABOUT HERE]

We focus the analysis on the investment climate surveys of India, Turkey, South Africa and Tanzania because they represent almost all the situations regarding the structure of missing data we may found.[9] For India in the complete case we lose 35% of the original sampling frame, while after replacing we only lose 16%. Turkey and Tanzania lose a similar percentage of observations, 70.9% and 60.7% respectively. South Africa only loses 29.2%.

Table 1.2 goes in depth in the description of the missingness problem of the four countries selected. In this case for the computation of the observations available in the complete case we use all those IC variables included in the survey likely to be used in a regression analysis framework. This means using more than 115 variables in India, 90 in Turkey, 168 in South Africa and 162 in Tanzania. For each country we consider two benchmark cases: the first one includes both PF and IC variables in the computation of the complete case, while the second only considers the IC variables. In the extreme case, when we consider all those IC variables the complete case reduces to 0% of the complete case in all the countries, it doesn't matter whether we include PF or not. Note that the observations available in the complete case increases as we exclude from the computation of the complete case those IC variables with the largest proportion of empty cells reported. In order to have a large enough number of observations we would need to exclude from the analysis those IC variables with a response rate lower than 95%. Even in this case, and considering also the PF variables, we should be forced to exclude 41.1% of the interviews in India, 76.9% in Turkey, 60.2% in South Africa and 66.2% in Tanzania. The evidence is overwhelming on the importance of the problem of missing information we have to deal with.

[TABLE 1.2 ABOUT HERE]

In the remaining of this section we first present the pattern of missing values observed in the four surveys considered. We also evaluate the representativity of the sample with replacement and the complete case with respect to the sampling frame.

2.1 Sampling and characteristics of the ICSs

The sampling of the ICSs is based on a World Bank template used in a large number of countries and customized in collaboration with regional statistical agencies to reflect country-specific issues and policy areas of interest. In order to ensure proper representation of the sectors of

[9] These datasets have been in turn analyzed in the following works Escribano, Guasch and de Orte (2009) for India, Escribano, Guasch, de Orte and Pena (2008b and c) for the case of Turkey and Escribano, Guasch and Pena (2009) for South Africa and Tanzania.

interest,[10] respondents are carefully selected. The sampling process is normally based on national industry databases and census of firms or establishments,[11] which provides the necessary information on the particular population of establishments. To ensure proper representation of firms, stratification is usually done based on three standards: size, sector and location.[12]

The information contained in the ICSs is composed by a wide set of around 400 variables. Eventually, the number of variables likely to be used in regression analysis reduces to a number around 120-200.[13] The Investment Climate Surveys capture firms' experience in a range of areas related with the economic performance: financing, governance, corruption, crime, regulation, tax policy, labor relations, conflict resolution, infrastructures, supplies and marketing, quality, technology, and training among others. The ICSs also provide information on the productivity (or production function) variables, says output (sales are used as measure of output), employment, intermediate materials, capital stock and labor cost. The resulting panel information is short in the time dimension, since includes only 2 or 3 years of productivity data (in our case 2 years for Turkey and 3 for India, South Africa and Tanzania), and has 1 year of information for the investment climate variables. Finally, it is important to note that all information is based on recall data and not in book values or accountings.

2.2 The missing information problem at first glance

Figures 1.1 to 1.4 show the complex and unbalanced patterns of missing values observed in the PF variables in the four countries considered. The most common case is to find observations with information for all the PF but one. In India, the percentage of establishments reporting information for all the PF variables but capital is 16.3%. In the rest of the countries this percentage is slightly lower but significantly high too. It is less common to observe data on all the PF variables but sales, materials or employment. Although in Tanzania the percentage of firms reporting all the figures but sales is relatively important, 9.8%. The cases for which data is collected for only two PF variables represent, in all the countries, less than 1% of total data. Finally, it is very common to have data collected only for labor, this percentage represents 13.3% in India, 27.9% in Turkey, 5.5% in South Africa and 15.7% in Tanzania.

[FIGURES 1.1 TO 1.4 ABOUT HERE]

---

[10] Here we focus only on the manufacturing sector. By classifying the establishments by their ISIC code we generally end up with establishments from the next eight sectors: a) Food and beverages; b) Textiles and apparels; c) Chemicals; d) Non-metallic mineral products; e) Metallic products; f) Machinery and equipment; g) Electrical machinery; h) Transport equipment.

[11] The unit of reference in the ICSs is the establishment, although in this paper we refer indistinctively to both establishments and firms.

[12] Concretely, the establishments are selected according to a random sampling by industry and region. Taking into account this issue we use standard errors allowing for clustering by industry and region (apart from the conventional correction for heteroskedasticity *a la* White). In some surveys there is also oversampling of large firms.

[13] We understand by *"likely to be used in regression analysis"* all those variables describing the investment climate in which firms operates and likely to be related with firms economic performance.

Tables from 2.1 to 2.4 of the appendix show the distribution of the number of observations available in the original sampling frame, in the complete case and in the sample with replacement, along with the percentage of observations lost with respect to the original sampling frame. From Table 2.1 the percentage of observations lost in India in the complete case varies when we move industry by industry and size by size. Flagrant cases of lost of observations are small firms operating in the non-metallic products sector (61.9%) or the medium firms of the food sector (55.37%). The replacement process allows retrieving for the analysis a considerable percentage of observations. After the replacement we only lost 28.6% and 22.6% in the two cells mentioned previously. In Turkey the percentages of observations lost by size and industry (see Table 2.2) ranges from 40% (medium firms in the transport equipment sector) to 87.3% (small firms in textiles and apparels industry). South Africa lost 50% of small firms in textiles and apparels and chemical, rubber and plastics sectors (see Table 2.3). Lastly, Tanzania lost more than 70% of small firms in paper, edition and publishing and machinery and equipment and 73% of large firms in textiles and apparels (Table 2.4).

[TABLES 2.1 TO 2.4 ABOUT HERE]

Tables 3.1, 3.2, 3.3 and 3.4 try to illustrate how the representativity of the sampling frame changes with respect to the complete case and the sample with replacement.[14] In all the cases the percentages slightly varies in the complete case with respect to the sampling frame. The percentages of the sample with replacement are more similar to the sampling frame. For instance, in India from Table 3.1, panel a), the percentage of '*food*' firms falls from 8.7% to 6.9%, while after the replacement it is 7.9%. Symmetrically, the percentage of '*apparel*' firms jumps from 12% to 14.3% in the complete case and to 12.4% in the sample with replacement. Similar patterns can be observed in the remaining countries. Finally, from these tables response rates do differ across countries, but within countries they are remarkably uniform across regions and industries.

[TABLES 3.1 TO 3.4 ABOUT HERE]

[14] In order to evaluate how representativity changes from the sampling frame to the complete case we would need to have information on the weight of each category over the reference population. Unfortunately, this information is not available. As a second best we can still demonstrate how representativity changes from the data we have. Let us suppose population is split into two strata, and that the original sample selects a given number of observations for strata 1 and 2, as a result X and Y are the percentages that represent the weight of each strata in the population. In the complete case we introduce the missing data problem so instead of X and Y we have X´, Y´. If we suppose that the sampling frame is representative of the population then the complete case is said to be representative if and only if the weights in the complete case are proportional to the weights in the sampling frame; that is X≈X´ and Y≈Y´.

3. Imputation of missing values: *ICA method*

Rubin (1976) rigorously defined the assumptions that might plausibly be made about missing data mechanisms (MDM).[15] When the MDM is ignorable, the objective of the replacement methods is not to augment the sample size, but to preserve the sample representativity, to gain efficiency in the estimation and to retrieve for the analysis a large number of very expensive interviews. The alternative to these methods is the *listwise deletion*, which is not a panacea even when the MDM is ignorable. Operating with the complete case is only acceptable if incomplete cases attributable to missing data comprise a small percentage, say 5% or less, of the number of total cases (Schafer, 1997), and when the complete case preserves the representativeness of the original sampling frame. In addition, in models with a large number of regressors missing data problem may encourage analysts to leave out of the regression some explanatory variables with high proportion of missing values. As Cameron and Trivedi (2005) point out, this practice may be misleading as it leads to an omitted variables problem, which is more serious than the missing data problem *per se*.

To see how the various mechanisms applied to deal with missing data perform it is useful to depart from a population model of interest. A repeated task that applied researchers do in the context of IC data is the estimation of production functions to perform a variety of productivity analyses. Concretely, let us suppose the extended production function as in Escribano and Guasch (2005 and 2008). The population model is given by

$$\log Y_{it} = \alpha_0 + \alpha_L \log L_{it} + \alpha_M \log M_{it} + \alpha_K \log K_{it} + \alpha'_{IC} IC_i + \alpha'_D D_{it} + u_{it}, \qquad (1)$$

where *logY, logL, logM* and *logK* represents output, labor, materials and capital all in logs, *IC* is the time-invariant vector of investment climate and other control variables and *D* is a vector of industry/region/size/time dummies. Since the usual time, industry, region and size fixed effects are included in the vector D, and the usual fixed effects are assumed to be observable and included in IC vector, u is assumed to be an usual *i.i.d* error.[16]

Equation (1) is of special interests for the purpose of this paper as it implies using a large proportion of the variables included in the ICSs. Furthermore, it is especially useful to illustrate the trade-off between plausible biases inherent to measurement errors that could arise after replacing missing data and the omitted variables bias associated to the complete case. Concretely, in the four cases considered the final vector of significant IC variables is intended to

[15] Data on Y variable is said to be missing completely at random (*MCAR*) if *P(Y missing | Y, X)= P(Y missing)*, where X is a matrix of other variables on data. Data is missing at random (*MAR*) if *P(Y missing | Y, X)= P(Y missing | X)*. Missing data is nonignorable if *P(Y missing | Y, X)= P(Y missing | X, Y)*.

[16] Concretely, equation (1) is based on the methodology proposed in Escribano and Guasch (2005 and 2008) with further developments in Escribano et al (2008a and b). The selection of variables is detailed in those papers, and it is based on a general to particular procedure. Although for the purpose of this paper we are not interested in the properties of the model, but we want to test the sensitivity of the results to the imputation method used, it is interesting to clarify that the underlying philosophy of this methodology is to use the time-invariant vectors of IC variables to correct for observable fixed effects.

include 27 variables in India, 18 in Turkey, 31 in South Africa and 25 in Tanzania.[17] The definition of the variables used, classified in five broad groups (infrastructures, red tape, finance, quality, and other), is in the appendix on definition of variables.

For identification in (1) if we observe all data and under regularity conditions it is clear that, following Wooldridge (2007), we need $E(u_{it} \mid \log L_{it}, \log M_{it}, \log K_{it}, IC_i, D_{it}) = 0$. Now let the pattern of missing values for each observation *i* at moment *t* be given by s_{it}, where s_{it}=0 if missing value and 1 otherwise. So what we observe is

$$s_{it} \log Y_{it} = s_{it}(\alpha_0 + \alpha_L \log L_{it} + \alpha_M \log M_{it} + \alpha_K \log K_{it} + \alpha'_{IC} IC_i + \alpha'_D D_{it}) + s_{it} u_{it} . \qquad (2)$$

If the pattern of missing values is *M.A.R* or *M.C.A.R* then the necessary conditions for equation (4) to be identified are $E(s_{it}u_{it}) = 0$, $E[(s_{it}J)(s_{it}u_{it})] = E[(s_{it}Ju_{it})] = 0$ with $J = \log L_{it}, \log M_{it}, \log K_{it}, IC_i, D_{it}$. In the additional case of *exogenous sample selection*, when the pattern of missing values is determined only by the explanatory variables of (1),—for instance the missing values have some patterns on time, size, industries, regions or even between exporters/non-Exporters firms, domestic/foreign, etc—we also need that

$$E(s_{it}u_{it} \mid s_{it} \log L_{it}, s_{it} \log M_{it}, s_{it} \log K_{it}, s_{it} IC_i, s_{it} D_{it}) = s_{it} E(u_i \mid s_{it} \log L_{it}, s_{it} \log M_{it}, s_{it} \log K_{it}, s_{it} IC_i, s_{it} D_{it}) = 0 .$$

That is, for the identification condition in this case to hold we need to control for any exogenous variable affecting the pattern of missing values, and this is the way we proceed in the estimation of the productivity equations. Note that once we have controlled for all these variables, we can estimate (2) in the *complete case* consistently, although at the cost of losing efficiency and in some cases the representativity of the original sampling frame.

When the pattern of missing values *s* is correlated with the dependent variable of (1) we are in the presence of self-selection case.[18] In this case the missing values are not ignorable and we cannot get rid of incomplete observations. In this case equation (2) must be estimated by other sample selection corrections, like the Heckman selection model.

In what follows we discuss the first imputation mechanism proposed to deal with the problem of incomplete data; *the ICA method*.

3.1 Imputation of missing values: ICA method

Our method of imputing missing data, which we call *ICA method*, shares the expectation step of the Expectation-Maximization (EM) algorithm proposed in the seminal paper of Dempster, Laird

---

[17] Although the initial set of IC vectors comprises more than 150 variables, a reduction process from the general to the specific were applied in order to find the final sets of significant variables. The final set of variables is required to be robust to 12 different TFP measures. More details are in Escribano and Guasch (2005 and 2008).

[18] Notice that as equation (1) is equivalent to: $\log Y_{it} - \alpha_L \log L_{it} - \alpha_M \log M_{it} - \alpha_K \log K_{it} = \alpha_0 + \alpha'_{IC} IC_i + \alpha'_D D_{it} + u_{it}$, where in the right hand side we have the productivity index. We are clearly concerned with the possible correlation of the MDM with productivity or TFP as it may induce biases in the estimators of the vector *β*.

and Rubin (1977), method that, within the maximum likelihood approaches, has been widely applied in several scientific fields (see McLachlan and Krishnan (1997) for a review). In particular, the replacement strategy used departs from the expectation of the production function variables conditional on the industry, region and size the corresponding observation belongs to ('*expectation step*'). Or equivalently, we replace the missing value by the expectation of the distribution of the variable conditional on the information on sector, region and size according to next equation

$$E(J_{it} \mid D_{R,it}, D_{I,it}, D_{S,it}) = \rho_0 + \rho'_{R,J} D_{R,it} + \rho'_{I,J} D_{I,it} + \rho'_{S,J} D_{S,it} \qquad J = Y, L, M, K \qquad (3)$$

where Y, L, M and K represents output, labor, materials and capital and D_R, D_I and D_S are vectors of region, industry and size dummies respectively. Notice that we choose (3) such that it represents the special features of the IC datasets—in IC surveys industry, region and size are the variables used to stratify the sample.

After excluding from the replacement process those observations with all the production function variables missing, [19] estimated values to replace incomplete data are given by

$$\tilde{J}_{it} = \hat{\rho}_0 + \hat{\rho}_{R,J} D_{R,it} + \hat{\rho}_{T,J} D_{I,it} + \hat{\rho}_{T,J} D_{S,it} \qquad J = Y, L, M, K \qquad (4)$$

Unlike EM algorithm,[20] ICA method has the advantage of separating the imputation of missing data from the estimation of the parameters of the population model. More precisely, separating the imputation mechanism of a population model is the main characteristic of the *multiple imputation* approaches, what allows using them with virtually any kind of data and any kind of model. *ICA method* is, in fact, a general multiple imputation mechanism in which we assume that each imputed variable can be represented as a linear function of the variables used to stratify the sample (dummies of industry, region and size) and therefore the fitted values can be used to replace missing data.

Hence, the first assumption we need is that the imputed variable can be represented as a multiple linear function of other variables. The second condition that needs to be held for multiple imputation to work well is that all the variables, including those replaced and those used to replace, have normal distributions (see Allison, 2001).[21]

[19] ICA method is conservative in the sense that we do not replace missing cells for those observations with all but one PF variables unobserved. We force the industry-region-size cells to have at least 18 values to estimate consistently the sample average. Moreover, in order to avoid biases caused by outlier observations we use the within-group median instead of the within-group mean.

[20] The EM algorithm imputes missing data conditional on a given population model, and therefore chooses the candidates values to replace the missing cells that maximizes the likelihood function conditional on a vector of parameters of that model.

[21] Although these are strong assumptions the imputation method seems to works well even when the variables have distributions that are manifestly not normal, see Schafer (1997).

According to equation (3) and (4), equation (2) represents the '*maximization step*', which is now given by

$$s_{it}^{*}\log\tilde{Y}_{it}=s_{it}^{*}(\alpha_0+\alpha_L\log\tilde{L}_{it}+\alpha_M\log\tilde{M}_{it}+\alpha_K\log\tilde{K}_{it}+\alpha_{IC}'IC_i+\gamma' D_{it})+s_{it}^{*}\tilde{u}_{it} \qquad (5)$$

where *y, l, m* and *k* with tilde on top represent the imputed variables and s^* is the new pattern of missing values after the replacement process.[22] With identification conditions in the MAR case given by $E(s_{it}^{*}\tilde{u}_{it})=0$, $E[(s_{it}^{*}J)(s_{it}^{*}\tilde{u}_{it})]=E[(s_{it}^{*}J\tilde{u}_{it})]=0$ with $J=\tilde{l}_{it},\tilde{m}_{it},\tilde{k}_{it},IC_i,D_{it}$, while in the case of *exogenous sample selection* we need that

$$E(s_{it}^{*}\tilde{u}_{it}\mid s_{it}^{*}\log\tilde{L}_{it},s_{it}^{*}\log\tilde{M}_{it},s_{it}^{*}\log\tilde{K}_{it},s_{it}^{*}IC_i,s_{it}^{*}D_{it})=s_{it}^{*}E(\tilde{u}_{it}\mid s_{it}^{*}\log\tilde{L}_{it},s_{it}^{*}\log\tilde{M}_{it},s_{it}^{*}\log\tilde{K}_{it},s_{it}^{*}IC_i,s_{it}^{*}D_{it})=0.$$

That is, we need to control for any explanatory variable correlated with s^* to get consistency either in the inputs or IC variables.

When the two assumptions mentioned above (normality and linearity of imputed variables on dummies of industry, region and size) do not hold the replacement strategy is no longer consistent. Very few can be said about the asymptotic distributions of the estimators obtained under these circumstances, because they have not been derived yet. In a general fashion, in these cases we can understand our replaced variables as the *classical* problem of *variables measured with error*. In order to illustrate this let our model be given by $y_i=x_i\beta+u_i$, where y_i represents sales and x_i is a vector of inputs. Suppose that in the population we have that $E(u_i\mid x_i)=0$, and that x_i is missing when $i\in S$. When we predict x_i $i\in S$ such that $\hat{x}_i=x_i+v_i$ where $\hat{x}_i$ is our predicted value, then the model becomes $y_i=\tilde{x}_i\beta+\tilde{v}_i\beta+u_i$. Where when $i\notin S$ $\tilde{x}_i=x_i$ and $v_i=0$, while if $i\in S$ $\tilde{x}_i=\hat{x}_i$ and $\tilde{v}_i=x_i-\hat{x}_i$. Therefore, consistency of estimates of β depends on whether $E(\tilde{v}_i\mid\tilde{x}_i)=0$. Consistency follows if the linear regression of the inputs on industry, region and size variables gives us a noisy measure of the true level of the variables. Otherwise we will have a v_i and the parameters obtained from regression analysis would be consequently downward biased, and the magnitude of the bias will depend on the standard deviation of the error term relative to the standard deviation of the variable and the proportion of replaced values.[23]

3.2 Performance of ICA method

The performance of the ICA method is illustrated by plotting the Kernel densities of the PF variables in the complete case and after imputing missing data. Those are in figures 2.1 to 2.4 in the appendix at the end of the paper. Overall, from these figures the distributions of ICA method

[22] Variables included in the IC and C vectors are imputed by using the same procedure. However, by means of illustration and simplification here we only discuss the identification condition as if only PF variables were imputed.
[23] We thank Ariel Pakes for useful suggestions at this point.

and the complete case tend to be similar when the proportion of missing values is not too high. Divergences appear as the proportion of unobserved sample becomes larger.

[FIGURES 2.1 TO 2.4 ABOUT HERE]

From a more detailed analysis of Figure 2.1, which illustrates the Indian case, it is clear that there are not significant differences in the distributions of any of the PF variables in the complete case and in the sample with replacement by the ICA method, what is supported by the Kolmogorov-Smirnov tests. Furthermore, both the sample mean and the standard deviation do not change significantly before and after the imputation process (especially important is the fact that the standard deviation does not declines after the imputation). These observations hold for all the PF variables, even for the case of the capital stock, for which the proportion of imputed values is much higher than in the remaining variables. South African case represented in Figure 2.3 preserves the same conclusions of the Indian sample.

In the other hand, the performance of the ICA method in the cases of Turkey and Tanzania show significant different behaviors than in the previous cases. Thus, in Turkey where the response rate of PF variables is below 40%, the kernel estimates suggest slight differences in the shape of the distributions, and although the sample means are rather similar the standard deviation estimated after imputing missing values decreases as the proportion of missing values increases. The same holds for the case of Tanzania, although in this case the problem becomes more acute as the sample distributions are far of being normal, rejecting the null hypothesis of the Kolmogorov-Smirnov tests.

The extent to what the ICA method give us a good approximation of the population distribution of the variables and therefore leading to a consistent estimation of equation (1) depends on the determinants of the MDM. Studying and analyzing the characteristics of the MDM is precisely the aim of sections 4 and 5, where we investigate the links between the patterns of missing values and productivity, sales and other key characteristics at the firm level like accountability, informality, corruption, crime, innovative activity, etc. This analysis will be significantly important in the remaining sections when we compare the ICA method with extensions and other different imputation mechanisms, which rely in different assumptions on the nature of the missingness mechanism.

4. Nature of missing data mechanism

The objective in what follows is to present a careful descriptive analysis of the characteristics of those firms having missing values. The aim of this exercise is to form a judgment about whether the missing data mechanism may be treated as missing at random or not.

4.1 Why do some establishments refuse or avoid providing some information?

At this point, one question of great concern is the nature of the generating data process: missing completely at random, missing at random or non-ignorable missing data. Different assumptions can be made about the nature of the mechanism generating missing values. In general, missing values may be considered as a consequence of some of the following causes: a) firms refuse to answer some questions (they do not have the information with them, they simply don't know the information, they don't want to report it, they forget to answer some questions, etc); b) the interviewer neglects to ask some questions; and c) the question does not apply to some firms.

Since missing data coming from an oversight of the interviewer or because the question simply does not apply represent a small share of the total number of missing values and may be assumed as random, we are clearly concerned with the case in which firms avoid, refuse or simply don't answer some questions. Here one can use some assumptions on why firms do not report some figures to the interviewer. Maybe, firms do not report data on production function variables because of lack of accountability. It could also be a matter of informality, those firms that do not report all sales to IRS authorities may have an incentive to avoid reporting these figures to the data collector as well, even though data is confidential. In this vein, one may also consider that missing values could be correlated with the level of corruption of the environment in which firms operate.

Productivity or the level of sales could also explain missing values, the more sales (or productivity) the less missing values. The explanation could be simply that weaker/less profitable firms do not keep a proper accountability, or maybe because weaker firms' managers are less likely to know PF figures (it is important to point out here that PF variables comes from *recall* data). At this point, the question is whether the pattern of missing values is directly correlated with sales or TFP or it is correlated indirectly through other variables such as share of exports, imports, access to infrastructures, capacity, innovation, R&D, quality, use of IC technologies, informality, corruption, accountability, etc, which are known to be strongly associated with sales and TFP.[24]

If the pattern of missing values is directly correlated with the dependent variable of our model—sales or TFP in our case—then the MAR or MCAR assumptions no longer hold. In this case, the missing value mechanism is said to be non-ignorable and the missing data mechanism needs to be modeled together with the structural model we are willing to estimate. In the other hand, when the missing data mechanism is related with sales or TFP indirectly through other—independent or exogenous—variables in the dataset, the missing data mechanism is considered to be missing at random, which under regularity conditions is equivalent to say that missing data is

[24] Notice that we are concerned with the correlation of the MDM with either sales or TFP. We use the extended production function of equation (1) where a wide set of IC and C variables is plugged into a general PF in order to control for observable fixed effects. The correlation of MDM with sales may introduce bias in the input-output elasticities estimates, whereas the correlation with TFP could imply biased IC parameters estimates.

ignorable.[25] In this case we can get rid of missing data and operate only with the complete case once we have controlled for the variables correlated with the missingness mechanism. However, some caveats need to be made over *casewise deletion* as we will see in later sections.

The descriptive analysis we propose in this section allows us to have a deeper and thorough knowledge of the MDM. This is especially useful when the MDM is non-ignorable (not MAR and therefore not MCAR). As Meng (2000) signals, ignorability is untestable from the observed data, so caution is required when drawing conclusions from models with imputed data. Further, sensitivity analysis and subjective knowledge of the nature of the MDM plays a critical role in this point, as Molerberghs et al. (1999) illustrate. In fact, modeling the MDM is a very active line of research with a number of unresolved problems (see e.g. Heitjan, 1994 and 1999; Ibrahim, et al., 1999). From now on, the aim is therefore to describe the characteristics of those firms reporting missing values. The types of questions we are willing to address are: has the missingness mechanism some relevant information about the parameters we are willing to estimate? Or in other words, are the parameters of the MDM related with the parameters of our model? And as a consequence, is the MDM ignorable?

4.2 Is it more likely to find a missing value within small firms?

Firstly, we are concerned with the possibility of systematic bias in the response rates to questions on sales and inputs. Table 4 shows the number of missing values in sales and inputs by sizes, which are known to correlate strongly with productivity (and also with sales).[26] The pattern in response rates is that small firms (those with less than twenty employees) tend to respond less often in India and South Africa. The pattern is somewhat different in Turkey and Tanzania where missing values in the inputs are uniformly distributed across categories of firms' sizes, with the exception of capital stock which has more proportion of missing values within small firms. At this point, these results could suggest the presence of some degree of systematic bias of the response rates in India and South Africa. Nonetheless, further investigation is needed to place additional insight on this question. The fact that small firms report less information also suggests that response rates to detailed sales and costs questions could have more to do with accounting and capacity—less affordable for small firms.

[TABLE 4 ABOUT HERE]

---

[25] A separate question is whether MAR is equivalent to ignorable missing data. Even when the missing data mechanism is assumed to be MAR, an additional assumption is needed to ensure that empty cells can be ignored: the parameters of the missing data process need to be unrelated with the parameters of the model we are willing to estimate. However, MAR and ignorability are almost always considered as equivalent assumptions in the literature, since the assumption that the parameters defining the missingness model are unrelated with the structural model is easily satisfied (see Allison, 2001 and Heitjan and Basu, 1996 for illustrations).

[26] Categories of size are: small, less than 20 employees; medium, in between 20 and 100 employees; large, more than 100 employees.

4.3 Are missing values distributed uniformly across different categories of firms?

Tables 5.1 to 5.4 offer further empirical underpinning on whether the MDM is related with firms' weakness, or rather are other firms' attributes what determines the probability of observing a missing value. Table 5.1 focuses on the case of India. It compares the share of firms reporting at least one missing value on PF variables in the whole sample, with the share of firms reporting missing values by categories of key IC variables. In the case of India 32.8% of firms report at least one missing value in PF variables. This percentage varies when we take into account categories of IC variables. Thus, those firms that do not use e-mail or suffer power outages tend to respond less often to PF questions, respectively 39.0% and 37.8% of firms with missing information within these two categories. It is indicative of the nature of the MDM that those firms hiding some share of sales and/or workforce to IRS tax authorities have more missing values in PF variables on average (see the rows corresponding to Informality (I) and Informality (II)). With regard to corruption those firms that operate in a more corrupt environment report less missing values. Similar conclusion can be obtained from crime; those firms having suffered criminal attempts also tend to avoid reporting PF figures.

Symptomatic of the nature of the MDM in India is the fact that firms with access to a credit line and with the annual statements reviewed by a external auditor report a lower proportion of missing values (PF information is lost for 40.4% of firms without access to a credit and 50.2% of firms with the annual statements not engaged in external audit report at least one missing value). This indicates that a plausible explanation of the missing values is the lack of proper accountability or even informality.

Continuing with Table 5.1 other indicative variables of the pattern of missing values are the exporting activity (only 18.2% of those firms exporting directly report any missing value) and the education of the manager (28.5% of firms with a manager with university education of manager report missing values, while 35.1% of the remaining firms report missing values). These two variables indicate that the level of competitiveness of the firm is another important factor explaining the pattern of missing values. However, other variables that are known to correlate strongly with competitiveness and productivity such as FDI or the introduction of new technologies and products does not provide any further information on the MDM.

[TABLES 5.1 TO 5.4 ABOUT HERE]

The case of Turkey is represented in Tables 5.2. The patterns are similar to those observed in India. The power outages suffered, e-mail usage, informalities and corruption are good indicators of the pattern of missing values. Again the proportion of missing values within firms having access to credit line and to an external auditory is larger relative to those that do not, what comes to corroborate the story of the accountability as a determinant of the MDM.

Other variables with important implications for the MDM are the exports, the FDI, the introduction of new technologies, the legal organization of the firm (incorporated company or not) and the percentage of capacity utilization.

Similar conclusions can be obtained for South Africa in Table 5.3. Missingness in this country appears to be associated with water outages, use of e-mail, informality and corruption, accountability, and the legal status, and in a lower extent with power outages, security expenses and introduction of new products and technologies.

These patterns are even more intense in Tanzania. Table 5.4 illustrates that, for instance, within those firms with access to a loan 39% report missing values, while within those firms without loan the percentage jumps to 48.2%. The same holds for informality, corruption, quality, technology, exporting activity, legal status, holdings or capacity utilization.

4.4 More on the relationship between the MDM and the investment climate variables

Continuing with the analysis presented so far and in order to go in depth into the relationship between the probability of observing a missing value in TFP and the IC variables we propose the next model for the probability of observing data on TFP in terms of IC and D variables

$$\Pr(s_i^a = 1 \mid D_i, IC_i) = \varphi(\rho_0^a + \rho_2^{a\prime} D_i + \rho_3^{a\prime} IC_i + \upsilon_i^a),$$

where s_{it}^a is a dichotomous variable taking value 1 if we do observe all sales, labor, materials and capital and zero otherwise. Symmetrically, for the case of sales we have the next equation

$$\Pr(s_i^b = 1 \mid D_i, IC_i) = \varphi(\rho_0^b + \rho_2^{b\prime} D_i + \rho_3^{b\prime} IC_i + \upsilon_i^b),$$

where in this case s_{it}^b takes value 1 if we observe data for sales.

Tables 6.1 to 6.4 present the estimating results by applying a LPM to model the probability of having a missing value conditional on the investment climate faced by firms. Concretely, we propose four models for each country. First we consider missing values in TFP conditioning in two different vectors of IC variables. The first specification includes the same set of IC variables than that included in equation (5); that is, the set of covariates statistically significant in the extended production function before imputing missing values by the ICA method. The second specification chooses the set of significant correlates starting from the whole set of IC variables and applying a general-to-specific procedure of selection of variables. The case of sales is symmetrical in the sense that model [3] uses the same set of IC variables than in equation (5), while the specification shown in column [4] selects the set of variables as we did in the case of column [2].

[TABLES 6.1 TO 6.4 ABOUT HERE]

The main motivation of these models, besides gathering evidence on which are the variables empirically associated with the MDM, is to know to what extent is needed to control for IC variables in the estimation of equation (5). Remember that even when the MDM is assumed to be MAR we still need the next moment condition:

$$E(s_{it}u_{it} \mid s_{it}\log L_{it}, s_{it}\log M_{it}, s_{it}\log K_{it}, s_{it}IC_i, s_{it}D_{it}) = s_{it}E(u_i \mid s_{it}\log L_{it}, s_{it}\log M_{it}, s_{it}\log K_{it}, s_{it}IC_i, s_{it}D_{it}) = 0,$$

and therefore independence between the set of IC variables we are interested in (those of equations (1) and (5)) and the MDM is achieved only before controlling for any variable correlated with the MDM. At this point, in the setting up of our model, the question is whether it is enough to use the matrix of IC variables of equations (1) and (5) or, as opposite, we have to find a better model for the MDM.

The results illustrate the clear relation between the MDM and the IC. Either we use missingness in TFP (model [2]) or in sales (model [4]), those IC variables are able to explain a large proportion of the variance of the MDM. Furthermore, the results come to confirm the analysis of section 4.3, auditing, innovative activity, financing, capacity, corruption or informality among others are significant covariates of the pattern of missing data in all the countries, even after controlling for size, industry and region effects.

Moreover, the IC variables used as covariates of equation (1) present high correlation with the MDM, especially in Turkey (see specifications [1] and [3]), supporting the assumption of exogenous sampling selection, with the IC variables influencing the data generating process. Thereby, controlling for those IC variables becomes a requisite.

The question that arises at this point is whether it is enough by controlling for the IC variables of equation (1)—those of specifications [1] and [3]—, or rather we have to select the set correlates of the MDM from the whole set of IC variables, as in specifications [2] and [4]. In this vein, we argue that models [1] and [3] incorporate most of the information on the IC we need. In order to test it we perform likelihood-ratio tests between model [1] in one side and [1] plus [2] in the other. Symmetrically, for the case of sales we compare model [3] with [3] plus [4]. In addition, we also compare the R^2, AIC and BIC criterions of model [1] with that of model [1] plus [2] ([3] with [3] plus [4] for of sales). Given these results in the remaining of the paper we only control for the IC variables included in equation (1).[27]

4.5 Some exhibits on the plausible correlation of PF variables and MDM

The descriptive analysis of the MDM is completed in figures 3.1 to 3.4. These figures compare the probability of picking an establishment with complete information for all production function variables with the probability of picking a establishment with information for sales (panel A) and at least one missing value in the remaining PF variables. Panels B, C and D, simply change sales

[27] We also believe that there exists a clear trade-off between parsimony and simplicity in the specification and adding additional controls for the MDM

by materials, capital and employment respectively. The aim of these figures is to know to what extent the pattern of missing values is correlated with PF variables. If the probability mass of picking a firm with missing value is accumulated around low values of sales, materials, capital and employment it could indicate that having a missing value is negatively related with the level of sales, materials, labor and/or capital. In other words, the probability of randomly draw a firm with information for sales and with, at least, one PF variable missing is higher within low sales firms. The same holds for materials and employment. The probability is lower for the case of capital. The same pattern is observed in India, Turkey, South Africa and Tanzania.

[FIGURES 3.1 TO 3.4 ABOUT HERE]

Figures 3.1 to 3.4 support the story of weaker firms reporting more missing values. However, the story is not conclusive yet. Low sales (and materials, capital and employment) firms usually don't need proper accountability, also tend to operate in more corrupt environments, are less innovative and dynamic, etc. In addition, as most of the firms are accumulated around low values it is easy to infer that the probability of picking a firm with any missing value in the PF variables will be higher within this range of values as well. From these figures we cannot conclude that low sales do not imply weakness or low productivity, and therefore higher probability of having missing values.

4.6 Can we relate the MDM and our endogenous variables by means of the ICSs?

So far we know that the MDMs in the countries analyzed are, in some way, related with a number of firms' attributes like accountability, corruption, openness, informality or size. However, we are not still able to conclude whether the MDM is determined independently of sales and TFP. The debate probably would end if we were able to construct a model of the probability of having a missing value and productivity (or sales) as RHS variable. Unfortunately this is not possible because, obviously, we do not observe either productivity or sales when we observe a missing value. However, we can still take advantage of the particular structure of the pattern of missing values to relate it with productivity or sales. Since the number of missing values reported increases when we move backward in time we can construct a model relating the probability of having a missing value in any PF variable in period t and productivity (tfp) in period t+1 plus other controls. That is, assuming that information in t+1 is better than in period t—remember that establishments report *recall* data—we propose the next model for the probability of having a missing value

$$\Pr(s_{it}^{a} = 1 \mid tfp_{it+1}, D_{it}, IC_{i}) = \varphi(\delta_{0}^{a} + \delta_{1}^{a} tfp_{it+1} + \delta_{2}^{a\prime} D_{it} + \delta_{3}^{a\prime} IC_{i} + \zeta_{it}^{a}),$$

where s^a takes value 1 if we observe all sales, labor, materials and capital and 0 otherwise.[28] Or alternatively we can also use the next model for sales

$$\Pr(s_{it}^b = 1 / y_{it+1}, D_{it}, IC_i) = \varphi(\delta_0^b + \delta_1^b y_{it+1} + \delta_2^{b\prime} D_{it} + \delta_3^{b\prime} IC_{it} + \zeta_{it}^b),$$

where s^b takes value 0 if we do not observe sales and y is the logarithm of firms' sales.

The questions we are willing to respond with these kinds of models is whether the probability of observing a missing value in period t-1 is correlated with the level of sales (productivity or TFP) in period t. Or in other words, are more productive/profitable firms more likely to keep track of their input/output accountability? Obviously, these models do not imply contemporaneous correlations but we think they might still be a good indicator between the actual relation between the level of sales/TFP and the MDM. On the other hand, an additional consideration have to be pointed out; there is a selection bias in the models as we are only able to use those observations with observable sales or TFP in t+1, so the resulting sub-sample is likely to be biased toward those responding firms. In order to reduce the degree of the bias we use those imputed values of sales or TFP in period t+1.[29]

[TABLE 7 ABOUT HERE]

The results of both equations for missingness in TFP and sales are in Table 7. Under endogenous sampling when the pattern of missing values is correlated with sales or TFP and if we were able to observe everything, we should expect a positive relation between contemporaneous TFP/sales and the missingness problem before controlling for other determinants such as IC and D variables. As a consequence the relation between missingness 'yesterday' and TFP/sales 'today' should be also positive. Table 7 supports this view for TFP (see Table 7 panel A) and for the cases of India and Turkey, where the $\hat{\delta}_1^a$ is positive and therefore more productive firms in year t+1 are associated with higher probability of being able of keeping track of proper accountability on output and inputs in past years. Note that we find this relation even before controlling for IC and D effects. However, the $\hat{\delta}_1^a$ for South Africa and Tanzania do not indicate any significant association between TFP and missingness in this countries. In the other hand, for the case of sales (panel B) we only observe a positive and significant effect of $\hat{\delta}_1^a$ in India, although the effect in Turkey is no longer significant. In South Africa and Tanzania the effect remains non-significant.

[28] In addition, if we assume a fist order Markov process for productivity, Pr(tfp_{t+1}/ tfp_t, tfp_{t-1},…)= Pr(tfp_{t+1}/ tfp_t) and therefore *tfp* in t+1 is a good proxy of *tfp* in period t the model is reduced to $\Pr(s_{it}^a = 1/tfp_{it}, D_{it}) = \varphi(\delta_0 + \delta_1 tfp_{it} + \delta_2' D_{it} + \upsilon_{it})$.

[29] Although by applying this strategy we reduce the degree of sample bias the problem remains to some extent. Nonetheless, we still believe that the models can be very informative on the relation of the plausible endogeneity of the MDM.

Therefore, Table 7 points to a plausible endogenous selection problem between missingness and TFP in India and Turkey, being the endogenous sampling selection problem corroborated in the case of sales in India but not in Turkey. As opposite, the analysis does not support this view in South Africa and Tanzania neither in the case of sales or TFP. Nonetheless, Table 7 does not allow concluding that there is a self-selection problem in India and Turkey, neither the MDM is MAR in South Africa and Tanzania. At this point caution is a requisite. All we are able to say is that we have four different patterns of data generating mechanisms, for some of them we find evidence of a more likely self-selection problem and under which we can test the performance of the various imputation methods, including the Heckman models.

4.7 Conclusions on the nature of the MDM

The question at the core of all the analysis of this section is whether the MDM in these countries is governed only by the level of sales or TFP (weakness) or rather the MDM can be explained by a number of firms attributes like the level of competitiveness, dynamism, corruption, informality, accountability and other indicators of the capacity of the firms: *MAR versus non ignorable missing data assumption*s.

According to the descriptive analysis presented the MDM mechanism has to do with informality and corruption but also with the capacity of the firms. More dynamic firms engaged in R&D, quality, innovation of new products, technologies and operating in more exigent and competitive export markets tend to report less missing values. Accountability can by itself explain a large share of missing data too. Much of these variables indicate that weaker firms tend to avoid reporting PF figures, and size is in some cases a good indicator of weakness as section 4.1 indicated. All these patterns are, in lower or greater extent, common to all the countries analyzed.

Notwithstanding this clear relation between IC and MDM, we cannot reject the hypotheses of non-ignorability in any of the cases. As already pointed out, this assumption is untestable from the available data. The preliminary descriptive analysis of section 4.5 points to a relation between the level of usage of inputs and output and missingness. Furthermore, previous econometric analyses of section 4.6 report a plausible relation between TFP and sales and missingness in t-1, especially in the cases of India and Turkey. In either case MAR or non-ignorable MDM, we believe that according with the analysis presented, controlling for those IC and D variables related with the missingness mechanism is a requisite, as can be shown from the LPM models presented for the probability of observing the required data to construct sales or TFP measures, and this is the way we proceed in the rest of the paper.

The aim of the next sections is to explore the dichotomy "*MAR versus non-ignorability*" of the MDM and their effects on the imputation mechanism proposed by comparing the sensitivity of the results of estimating the extended production function (1) under two assumptions: first, MDM is ignorable and therefore it may be explained by a number of exogenous firms' characteristics; and second, the MDM is endogenous and intimately linked to the level of sales and TFP of the firms. We also take advantage of the heterogeneity of the aprioristic relations observed between the MDM and their determinant in the four countries considered. This will allows us to illustrate how sensitive the results are under very different assumptions.

In addition, besides of testing the non-ignorable MDM, the analysis we present in what follows also allows us to study how the sensitivity of the imputations from the ICA method responds to: first, additional assumptions, such as randomness, or the amount of information embodied in the ICA method, all of them requiring the MAR assumption; and second, to different patterns of missing data: Turkey and Tanzania with a response rate for sales and TFP lower than 40% and India and South Africa with more than 70% of observations reported.

5 Robustness analysis

As pointed out, the aim of the paper is to compare the results of estimating equation (1) under the ICA method and several alternative imputation procedures. The methods presented to test the robustness of the results have their origins in two distinct bodies of statistical literature. The first one is related with likelihood-based inference with incomplete data, and, in particular, the EM algorithm. The second concerns techniques of Markov Chain Monte Carlo (MCMC), generally referred to as multiple imputation. We also consider extensions of the ICA method, allowing for additional randomness in the imputation procedure and the selection of the explanatory variables in equation (3). Lastly, we consider the estimation of (1) by sample selection estimation, such as different Heckman models.[30]

The literature on missing data points to the advantages of modern imputation mechanisms—EM-type algorithms and MCMC simulations—over other simpler methods based on basic standard regression techniques (such as the ICA method presented), see Allison (2001) and Little and Rubin (1987) for a review. Nonetheless, while most of this techniques has been widely evaluated under univariate missing data patterns (missingness for only one variable), or simple patterns of missinness in some of the variables of the dataset, the patterns of missing data observed in ICSs are very complex and unbalanced, even although we only consider PF variables and not the remaining IC variables. As an additional objective, it raises the possibility of evaluating the performance of modern imputation mechanisms under the complex and very different patterns of missing data observed in ICSs.

5.1 The *ICA Method* as an EM type algorithm

The EM algorithm has been widely applied in a broad range of applications, from missing data to latent variables models. Here we present several EM algorithms that will serve as benchmark to be compared with the *ICA method* proposed.

[30] Notice that although in this section we only analyze the behavior of PF variables as if they were the only set of imputed variables, in must be pointed out that IC variables are in all the cases imputed by the ICA method.

In particular, the aim is to test the sensitivity of the results obtained from the ICA method to other more sophisticated imputation mechanism allowing for additional randomness and amount of information embodied in the imputation mechanism. EM-type algorithms are based on an underlying likelihood function of the process generating data, and as a consequence impute missing data is based on draws from the posterior predictive distributions of the postulated missing data mechanism (or data generating process). A key issue under these mechanisms is whether the MDM may be considered as MAR or not.

5.1.1 EM-Algorithm on size, industry and region

Let J denote the vector dependent variable of interest, determined by the underlying unobserved vector variable J_{Mis}. Let $f^*(J_{Mis} \mid \mathbf{X}, \theta) = 0$ be the joint density of the latent variables conditional on the matrix of observed regressors **X**, and let $f(J \mid \mathbf{X}, \theta) = 0$ be the joint density of the observed variables. In essence, the maximum likelihood estimator (MLE) in this case maximizes

$$Q_N(\theta) = \frac{1}{N} L_N(\theta) = \frac{1}{N} \ln f^*(J_{Mis} \mid \mathbf{X}, \theta) - \frac{1}{N} \ln f(J_{Mis} \mid J, \mathbf{X}, \theta) \text{ .[31]} \tag{6}$$

The first term is not observed and therefore it is ignored, the second term is replaced by its expected value which does not involve J_{Mis}. The process is iterative, at the *r-th* round the expectation of the second term is evaluated at $\theta = \hat{\theta}_r$. The Expectation step of the algorithm therefore calculates

$$Q_N(\theta \mid \hat{\theta}_r) = -E\left[\frac{1}{N} \ln f(J_{Mis} \mid J, \mathbf{X}, \theta) \mid J, \mathbf{X}, \hat{\theta}_r\right]. \tag{7}$$

The Maximization step simply maximizes $Q_N(\theta \mid \hat{\theta}_r)$ to compute $\hat{\theta}_{r+1}$. Note that the iterative process continues until convergence is achieved.

In this paper we follow Cameron and Trivedi (2005) and propose the next EM type algorithm with our model rewritten as

$$\begin{bmatrix} J_1 \\ J_{Mis} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \beta + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}. \tag{8}$$

Where N_1 are the available observations and N_2 the missing observations and **X** denotes the explanatory variables. The EM algorithm consists on (1) estimate $\hat{\beta}$ using the N_1 available observations; (2) generate $\hat{J}_{Mis} = \mathbf{X}_2 \hat{\beta}$; (3) in order to mimic the distribution of J_1 generate adjusted values of $\hat{J}^a_{Mis} = (\hat{\mathbf{V}}^{-1/2} \hat{J}_{Mis}) \otimes \mathbf{u}_m$, where $\mathbf{u}_m$ is a Monte Carlo draw from the $N(0, s^2)$ distribution, being s^2 the variance of u_1 and a estimate of V can be obtained as

[31] Note that J* uniquely determines J but the inverse is not true, that is J does not uniquely determines J*, from the Bayes Rule it follows that $f(J \mid \mathbf{X}, \theta) = f^*(J^* \mid \mathbf{X}, \theta) / f^*(J^* \mid J, \mathbf{X}, \theta)$ (see Cameron and Trivedi, 2005).

$\hat{\mathbf{V}}(\hat{J}_{Mis}) \equiv \hat{\mathbf{V}}(\hat{J} \mid \mathbf{X}_2) = s^2(I_{N_2} + \mathbf{X}_2[\mathbf{X}_1{}'\mathbf{X}_1]^{-1}\mathbf{X}_2{}')$, and $\otimes$ denotes element by element multiplication; (4) using the augmented sample obtain a revised estimate of $\hat{\beta}$; (5) repeat steps (1) to (4) until convergence is achieved in the sense that the change in the sum of the square residuals becomes arbitrarily small.

Note that steps (3) and (4) are simply random draws from the conditional distributions of J given β in the case of step (3), and of β given s^2 in the case of step (4). In this first case, by means of direct comparisons with the ICA method, we include in the matrix **X** only the industry, region and size dummies. We also exclude of the imputation those observations with all production function variables missing.

Note the advantages of the EM algorithms over the ICA method. Since the EM algorithm work over the posterior predictive density, after each replication the new estimation of $\hat{\beta}$ improves the previous one—because in each iteration we are approaching the postulated distribution of the mechanism generating data. In addition, theoretically the estimates of s^2 improves the ones obtained in the *ICA method* as those are likely to be downward biased as they do not make allowance for the uncertainty inherent to J_{Mis}. Obviously, these advantages highly depend on the specification (model) chosen for the EM algorithm.

5.1.2 Extended EM-Algorithm on PF variables

The first alternative model for the EM algorithm is to extend matrix **X** to contain industry, region size, dummies and production function variables. The imputation now has two iterative processes. The first iteration process is the iterative EM algorithm per se, while the second one consists in replacing missing cells conditional on the information available for the remaining production function variables and the patterns of missing values observed (see Figures 1 to 4). We start by replacing that production function variable with the larger amount of missing values where **X** contains the remaining PF variables. We continue by applying the EM algorithm to the remaining PF variables.

5.1.3 Extended EM-Algorithm on PF and IC variables

In order to check the sensitivity of the results to the matrix **X** used, and therefore to the amount of information embodied in the EM algorithm, we include in this case industry/region/size dummies, PF variables and a large set of IC variables. Concretely the set of IC variables comes from the significant IC variables of equation (1). The idea is to check how EM algorithm responses to the amount of information incorporated in the imputation mechanism. Different results with respect to EM algorithms in sections 5.1.1 and 5.1.2 would pose some doubts on the validity of the ICA method as it does not incorporate enough information in the mechanism of imputation.

5.2 Further extensions of the *ICA method*

We now extend the ICA method to meet with additional assumptions on the MDM. In particular we develop the ICA method to incorporate some degree of randomness in the imputation. We also propose an ICA method in which the dependent variable of the model (sales or logY) is excluded from the imputation procedure.

5.2.1 Random industry-region-size replacement: *random ICA Method*

Under the two assumptions mentioned in section 3 (normality of replaced variables and linearity, apart from the MAR assumption) ICA method leads to consistent estimation of the parameters of equation (1). However, it can be argued that a more efficient method can be used. Notice that by imputing missing values we are modifying the population distribution of replaced variables. In particular, if the two conditions mentioned in section 3 hold the sample average of the modified distribution of the variable converges to the population expectation. Unfortunately, this is not true for the case of the standard deviation. With the replacement strategy we are reducing the variability of the distribution of those variables with missing values and therefore any statistical inference will be based on downward biased standard errors. Moreover, the bias in the standard errors will be higher as the proportion of missing values increases and the sample size decreases.

This problem will arise whenever we use imputed data as if it were real data. It has to do with the lack of uncertainty in the estimation of the parameters of estimating regressors equations and reflects the fact that conventional formulas to compute standard errors does not correct for imputed data.

The ICA method, although deterministic, introduce variability in the imputation of missing data by replacing missing cells by industries, regions and sizes with the variability given by *I*R*S* being *I, R* and *S* the numbers of industries, regions and sizes respectively. A good question is therefore whether this variation is enough or the ICA method lead to downward biased standard errors. To do it we propose an alternative variation of the ICA method consisting in adding a random part to each imputed value.

The new replacement strategy is again based on the expectation of equation (3), but in this case a random term is added in order to embody uncertainty to the imputation mechanism

$$\tilde{J}_{it} = \hat{\rho}_0 + \hat{\rho}_{R,J} D_{R,it} + \hat{\rho}_{I,J} D_{I,it} + \hat{\rho}_{S,J} D_{S,it} + \hat{\sigma}_{J,\varepsilon} \xi_{J,it} \qquad J = Y, L, M, K \tag{9}$$

where $\hat{\sigma}_{J,\varepsilon}$ is the standard error of the residual $\varepsilon_{J,it}$ from

$$J_{it} = \rho_0 + \rho_{R,J} D_{R,it} + \rho_{I,J} D_{I,it} + \rho_{S,J} D_{S,it} + \varepsilon_{J,it} \qquad J = Y, L, M, K$$

and $\xi_{J,it}$ is a random draw from $\varepsilon_{J,it}$. In particular we take 100 random draws from $\varepsilon_{J,it}$ constructing 100 candidate values to replace each missing cell in the data matrix. To do the definite replacement we compute the average across the 100 candidate values.

5.2.2 Random industry-region-size replacement: *bootstrap ICA Method*

Another problem coming from the lack of uncertainty inherent to deterministic imputation methods is that, generally, when some instruments and or regressors are estimated in a first stage (our case for production function variables) the asymptotic variance needs to be adjusted because of the generated instruments, see Pagan (1984), Newey (1984), Murphy and Topel (1985) and Newey and McFadden (1994).[32]

A plausible solution for this problem is to compute the bootstrap estimate of the standard errors of the estimated coefficients of equation (5). The idea is to create '*r*' replications of the original sample using as strata industry and region. In the next step and for each replication, we apply equation (4) to replace the missing data and to estimate equation (5). The result will be a bootstrap distribution of the estimators of equation (4) under different replacements of missing data that can be used to compute the bootstrap estimates of the standard errors.

5.2.3 *ICA method* on the inputs

One can also be concerned with the imputation of missing data in the dependent variable of equation (1), sales. At this respect it can be argued that the MDM may be correlated with the dependent variable of (1), so imputing missing values in sales and estimate (2) by OLS or standard econometric techniques is not a valid solution. In this case, when s depends on log$Y$, it is clear that s and u are no longer uncorrelated, even though we control for IC and D variables. In particular when s is correlated with $logY$ in equation (2) there is a self selection problem that should be handled with other sample selection corrections like the Heckman model as we will see later on.

Here we propose the same replacement mechanism than in section 3, but in this case excluding sales of the replacement process. The extended production function to be estimated is therefore

$$s_{it}^{**}\log Y_{it} = s_{it}^{**}(\alpha_0 + \alpha_L \log \tilde{L}_{it} + \alpha_M \log \tilde{M}_{it} + \alpha_K \log \tilde{K}_{it} + \alpha_{IC}' IC_i + \alpha_D' D_{it}) + s_{it}^{**}\tilde{\tilde{u}}_{it}, \quad (12)$$

with identification conditions symmetrical to those of equation (5).

Note that when there is no sample selection, incomplete data is MAR, the incompleteness of logY is not as large that it makes the complete case no representative of real population and we are not concerned with efficiency, estimating (12) by standard techniques is equivalent to estimating (5) or (2). On the contrary, when there is a sample selection problem the point of reference to compare with (12) would be the Heckman selection model.

[32] More precisely, the problem appears when testing the null hypotheses $H_0: \psi = 0$, where $\psi = \alpha, \beta, \delta, \omega$ are the coefficients of generated regressors (see equation 1). Before including the generated regressors in (1), the usual test statistic on ψ has a limiting standard normal distribution under H_0. However, when $\psi \neq 0$ standard t statistics will not be asymptotically valid and an adjustment is needed for the asymptotic variances of all estimators of generated regressors.

5.3 Multiple imputation via switching regression

The aim now is to propose different imputation mechanisms to compare their performance with the ICA method and its variations. The following imputation mechanism was first proposed by van Buuren, Boshuizen and Knook (1999) and it has been chosen because it fits very well with datasets with a large amount of missing values in many variables, like IC datasets. See also Schafer (1999) for a tutorial on multiple imputation, and Schafer (1997) and Gelman, King and Liu (1998) for applications.

The basic idea is to create a small number of copies of data, each of which has the missing values suitably imputed. Each imputed dataset is then analyzed independently. Estimates of the parameters of interest are properly averaged across the copies of data, while standard errors are computed according to *'Rubin rules'*, see Rubin (1987). In particular this multiple imputation mechanism is accomplished in the following steps:

1. Specify the posterior predictive density of incomplete data as $p(J_{MIS}|X,s)$ given the non response mechanism is $p(s \mid J, IC, C, D)$ and the complete data model is $p(J, IC, C, D)$, where X is the set of covariates used in the imputation mechanism and s is the pattern of missing values. The posterior predictive density is generally given by

$$p(J_{MIS} \mid X, s) = \int p(J_{MIS} \mid X, s, \theta) p(\theta \mid X, s) d\theta \tag{13}$$

   where the standard procedure to impute missing data consist in, first, draw a value of θ^* from $p(\theta \mid X, s)$ and, second, draw a value J_{Mis}* from $p(J_{MIS} \mid X, s, \theta = \theta^*)$.
2. The next step is to draw imputations from this density to produce m complete datasets. Here we follow van Buuren et al. (1999) and we produce m=5 datasets.
3. Estimate equation (1) m times.
4. Pool the m results.

This imputation mechanism involves the choice of the form of the linear model and the predictor variables. In particular, we use a linear regression of each J_{MIS}= Y, L, M *and* K on a set X of predictor variables, where the set of predictor variables is given by X=$Y, K, L, M,$ and D. Note that each J is used as predictor variable and as imputed variable in (10), while D are used only as predictor variables.

5.4 Sample selection correction (I): Heckman on complete case

If the pattern of missing values is endogenously determined (it is correlated with output ($\log Y$) in equation (4)), thereby appearing a self-selection problem, the ICA method may lead to inconsistent estimates of parameters of (1). In these cases one has to implement the Heckman (1976) or Heckit method to correct for self-selection, since OLS applied either on the complete case or on the sample with replacement is inconsistent. In particular, Heckman model over the complete case is given by

$$E(\log Y_{it} \mid \log L_{it}, \log M_{it}, \log K_{it}, IC_i^H, D_{it}, s_{it}=1) = \alpha_0 + \alpha_L \log L_{it} + \alpha_M \log M_{it} + \alpha_K \log K_{it} + \beta' IC_i + \omega' D_{it} + \rho\lambda(\gamma_L \log L_{it} + \gamma_M \log M_{it} + \gamma_K \log K_{it} + \gamma_{IC}' IC_i^H + \gamma' D_{it}), \quad (14)$$

where as usual $\rho\lambda(.)$ is simply the inverse of Mills ratio or Heckman's lambda given by the next Probit

$$\Pr(s=1 \mid l_{it}, m_{it}, k_{it}, IC_i^H, D_{it}) = \Phi(\gamma_L \log K_{it} + \gamma_M \log M_{it} + \gamma_K \log K_{it} + \gamma_{IC}' IC_i^H + \gamma' D_{it}), \quad (15)$$

with the next moment condition $E(u \mid \log L_{it}, \log M_{it}, \log K_{it}, IC_i, IC_i^H, D_{it}) = 0$.

Heckman method is subject to high sensitivity to model choice, requiring a good knowledge on the nature of the missing data mechanism. For this reason, the selection of the Probit model in (12) goes from the general to the specific to select the variables with a significant effect on the probability of having a missing value. Concretely, the selection of variables starts with a wide set of more than 120 IC and D variables in each country. Eventually, the final set of significant variables is reduced to a number around 15 and 25.

5.5 Sample selection correction (II): Heckman imputing inputs with *ICA method*

In 3.4 the selection of Heckman model is based on the complete case. In this section we propose to perform the same model on the sample after replacing missing values in employment, materials and capital according to equations (10) and (11). Heckman model in this case is given by

$$E(\log Y_{it} \mid \log \tilde{L}_{it}, \log \tilde{M}_{it}, \log \tilde{K}_{it}, IC_i^H, D_{it}, s_{it}=1) = \alpha_0 + \alpha_L \log \tilde{L}_{it} + \alpha_M \log \tilde{M}_{it} + \alpha_K \log \tilde{K}_{it} + \beta' IC_i + \omega' D_{it} + \rho\lambda(\gamma_L \log \tilde{L}_{it} + \gamma_M \log \tilde{M}_{it} + \gamma_K \log \tilde{K}_{it} + \gamma_{IC}' IC_i^H + \gamma' D_{it}), \quad (16)$$

with Heckman's Lambda and moment condition obtained symmetrical to previous sub-section. Note that equation (17) is directly comparable with equation (12).

In addition, in sections 5.2.1 and 5.2.2 we introduced the problem of lack of uncertainty in the estimation of the standard errors of estimating regressors equations. A solution proposed was to obtain the bootstrap standard errors under replacement of missing values in each resampling. The solution here is similar: we obtain the bootstrap standard errors to make statistical inference and to correct for the problem aforementioned. More precisely, we will compare the standard errors from the estimating sample with the bootstrap estimator of the standard errors, which will give us a benchmark on how serious this issue is in our case.

6 Empirical results

The objective of this section is to evaluate to what extent the results obtained from the ICA method are influenced by different assumptions on the MDM. In particular, as we pointed out in section 5 under the ICA method we have to consider two different key assumptions on the patterns of missing data. First, if we can assume MDM as MAR, in which case we test the goodness of the fit of the ICA method to other more sophisticated mechanisms that are supposed to work better as they consider the randomness issue and are able to include more information in the imputation mechanisms. And second, the MDM is non-ignorable and therefore we are forced to apply sample selection corrections such as Heckman models.

The evaluation of the ICA method is based on the kernel estimates of inputs and output and the underlying TFP densities under all the imputation mechanism proposed. We also present the empirical results from estimating the extended production function (1) under different imputation methods. In all the cases we use the ICA method as benchmark for comparison purposes. In all the regressions, outliers, defined as those observations with ratios of labor cost to sales and/or materials to sales greater than one, are excluded.

6.1 Evaluation of imputation mechanism: Comparison of estimated inputs and output densities

The kernel densities of $\log\tilde{Y}_{it}$, $\log\tilde{L}_{it}$, $\log\tilde{M}_{it}$, $\log\tilde{K}_{it}$ for each country and for the complete case, the ICA method, the random ICA method and the three EM-type algorithms considered are in figures 4.1 to 4.4. In turn, the descriptive statistics of the variables under each imputation mechanism are in tables 8.1 to 8.4.

[FIGURES 4.1 to 4.4 & TABLES 8.1 TO 8.4 ABOUT HERE]

We find that the proportion of missing values is an important factor in the observed underlying distributions after imputing missing values. So by means of explanation it is useful to discuss the results by groups of countries. The first group, with India and South Africa, comprises those countries with the largest response rate of PF variables, 65% in India and 70% in South Africa. The second group includes Tanzania and Turkey whose response rates are only 40 and 30% respectively.

As shown in the kernel densities the response rate dramatically determine the shape of the densities after imputing missing values. In India (see Figure 4.1) where the response rate is reasonably high in all the variables but capital, all the methods lead to estimated densities similar to those of the complete case. However, in the case of capital where the response rate is considerably lower we observe a dramatic change in the distribution of the imputed values by the Random ICA method. Concretely, the distribution appears to have two modes, moving a considerable proportion of density from the center of the distribution to the right. This misleading behavior is already pointed out in the case of materials, although in a lower extent.

In what refers to the estimated distributions of the remaining imputation mechanism, all of them lead to results similar to those of the complete case, including the ICA method and EM algorithms. Nonetheless, in terms of descriptive statistics, it is noticeable that in spite of the uncertainty inherent to the EM algorithm [1], it slightly reduces the estimated standard deviation of all PF variables, even with respect to the ICA method case, probably due to the higher amount of imputed cells than under other mechanisms. Nonetheless, it must be also pointed out that the reduction of the standard deviation is only of the order of one decimal point. In this sense, the Random ICA method, and the remaining EM algorithms increase to some extent the estimated standard errors with respect to the ICA method.

The case of South Africa is virtually symmetrical to that of India. Again the Random ICA method performs badly in the case of capital. Likewise, due to the larger proportion of missing values imputed, the EM algorithm [1] lead to estimated standard errors that slightly reduce those of the complete case.

As the response rate of PF variables decreases, the estimated densities obtained from the EM algorithms and Random ICA method tend to be different than those of the complete case and the standard ICA method, especially in the case of the Random ICA method. This is illustrated in the cases of Turkey and Tanzania in figures 4.2 and 4.4. Nonetheless, the estimated descriptive statistics are rather homogeneous among imputation methods, as shown in tables 8.2 and 8.4. The estimated means are virtually equal in all the cases, and the standard errors show great consistency across specifications, but in the EM algorithm [1] where, again due to the larger proportion of values imputed, the standard errors are slightly lower.

It is useful to recapitulate the main conclusions of this subsection before introducing the results of estimating equation (1). Overall, there are small differences in the imputation of PF variables. Nonetheless, these differences become more marked as the number of missing values increases and when the variables are far of being normally distributed.

6.2 Evaluation of imputation mechanism: Comparison of estimating results of equation (1)

6.2.1 Comparison of *ICA method* and other EM algorithms

Tables 9.1, 9.2, 9.3 and 9.4 show the results of estimating equation (5) after imputing missing values by the ICA method and by the three EM algorithms proposed in section 5.1. A key conclusion is that when the proportion of missing values is not large enough there are no remarkably differences between applying the ICA method or the EM algorithm [1], neither in the point estimates of the input-output (I-O) elasticities, nor in the standard errors (recall that uncertainty is a key issue under EM algorithms). Another interesting observation is that we do not gain too much by extending the EM algorithm to include the IC variables among the information set.

[TABLES 9.1 TO 9.4 ABOUT HERE]

Table 9.1 focuses on the Indian case, in which the ICA method and the EM algorithm on industry, region and size variables (EM algorithm [1]) lead to similar results in terms of input-output elasticities. However, there are divergences in the input-output elasticities estimated for the remaining two EM-algorithms. Concretely, employment coefficient decreases from 0.1 in the ICA method and EM algorithm [1], to 0.05 in the EM algorithms [2] and [3]. Similarly, it is worth mentioning that the estimates of the standard errors of the coefficients of the input-output elasticities do not improve in the EM algorithm [1] with respect to the ICA method, being even lower in the EM algorithms [2] and [3].

It is important to note that most of the differences between the ICA method and the EM algorithm [1] in one side and the EM algorithms [2] and [3] in the other side can be explained by the higher amount of information embodied in the imputation process—production function variables in the EM [2] and production function, IC, and D variables in EM [3]; and not by the iterative process based on posteriors predictive densities as in the EM algorithms. When the pattern of missing data is very unbalanced and we are able to observe only one or two PF variables for each cross-sectional observation, those EM algorithms including additional variables, beyond of the region/industry/size dummies, are more likely to lead to heterogeneous results as they include a different amount of information for each cross-section. This becomes more patent in case of the EM algorithm [3], in which we include also IC variables in the imputation.

Apart from this observation, the elasticities and semi-elasticities of IC variables show a reasonable robustness to the imputation mechanism used. In general terms, the ICA method is more consistent with the results from the EM algorithm [1], whereas EM algorithms [2] and [3] show more differences. For example, out of 6 IC variables significant in the ICA method case, 5 are also significant in the EM algorithm [1], while only 3 in the EM algorithms [2] and [3] (see Table 12). Nonetheless, the changes observed are only in the magnitude of the coefficients estimated, and never in the direction of the effects. All the estimated IC coefficients move within a reasonable range of values in the four cases.

[TABLE 12 ABOUT HERE]

The case of South Africa in Table 9.3, with a pattern of missing values similar to that of India, leads to analogous conclusions. Again the I-O elasticities estimated under the ICA method are rather similar to those we get under the EM algorithm [1], whereas the EM algorithms [2] and [3] diverges in the sense that the estimated I-O elasticity for employment is almost one percent point lower than in the ICA method and EM algorithm [1]. The patterns observed for the standard errors estimated are the same to those of India: almost equal standard errors between the ICA method and the rest of EM algorithms, so no improvements of efficiency can be observed from using the EM algorithms in this case. Concretely, from Table 12 there are 10 significant IC variables under the ICA method, the same variables are significant again under the EM algorithm

[1] (plus other three new significant IC variables). In the EM algorithms [2] and [3] only 7 IC variables out of 10 repeat significance.

The patterns observed in India and South Africa are not supported by the Turkish case in Table 9.2. Recall that the proportion of missing values among PF variables reach 70%, and therefore the affects of the imputation mechanism used will be quite different than those applied to patterns of missing data with only a 20 or 30% of response rate. In this case it is remarkable that I-O elasticities in the EM algorithms [1], [2] and [3] are closer to constant returns to scale (CRS) than the ICA method is. In this sense, and in terms of I-O elasticities, the results from the ICA method are different from the EM algorithms, with materials and capital elasticities significantly lower than in the remaining cases. However, the estimated standard errors do not change much and the significance of the PF variables is not modified in any of the cases. In spite of these changes in the I-O elasticities, it is important to note that again the IC parameters appear to be robust to the imputation method used. 10 IC variables turned out to be significant in the ICA method case, 12 in the EM algorithm [1] and 14 in the EM algorithms [2] and [3]. Apart from minor changes of the magnitude of the coefficients, and in some cases in the significance of some variables, we do not observe changes in the estimated directions of the effects of the IC variables.

Finally, the case of Tanzania is presented in Table 9.4. The proportion of missing values in PF variables in this country is more than 70% of the original sampling frame, similar to that of Turkey. However, unlike Turkish case, EM algorithms [2] and [3] do not improve the results obtained from the ICA method. Again, ICA method and EM algorithm show a symmetrical behavior with similar I-O elasticities, whereas in EM algorithms [2] and [3] the estimated elasticity for employment is three times lower than in the ICA method, increasing in turn the elasticity of materials. In the other hand, almost all of those IC variables significant in the ICA method repeat significance in the EM algorithms, and what is more important, the coefficients are robust to all the imputation mechanisms, apart from marginal differences in some variables (see Table 12).

6.2.2 Comparison of ICA method with complete case, extensions of ICA method and multiple imputation

In this section we compare the results obtained from the ICA method with those from the complete case, other extensions of the ICA method (see section 5.2) and multiple imputation (see section 5.3) in tables 10.1 to 10.4. Table 10.1 focuses in the case of India. The fourth column comprises the results of the complete case, for which the number of observations is considerably reduced with respect to the ICA method case, from 5211 to 3943. In spite of the reduced number of observations used, there are not significant changes either in the estimated I-O elasticities, or in its level of significance. In what refers to the IC parameters, it is worth mentioning that, although there are no changes in the directions of the estimated effects, and the coefficients are rather robust in both specifications, some of the variables lost their significance in the complete case with respect to the ICA method. Thus, out of the 6 significant IC variables in the ICA method, only 1 is also significant in the complete case.

[TABLES 10.1 TO 10.4 ABOUT HERE]

Especially interesting is the comparison of the ICA method with the Random ICA method—introduced in section 5.2.1—in which we introduce a random component to the imputation procedure in order to test the role played by the uncertainty inherent to the imputation mechanism. In this vein, another interesting point is to check the sensitivity of the significance level of the variables to using bootstrap standard errors to correct for the problem of generated regressors (see section 5.2.2). Only 2 IC variables lose their significance in the ICA method with bootstrap standard error with respect to the regular case, and 2 new variables became significant. A similar pattern is observed in the Random ICA method with 6 significant IC variables, out of which 3 were also significant in the ICA method (Table 12 includes the summary of significant IC variables in each case).

Finally, the ICA method on inputs and the multiple imputation cases lead to similar results in the I-O elasticities, with the exception of a slight decline of the capital elasticity. In both cases, the significance of some IC variables is lost, although the direction of the estimated effects never changes.

Similar conclusions can be obtained in the case of South Africa, which results are presented in Table 10.3. In this case the number of observations used in the complete case only differs by 250 with respect to the ICA method. As expected from the larger response rate of PF variables in this country, there is not significant efficiency lost in the complete case and most IC variables remain significant. Alike Indian case, the Random ICA method and the bootstrap standard errors change the significance of some variables, and while some variables lose its significance, a small group of other IC variables become significant. Finally, both the ICA method on inputs and multiple imputation show robust results with respect to the ICA method. We only observe changes in the second or third decimals.

The cases of Turkey and Tanzania (tables 10.2 and 10.4 respectively) are rather different that the two previously commented. In both cases using the complete case implies using less than 50% of the sample under the complete case. This implies a clear efficiency lost, what is translated to four less significant IC variables in the complete case in Turkey and three in Tanzania. By means of significance of IC variables the results from the Random ICA, Bootstrap ICA method and ICA on inputs cases are more consistent with those from the standard ICA method. At this respect, introducing more uncertainty in the imputation procedure used in Turkey does not change the significance of 6 and 9 IC variables depending on whether we focus in the Bootstrap ICA or in the Random ICA respectively. In Tanzania the patterns are similar, 4 IC variables lost its significance in both the Bootstrap ICA and the Random ICA. Lastly, in both cases, Turkey and Tanzania, the ICA method on inputs and the multiple imputation do not modify the results of the ICA method.

In the other hand, regarding I-O elasticities and for the case of Turkey, it is important to note that, although we only observe changes in the I-O estimate for materials, the I-O elasticity of employment is non-significant under the ICA method with bootstrap standard errors and the Random ICA method.

6.2.3 Comparison of ICA method and Heckman selection model

We now focus in the comparison of the ICA method and the Heckman models proposed in section 5.4 and 5.5. The estimating results are in tables 11.1 to 11.4. The main conclusions are summarized in Table 12.

[TABLES 11.1 TO 11.4 ABOUT HERE]

First of all, we consider important to note that the Heckman's Lambda is significant in none of the four cases. Thereby, the plausible selection bias is not supported by the Heckman model in any country.

Besides the significance of the Heckman's Lambda the results are rather similar when we correct for the endogenous selection and when we do not. In India and South Africa there are no significant changes in the I-O elasticities. Nonetheless, the larger proportion of missing observations in Turkey and South Africa introduces some degree of heterogeneity between the results of the ICA method and the Heckman models. Even under very different estimated I-O elasticities, the IC parameters moves within a reasonable range of values and there are not changes in the estimated direction of the effects. Overall, there are more IC variables significant in the Heckman model, even when we consider bootstrap standard errors.

6.3 Evaluation of imputation mechanism: Comparison of estimated TFP densities

We finish this section with the evaluation of the estimated densities of the TFPs for each country. The estimated kernel densities of the different TFP measures obtained after applying the different imputation mechanism are obtained from equation (1) according to the next expression $\log TFP_{it} = s^{*}_{it}[\log \tilde{Y}_{it} - (\hat{\alpha}_L \log \tilde{L}_{it} + \hat{\alpha}_M \log \tilde{M}_{it} + \hat{\alpha}_K \log \tilde{K}_{it})]$, where $\log TFP_{it}$ is the measured productivity after the imputation process, $\log \tilde{Y}_{it}$, $\log \tilde{L}_{it}$, $\log \tilde{M}_{it}$, $\log \tilde{K}_{it}$ are the imputed inputs and output, the alphas with a hat on top denotes the different estimated input-output elasticities after imputing missing values and $s*$ is the pattern of missing values in PF variables after the imputation process. The results are in figures 5.1 to 5.2, along with the descriptive statistics of each TFP measure and the correlation matrix among productivities.

Again it is needed to distinguish between two blocks of countries. In the first one, say that compounded by India and South Africa, the estimated TFP measures show a similar shape of the kernel densities, although with different estimated mean, especially in the case of EM algorithm [1] in India. In South Africa this pattern is more marked, with more ostensible differences in the first moment of the distribution of the different TFP measures, although all the kernel densities have a similar shape, indicating that the standard deviations do not differ much among them. What is corroborated in panels B and C, where the descriptive statistics and the matrix of correlations are shown.

[FIGURES 5.1 TO 5.4 ABOUT HERE]

In Turkey and Tanzania the results are somewhat different. The larger proportion of missing values in these two countries, results in two different blocks of TFP measures. The first block comprises the TFP measures from the complete case, the ICA method on inputs, and the EM algorithms [2] and [3]. The second block includes the remaining measures, that is, those from the ICA method, the EM algorithm [1] and the Random ICA method. TFP measures are similar within each group, however between blocks there are evident differences in all the shape of the distribution, the skewness, the kurtosis, as well as in the estimated means and standard errors, as panel B shows. In spite of all these differences, panel C shows that the correlations of the TFP measure from the ICA method with the remaining cases are in between .8 and .99. Likewise, the correlation among the remaining measures is considerably high.

6.4 Summary and main conclusions

The ICA method performs reasonably well. Even under very different patterns of missing data and assumptions we are able to get robust results among different methods of handling missing data after controlling for IC variables in the estimation. When we assume that the MDM is MAR then there are two main issues we should consider: uncertainty and amount of information used in the imputation. In the other hand, if a non-ignorable pattern of missing data is assumed, then we are forced to test the robustness of the results of the ICA method with the Heckman models.

We find that, overall, the ICA method is a good alternative even when the proportion of missing values is relatively high and the underlying variables are manifestly non-normal., leading to rather homogenous results than other more sophisticated methods. We also observe that none of uncertainty, amount of information and non-ignorability of the MDM are big issues in the context of ICSs. Or at least they are not as serious as to invalidate the results of the ICA method. Lastly, we find that in order to get robust results under different imputation mechanism it is key to control for the same set of IC variables, as they contain much information on the MDM.[33]

[33] Obviously, this assertion is conditional on the objectives one may have.

The main conclusions of this section can be summarized as follows:

- Overall, there are small differences in the estimated distribution of the imputed PF variables. Nonetheless, these differences become more marked as the number of missing values imputed increases and when the variables are not normally distributed. In particular, The Random ICA method, is the mechanism with the worst performance under large proportion of missing values, followed by the EM algorithms. The ICA method preserves with reasonable precision the main moments of the distribution of the variables in the complete case.[34]
- These differences in the estimated distributions become even clearer if we focus in the TFP. However, the conclusions are the same whether we focus on inputs and output or TFP.
- We found reasonably robust elasticities in equation (1) under all the imputation methods proposed. However, there are important differences in the I-O elasticities and in the significance of the IC variables.
- The ICA method, EM algorithm [1], Random ICA method and Bootstrap ICA method lead to homogeneous results among them. That is, introducing uncertainty to the ICA method, besides whether in order to get it we use the EM algorithm [1], Random ICA method or Bootstrap ICA method, does not change significantly either the estimated effects or the level of significance of IC variables This suggest that uncertainty is not a big issue. Obviously, there are slight differences in the standard errors, but we argue that they are not as serious as to invalidate the results of the ICA method.
- In all the cases EM algorithms [2] and [3] lead to differences in the I-O estimates, although the IC parameters are again quite robust and do not vary much, the level of significance is affected in a higher proportion of cases than in the EM algorithm [1].
- More importantly, EM algorithms [2] and [3] are not homogeneous among them, suggesting that the amount of information embodied in the imputation algorithm does not improve the results consequently.
- Another interesting observation is that the performance of the EM algorithms [2] and [3] highly depends on the structure of the MDM. When the pattern of missing data is very unbalanced, meaning by that that it is common to observe only one or two PF variables in each cross-sectional observation, these two EM algorithms lead to rather different results than the ICA method and EM algorithm [1]. Intuitively, this is probably due to the unbalanced amount of information included in each cross-sectional observation.

[34] This would imply that the ICA method performs well when the MDM is MCAR or MAR, since in that case, under regularity conditions, the distribution in the complete case share the same characteristics of the population distribution. Nonetheless, at this point if the MDM is non-ignorable we cannot say anything about the goodness of the fit of the ICA method, since it could be replicating any distribution different from the population one.

- Only in Tanzania and Turkey, when the proportion of missing values is larger than in the other two countries, we observe significant changes in the estimated I-O elasticities under the Heckman models with respect to the ICA method.
- As a general rule, there are more significant IC variables under the Heckman models than under the ICA method.
- The Heckman's Lambda is never significant. What does not support the story of non-ignorable MDM. And comes to confirm that correcting for endogenous selection does not change considerably the results.
- It is also important to note that it doesn't matter whether we replace only the independent variables, the dependent variable or both. In all the cases, the results are similar among them. More importantly, the Heckman model with the inputs replaced by the ICA method and the case of the ICA method on the inputs are similar in both cases.
- Finally, we find of key importance to control for IC variables in the estimation in all the cases. We believe that this is what allows us to get so robust results under very different assumptions and patterns of missing data. This is supported by section 4.4, where we saw that IC variables are able to explain a rather important proportion of the variability of the MDM in all the countries.

7 Conclusions

When the missing data mechanism (MDM) is ignorable, the objective of the imputation methods is not to augment the sample size, but to preserve the sample representativity, to gain efficiency in the estimation and to retrieve for the analysis a large number of very expensive interviews. The alternative to these methods is the complete case or *listwise deletion*, which is not a panacea even when the MDM is ignorable. Operating with the complete case is only acceptable if incomplete cases attributable to missing data comprise a small percentage, say 5% or less, of the number of total cases (Schafer, 1997), and when the complete case preserves the representativeness of the original sampling frame. In addition, in models with a large number of regressors missing data problem may encourage analysts to leave out of the regression some explanatory variables with high proportion of missing values. As Cameron and Trivedi (2005) point out, this practice may be misleading as it leads to an omitted variables problem, which could be more serious than the missing data problem *per se*. The first question we raise in this paper is hence whether the researcher has to do something with regard to the missing values when dealing with investment climate surveys (ICSs).

In the context of ICSs a large proportion of the sample size is lost in the complete case and the representativeness of the original sample frame is to some extent modified. Given these results, the MDM is far of can be considered as missing completely at random (MCAR), and consequently complete case could lead to inconsistent and inefficient results. In order to overcome this problem we propose a imputation mechanism that fits well with the characteristics of ICSs—with unbalanced patterns of missing data and low proportion of available observations in the complete case—likely to be used to construct structural models composed by single, or even systems of, equations with a large number of explanatory variables, all of them containing missing data.

The imputation method proposed, which we call *ICA method*, departs from the class of EM type algorithms and lies on the expectation of the imputed variables conditional to the sector, region and size they belong to. The performance of the ICA method depends on several characteristics of the MDM like the number of variables replaced or of the proportion of missing values in the complete case; but specially it depends on the nature of the MDM: *missing at random (MAR) or non-ignorable*. Taking into account this, we analyze the MDM of four countries with very different patterns of missing data (India, Turkey, South Africa and Tanzania) to know to what extent the MDM can be treated as MAR or not. Although not conclusive on the nature of the MDM, the descriptive analysis shows that this has to do with a variety of IC determinants such as informality and corruption and also with the capacity of the firms. More dynamic firms engaged in R&D, quality, innovation of new products, technologies and operating in more exigent and competitive export markets tend to report less missing values. Accountability and size can by themselves explain a large share of missing data too. In the other hand, the analysis does not allow us to reject the non-ignorability assumption on the MDM in any case.

In addition, given the results of the descriptive analysis and beyond of the discussion between MAR and non-ignorable MDM, an interesting results is the need of controlling for those variables related with the MDM. Inconsistency would follow if we do not control for the large set of IC variables in the estimation.

In the next step of the analysis presented in the paper, we estimate an extended production function under imputation of missing values by the ICA method and we test the estimating results to other imputation mechanisms. We first considered imputation mechanisms requiring the MAR assumption alike the ICA method, including complete case, EM algorithms, extensions of the ICA method and multiple imputation. We then included in the analysis methods considering the non-ignorable assumption on the MDM, essentially we consider the Heckman model under different specifications.

Although caution is always a requisite when drawing conclusions from a model with imputed data, the *ICA method* leads to robust results to more sophisticated imputation methods requiring also the MAR assumption. We observe that more complex imputation mechanisms are rather sensitive to both the proportion of missing values and how these missing values are distributed among variables. When the MDM is very unbalanced, in the sense that we may observe only one or two PF variables for each cross-sectional observation those EM algorithms including additional explanatory variables like inputs or IC variables lead to changes in the results compared with the more linear, parsimonious and simpler ICA method and EM algorithm [1], both including only industry/region/size variables always available. Suggesting that more complex imputation methods based on simulations, especially EM algorithms and multiple imputation based on Markov Chain Monte Carlo, require a deeper and thorough knowledge of MDM that would allow us to handle proper assumptions on the unknown densities of data generating processes. The issue of the sensitivity of the results to the selection of a proper model for the MDM constitutes an interesting question to be handled in further research regarding ICSs.[35]

In this sense, we believe that incorporating systematically more information in the imputation mechanism does not constitute, per se, an improvement of the estimates. Rather, given the sensitivity of the results to the model choice for the MDM, extending the matrix of covariates used to impute missing values require a careful and deep knowledge of the determinants of the MDM, and this is likely to vary country by country.

Regarding the lack of uncertainty inherent to the ICA method as a deterministic imputation method, we find that using other mechanisms allowing for additional uncertainty in the imputation mechanisms such as the so called Random ICA method, Bootstrap ICA method or EM algorithms does not change significantly the results. Despite there are changes in the level of significance of some coefficients, most of the variables remain significant when incorporating additional randomness. Nonetheless, we also observe that the randomness issue becomes more important as the proportion of missing values increases (cases of Turkey and Tanzania).

In the other hand, provided we control for the same set of IC variables in all the specifications, the results under the complete case and the ICA method are reasonably consistent between them. Even in those cases in which the complete case represents less than half of the original sampling frame, the estimated parameters of production function (PF) and IC variables ranges within a reasonable range of values. This illustrates the importance of using the large set of IC variables in order to control for the data generating process in the estimation.[36]

Likewise, the ICA method shows reasonable robustness to the endogenous sampling case. The Heckman's lambda is non-significant in all the cases, what does not support the endogenous sampling selection hypotheses. The results of the ICA method are similar to those of the Heckman regressions, indicating that although even if there would be an endogenous sampling selection problem this is not as serious as to bias the final results. In this sense, replacing only those RHS variables and not the dependent variable (sales in our case) do not change the results provided the endogenous sample selection is not supported by the models and the robustness among the results.

[35] ICSs in particular and data collected from developing countries in general present the missingness issue as an additional challenge for applied researchers. We consider that a proper and systematic methodology to deal with this problem is a need, especially if more sophisticated imputation mechanisms are applied.

[36] In order to go in depth into this issue further research is needed. Nonetheless, once proved the relation between IC variables and the MDM using them to get independency between our model and the MDM is a requisite. We believe that this procedure is what balances the results, in the sense that it is what allows us to get robust results among specifications.

As the use of the Investment Climate Surveys become more and more important among policy makers, scholars and applied researchers a careful investigation on the causes of the missingnes problem in order to improve the quality of the data is becoming a requisite. The parsimonious methodology we propose here is intended to be a first step to help preparing the way forward and going in depth into this line of research.

References

Allison, P.D (2001): "Missing Data," *Quantitative Applications in the Social Sciences*. Sage University Paper.

Buuren v S., H.C. Boshuizen and D.L. Knook (1999): "Multiple imputation of missing blood pressure covariates in survival analysis," *Statistics in Medicine*; 18(6):681-94.

Cameron, A.C. and P.K Trivedi (2005): "Microeconometrics: Theory and Applications," *Cambridge University Press.*

Dempster A.P., N.M. Laird and D.B. Rubin (1977): "Maximum Likelihood Estimation for Incomplete Data Via the EM Algorithm," *Journal of the Royal Statistical Society*, Series B, 39, 1-38.

Escribano, A., and J. L. Guasch (2005): "Assessing the Impact of Investment Climate on Productivity Using Firm Level Data: Methodology and the Cases of Guatemala, Honduras and Nicaragua," *World Bank Policy Research Working Paper* #3621, Washington, DC.

Escribano, A., and J. L. Guasch (2008): "Robust Methodology for Investment Climate Assessment on Productivity: Application to Investment Climate Surveys from Central America," *Working Paper* # 08-19 (11), *Universidad Carlos III de Madrid*.

Escribano, A., J. L. Guasch, and M. de Orte (2009): "INDIA: Investment Climate Assessment on Productivity, Allocative Efficiency and Other Economic Performance Measures of the Manufacturing Sector,"mimeo, *Universidad Carlos III de Madrid*.

Escribano, A., J.L. Guasch, M. de Orte and J. Pena (2008a): "Investment Climate and Firm's Performance: Econometric and Applications to Turkey's Investment Climate Survey," *Working Paper* # 08-20 (12), *Universidad Carlos III de Madrid*.

Escribano, A., J.L. Guasch, M. de Orte and J. Pena (2008b): "Investment Climate Assessment Based on Demean Olley and Pakes Decompositions: Methodology and Applications to Turkey's Investment Climate Survey," *Working Paper* # 08-20 (12), *Universidad Carlos III de Madrid*.

Escribano, A., J. L. Guasch and J. Pena (2009): "Assessing the Impact of Infrastructure Quality on Firm Productivity in Africa," *World Bank Policy Research Working Paper* #(forthcoming), Washington DC.

Escribano, A., M. de Orte, J. Pena and J. L. Guasch. (2009): "Investment Climate Assessment on Economic Performance Using Firm Level Data: Pooling Manufacturing Firms from Indonesia, Malaysia, Philippines and Thailand from 2001 to 2002," *Singapore Economic Review, Special Issue on the Econometric Analysis of Panel Data*, Vol. 54, #3, August 20092009.

Gelman, A., G. King and C. Liu (1998): "Not Asked and Not Answered: Multiple Imputation for Multiple Surveys," *Journal of the American Statistical Association*, 93, 846-874.

Griliches, Z. (1986): "Economic Data Issues," *Handbook of Econometrics*, Vol, III. Ed. R.F. Engle and D. McFadden. Amsterdam: North Holland, 1464-1514.

Heckman, J.J. (1976): "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement* 5, 475-492.

Heitjan, D.F. (1994): "Ignorability in General Incomplete-Data Models," *Biometrika*, 81, 701-708.

___________(1999): "Causal Inference in Clinical Trials, A Comparative Example," *Controlled Clinical Trials*, 20, 309-318.

Heitjan D.F. and S. Basu (1999): "Distinguising 'Missing at Random' and `Missing Completely at Random'," *The American Statistician*, 50, 207-213.

Ibrahim, J.G., S.R. Lipsitz and M-H Chen (1999): "Missing Covariates in Generalized Linear Models When the Missing Data Mechanism is Ignorable," *Journal of the Royal Statistical Society*, Ser. B, 61, 173-190.

Little R.J.A. and D.B. Rubin, (1987): "Statistical Analysis with Missing Data," *Wiley Series in Probability and Mathematical Statistics*. John Wiley and Sons Eds.

McLachlan G.J and T. Krishnan (1997): "The EM Algorithm and Extensions," New York: *Wiley*.

Meng, X.L. (2000): "Missing Data: Dial M for?" *Journal of the American Statistical Association*, Vol. 95, No. 452, (Dec., 2000), pp. 1325 -1330.

Molerberghs, G., E.J. Goetghebeur, S.R. Lipsitz, M.G. Kenward (1999): "Nonrandom Missingness in Categorical Data: Strengths and Limitations," *The American Statistician*, 53, 110-118.

Murphy, K. M. and R. H. Topel (1985): "Estimation and inference in two-step econometric models," *Journal of Business and Economic Statistics,* 3(4): 370–379.

Newey, W.K (1984): "A Method of Moments Interpretation of Sequential Estimators," *Economic Letters,* 14, 201-206.

Newey, W.K and D. McFadden (1994); "Large Sample Estimation and Hypotheses Testing," *Handbook of Econometrics*, Vol, 4. Ed. R.F. Engle and D. McFadden. Amsterdam: North Holland, 2111-2245.

Pagan, A.R (1984): "Econometric Issues in the Analysis of Regressions with Generated Regressors," *International Economic Review,* 25, 221-247.

Rubin, D.B. (1976): "Inference and Missing Data," *Biometrika*, 63, 581-592.

Rubin, D.B. (1987): "Multiple Imputation for Nonresponse in Surveys," *New York: Wiley*.

Schafer, J.L (1997): "Analysis of Incomplete Multivariate Data," *London: Chapman and Hall*.

Schafer, J.L (1999): "Multiple Imputation: A Primer," *Statistical Methods in Medical Research*, 8, 3-15.

Wooldridge, J.M (2007): "Econometric Analysis of Cross Section and Panel Data," *The MIT Press*. Cambridge, Massachusetts.

Appendix I: definition of IC variables

I Infrastructures			
IC variables	**Country**	**Measurement units**	**Definition**
Longest days to clear customs to export	(IND)	Log	Longest number of days that it took to clear customs when exporting
Days to clear customs to imports	(TUR, SA)	Log	Average number of days that it takes to clear customs when importing
Dummy for own generator	(IND)	0 or 1	Dummy variable taking value 1 if the firm has an own generator
Electricity from own generator	(TUR, TZA)	Percentage	Percentage of total electricity used that came from oen generators
Losses due to power outages	(IND, TUR, SA, TZA)	Perc	Percentage of total annual sales lost as a result of power outages
Wait for electric supply	(SA)	Log	Average number of days that it takes to obtain a power supply
Water supply from public sources	(IND)	Perc.	Percentage of the water used by the establishment that came from public sources
Water from own well or water infrastructure	(SA)	Perc.	Percentage of the water used by the establishment that came from own well or water infrastructures
Losses due to water outages	(TUR, TZA)	Perc.	Percentage of total annual sales lost as a result of waqter outages
Water outages	(SA)	Log	Total number of water outages suffered by year
Wait for a water supply	(TUR, TZA)	Log	Average number of days that it takes to obtain a water supply
Shipment losses in the domestic market	(IND, TUR)	Perc.	Percentage of products shipped that were lost as a consequence of theft, breakage, or spoilage
Dummy for own transport	(IND)	0 or 1	Dummy variable taking value 1 if uses own transport services
Average duration of transport failures	(SA)	Log	Average duration in hours of transport failures
Transport outages	(TZA)	Log	Total number of transport failures by year
Losses due to transport delay	(IND, TZA)	Perc.	Percentage of total annual sales lost as a consequence of transport delays
Losses due to phone outages	(TZA)	Perc.	Percentage of total annual sales lost as a consequence of phone interruptions
Dummy for web page	(IND, SA, TZA)	0 or 1	Dummy variable taking value 1 if the firm uses web page to communicate with clients or suppliers
Dummy foe e-mail	(IND, TUR, SA)	0 or 1	Dummy variable taking value 1 if the firm uses e-mail to communicate with clients or suppliers
Sales lost due to delivery delays	(SA)	Perc.	Percentage of total annual sales lost as a consequence of delivery delays
Dummy for own roads	(TZA)	0 or 1	Dummy variable taking value 1 if the firm has own roads.
Low quality supplies	(TZA)	Perc.	Percentage of total supplies that were of lower of upon agree quality by year
Days of inventory of main supply	(TZA)	Log	Days of inventory that the establishment kept its main supply in storage in average during the last year

II Red tape, corruption and crime			
IC variables	**Country**	**Measurement units**	**Definition**
Crime losses	(TUR, SA)	Perc.	Percentage of total annual sales lost as a consequence of crime, vandalism or arson
Dummy for security	(IND)	0 or 1	Dummy variable taking value 1 if the firm has security expenses
Security expenses	(TUR, SA, TZA)	Perc.	Security expenses as a percentage of total annual sales
Illegal payments in protection	(SA, TUR)	Perc.	Illegal payments in protection (e.g. to organized crime) to prevent violence as a percentage of total annual sales by year
Manager's time spent in bur. Issues	(TUR, SA)	Perc.	Percentage of manager's time spent in dealing with bureaucratic issues
Payments to deal with bureaucratic issues	(TUR, SA, TZA)	Perc.	Payments to deal with bureaucratic issues as a percentage of total annual sales
Payments to obtain a contract with the government	(TUR, SA, TZA)	Perc.	Payments to obtain a contract with the government as a percentage of total annual sales
Dummy for payments to speed up bureaucracy	(IND)	0 or 1	Dummy variable taking value 1 if the establishment declared to make payments to 'speed up' bureaucratic issues
Dummy for payments to deal with bur. Issues	(IND)	0 or 1	Dummy taking value 1 if the firm declared to make 'irregular' payments to deal with bureaucratic issues
Dummy for interventionist labor regulation	(IND)	0 or 1	Dummy taking value 1 if the firm considers that regulation affected its decisions to hire or fire employees
Gift to obtain a operating license	(TZA)	Perc.	Gifts as a percentage of total annual sales paid to get an operating license
Number of inspections	(TUR)	Log	Total number of inspections suffered by the firm by year
Days in inspections	(TZA)	Log	Total number of days that the firm received inspections from public officials during the last year
Sales reported to taxes	(IND, TUR, SA)	Perc.	Percentage of total annual sales reported to IRS tax authorities
Workforce reported to taxes	(IND)	Perc.	Percentage of total workforce reported to IRS tax authorities
Production lost due to absenteeism	(IND, TUR)	Log	Days production lost as a consequence of employees absenteeism
Dummy for informal competition	(TUR)	0 or 1	Dummy variable taking value 1 if the firm declared to compete against informal competition
Dummy for lawsuit	(TUR)	0 or 1	Dummy variable taking value 1 if the firm had any lawsuit during the last year

III Finance			
IC variables	**Country**	**Measurement units**	**Definition**
Dummy for external audit	(IND, TUR, SA)	0 or 1	Dummy taking value 1 if the firm has its annual statements reviewed by an external auditor
Dummy for trade association	(IND)	0 or 1	Dummy variable taking value 1 if the firm belongs to a trade association
Dummy for loan	(IND, SA)	0 or 1	Dummy taking value 1 f the firm has access to a loan from any financial institution
Largest shareholder	(IND, SA)	Perc.	Percentage of firm's equity that belongs to the largest shareholder
Dummy for credit line	(TUR, SA, TZA)	0 or 1	Dummy taking value 1 if the firm has access to a credit line from any financial institution
Percentage of credit unused	(SA)	Perc.	Percentage of the credit line that is currently unused
Dummy for loan with collateral	(IND)	0 or 1	Dummy taking value 1 if the firm has a loan with a collateral associated
Value of the collateral	(SA)	Perc.	Value of the collateral as a percentage of the total value of the loan
Borrows denominated in foreign currency	(IND, TUR, SA, TZA)	Perc.	Percentage of total firm's borrows that were denominated in foreign currency
Dummy for loan denominated in Turkish Lira	(TUR)	0 or 1	Dummy taking value 1 f the firm has access to a loan denominated in Turkish Lira
Dummy for loan denominated in foreign currency	(TUR)	0 or 1	Dummy taking value 1 f the firm has access to a loan denominated in foreign currency
Dummy for long-term loan	(TUR)	0 or 1	Dummy taking value 1 f the firm has access to a loan for more than 1 year
Interest rate of the loan	(TZA)	Perc.	Interest rate of the last loan obtained by the firm
Dummy for new land purchased	(TUR)	0 or 1	Dummy taking value 1 if the firm obtained a new land in the last year
Charge to clear a check	(SA)	Perc.	Charges to clear a check as a percentage of the value of the check
Delay to clear a domestic currency wire	(TZA)	Log	Average number of days that it take to clear a domestic currency wire
Working capital financed by domestic private banks	(IND)	Perc.	Percentage of working capital financed by funds from domestic private banks
Working capital financed by commercial banks	(TZA)	Perc.	Percentage of working capital financed by funds from commercial banks
Working capital financed by foreign commercial banks	(SA)	Perc.	Percentage of working capital financed by funds from foreign commercial banks
Working capital financed by informal sources	(SA)	Perc.	Percentage of working capital financed by funds from informal sources
Working capital financed by leasing	(TZA)	Perc.	Percentage of working capital financed by funds from leasing arrangement
Dummy for current or saving account	(TZA)	0 or 1	Dummy taking value 1 if the firm has acces to a current or saving account
Inputs bought on credit	(TZA)	Perc.	Percentage of inputs bought on credit by year
Sales bought on credit	(TZA)	Perc.	Percentage of sales bought on credit by year

IV Quality innovation and labor skills			
IC variables	**Country**	**Measurement units**	**Definition**
Dummy for R&D	(IND)	0 or 1	Dummy taking value 1 if the firm invests in R&D
Dummy for new technology	(TUR, TZA)	0 or 1	Dummy taking value 1 if the firm introduced a new technology inherent to the production process during the last year
Dummy for new product	(SA, TZA)	0 or 1	Dummy taking value 1 if the firm introduced a new product of product line during the last year
Dummy for product innovation	(IND, TZA)	0 or 1	Dummy taking value 1 if the firm introduced a product innovation during the last year
Dummy for discontinued product line	(SA)	0 or 1	Dummy taking value 1 if the firm discontinued the production of any product during the last year
Dummy for foreign license	(IND, TUR, TZA)	0 or 1	Dummy taking value 1 if the firm has a technology licensed from a foreign company
Dummy for internal training	(IND, SA, TZA)	0 or 1	Dummy taking value 1 if the firm provide training to its employees
Training to unskilled workers	(SA)	Perc.	Percentage of unskilled workers that received training during the last year
Workforce with computer	(IND, TZA)	Perc.	Percentage of workers in the staff that regularly uses computer at job
Dummy for ISO quality certification	(IND, TUR, SA)	0 or 1	Dummy taking value 1 if the firm has an ISO quality certification
Dummy for outsourcing	(IND, SA, TZA)	0 or 1	Dummy taking value 1 of the firm outsourced any part of production in the last year
Dummy for brought in house	(TZA)	0 or 1	Dummy taking value 1 if the firm brought in house any part of the production process previously outsourced
Dummy for external training	(IND)	0 or 1	Dummy taking value 1 if the firm provided external training to its employees
Staff - skilled workers	(TZA)	Perc.	Percentage of skilled workers in staff
Staff - professional workers	(TZA)	Perc.	Percentage of professional workers in staff
Unskilled workforce	(IND)	Perc.	Percentage of unskilled workforce in staff
Staff with university education	(TUR, SA)	Perc.	Percentage of staff with at least one year of university education
Staff-part time workers	(TUR)	Perc.	Percentage of part time workers in staff
Staff - management	(SA)	Perc.	Percentage of management in the staff
Staff - non-production workers	(SA)	Perc.	Percentage of non-production workers in staff
Manager's experience	(SA)	Log	Manager's experience in years
Dummy for closed plant	(SA)	0 or 1	Dummy taking value 1 if the firm closed a plant during the year previous to the survey
Dummy for joint venture	(TZA)	0 or 1	Dummy taking value 1 if the firm agreed to do a joint venture during the last year

V Other control variables			
IC variables	**Country**	**Measurement units**	**Definition**
Dummy for incorporated company	(IND, TUR, TZA)	0 or 1	Dummy taking value 1 if the firm is constituted as an incorporated company
Age	(IND, TUR, SA)	Log	Age of the firm in years
Share of exports	(IND)	Perc.	Percentage of total annual sales exported
Trade union	(IND)	Perc.	Percentage of workers that belong to a trade union
Strikes	(IND, TUR)	Log	Days of production lost due to strikes
Market share	(TUR, SA)	Perc.	Share of market share
Dummy for recently privatized firm	(TUR)	0 or 1	Dummy taking value 1 if the firm was privatized within the last five years
Dummy for competition against imported products	(TUR)	0 or 1	Dummy taking value 1 if the firm competes against imported products
Capacity utilization	(SA)	Perc.	Percentage of total capacity used by the firm the last year
Dummy for FDI	(TZA)	0 or 1	Dummy taking value 1 if the firm receive FDI inflows
Dymmy for industrial zone	(TZA)	0 or 1	Dummy taking value 1 if the firm is located in an industrial zone

Appendix II: Tables and figures

Table 1.1: Observations available for regression analysis after and before imputing missing values and outliers in 43 ICSs

		Year of the survey	Obss. In the sampling frame	Complete case		After imputing missing cells	
				#Obs.	% respect to sampling frame	#Obs.	% respect to sampling frame
Latin America	Argentina	2006	746	372	*49.9*	664	*89.0*
	Bolivia	2006	409	209	*51.1*	336	*82.2*
	Colombia	2006	649	525	*80.9*	618	*95.2*
	Mexico	2006	1,161	778	*67.0*	1,093	*94.1*
	Panama	2006	243	97	*39.9*	223	*91.8*
	Peru	2006	361	230	*63.7*	337	*93.4*
	Paraguay	2006	440	111	*25.2*	315	*71.6*
	Uruguay	2006	396	155	*39.1*	304	*76.8*
	Chile	2006	697	382	*54.8*	629	*90.2*
	Costa Rica	2005	1029	643	*62.5*	970	*94.3*
	Ecuador	2006	394	235	*59.6*	346	*87.8*
	Salvador	2006	467	296	*63.4*	439	*94.0*
	Honduras	2006	263	189	*71.9*	243	*92.4*
	Guatemala	2006	328	262	*79.9*	316	*96.3*
	Nicaragua	2006	365	230	*63.0*	341	*93.4*
Africa	Algeria	2002	1,904	1,114	*58.5*	1,412	*74.2*
	Benin	2004	591	364	*61.6*	475	*80.4*
	Botswana	2006	114	109	*95.6*	113	*99.1*
	Cameroon	2006	119	117	*98.3*	118	*99.2*
	Egypt	2004	2,931	1,317	*44.9*	2,629	*89.7*
	Eritrea	2002	237	61	*25.7*	179	*75.5*
	Ethiopia	2002	1,281	1,048	*81.8*	1,142	*89.1*
	Kenya	2003	852	360	*42.3*	585	*68.7*
	Madagascar	2005	870	383	*44.0*	623	*71.6*
	Malawi	2005	320	208	*65.0*	288	*90.0*
	Mali	2003	462	242	*52.4*	309	*66.9*
	Mauritius	2005	636	271	*42.6*	417	*65.6*
	Morocco	2003	2,550	2,352	*92.2*	2,422	*95.0*
	Namibia	2006	106	100	*94.3*	104	*98.1*
	Senegal	2003	783	253	*32.3*	535	*68.3*
	South Africa*	2003	1,737	1,229	*70.8*	1,492	*85.9*
	Tanzania*	2003	828	325	*39.3*	561	*67.8*
	Uganda	2003	900	368	*40.9*	695	*77.2*
	Zambia	2002	564	391	*69.3*	417	*73.9*
Asia	Indonesia	2003	1,214	486	*40.0*	1,041	*85.7*
	Malaysia	2001	1,732	605	*34.9*	1,317	*76.0*
	Philippines	2003	1,432	1,092	*76.3*	1,272	*88.8*
	Thailand	2004	2,766	646	*23.4*	1,502	*54.3*
	Pakistan	2007	2358	990	*42.0*	2,144	*90.9*
	Bangladesh	2006	4804	2,533	*52.7*	3,946	*82.1*
	India*	2005	6849	4448	*64.9*	5750	*84.0*
Europe	Croatia	2007	419	219	*52.3*	372	*88.8*
	Turkey*	2005	2646	771	*29.1*	1,619	*61.2*

Complete case includes those observations without missing values and or outliers in sales, materials, capital, labor cost and labor
Source: Authors' calculations with IC data.

Table 1.2: Missing values in IC variables and its incidence on complete case

A. India

IC variables included	# variables	[1]		[2]	
		# obs. Available	% over total	# obs. Available	% over total
All IC variables [(a)]	115	0	0.0	0	0.0
those IC vars. with response rate >70% [(b)]	80	500	7.3	588	8.6
those IC vars. with response rate >80% [(c)]	71	942	13.8	1188	17.3
those IC vars. with response rate >90% [(d)]	63	1663	24.3	2202	32.2
those IC vars. with response rate >95% [(e)]	40	2109	30.8	2817	41.1

B. Turkey

IC variables included	# variables	[1]		[2]	
		# obs. Available	% over total	# obs. Available	% over total
All IC variables [(a)]	90	1	0.0	4	0.2
those IC vars. with response rate >70% [(b)]	78	426	16.1	740	28.0
those IC vars. with response rate >80% [(c)]	77	472	17.8	1226	46.3
those IC vars. with response rate >90% [(d)]	75	523	19.8	1394	52.7
those IC vars. with response rate >95% [(e)]	65	697	26.3	2034	76.9

C. South Africa

IC variables included	# variables	[1]		[2]	
		# obs. Available	% over total	# obs. Available	% over total
All IC variables [(a)]	168	0	0.0	0	0.0
those IC vars. with response rate >70% [(b)]	112	93	5.1	114	6.3
those IC vars. with response rate >80% [(c)]	108	391	21.6	451	24.9
those IC vars. with response rate >90% [(d)]	92	620	34.3	769	42.5
those IC vars. with response rate >95% [(e)]	81	828	45.8	1089	60.2

D. Tanzania

IC variables included	# variables	[1]		[2]	
		# obs. Available	% over total	# obs. Available	% over total
All IC variables [(a)]	162	0	0.0	0	0.0
those IC vars. with response rate >70% [(b)]	98	6	0.7	9	1.1
those IC vars. with response rate >80% [(c)]	89	32	3.9	69	8.3
those IC vars. with response rate >90% [(d)]	71	118	14.3	251	30.3
those IC vars. with response rate >95% [(e)]	40	227	27.4	548	66.2

[1] PF variables are also included In the computation of the final number f observations available in the complete case.
[2] PF variables do not included In the computation of the final number f observations available in the complete case.
[(a)] All IC variables are included in the computation of the number of observations available in the complete case.
[(b)] Only those IC variables with a response rate larger than 70% are included in the computation of the number of observations available in the complete case.
[(c)] Only those IC variables with a response rate larger than 80% are included in the computation of the number of observations available in the complete case.
[(d)] Only those IC variables with a response rate larger than 90% are included in the computation of the number of observations available in the complete case.
[(e)] Only those IC variables with a response rate larger than 80% are included in the computation of the number of observations available in the complete case.
Source: Authors' estimation with ICSs.

Table 2.1: INDIA, Percentage of observations lost due to missing values by industry and size

	Size	Small		Medium		Large		Total	
Industry		#Obs	%Lost[(d)]	#Obs	%Lost	#Obs	%Lost	#Obs	%Lost
Food	Sampling frame[(a)]	333		177		87		597	
	Complete case[(b)]	177	46.9	79	55.4	51	41.4	307	48.6
	With replacement[(c)]	248	25.5	137	22.6	69	20.7	454	24
Textiles & Leather	Sampling frame	426		255		207		888	
	Complete case	251	41.1	210	17.7	139	32.9	600	32.4
	With replacement	325	23.7	235	7.8	178	14	738	16.9
Apparels	Sampling frame	360		315		150		825	
	Complete case	247	31.4	267	15.2	120	20	634	23.2
	With replacement	287	20.3	290	7.9	138	8	715	13.3
Chemicals & Chemical prds	Sampling frame	426		333		171		930	
	Complete case	262	38.5	218	34.5	130	24	610	34.4
	With replacement	337	20.9	282	15.3	150	12.3	769	17.3
Plastics & Rubbers	Sampling frame	279		189		12		480	
	Complete case	193	30.8	112	40.7	11	8.3	316	34.2
	With replacement	243	12.9	157	16.9	11	8.3	411	14.4
Non-metallic products	Sampling frame	105		63		48		216	
	Complete case	40	61.9	38	39.7	32	33.3	110	49.1
	With replacement	75	28.6	50	20.6	39	18.8	164	24.1
Structural metal & metal prds	Sampling frame	618		252		39		909	
	Complete case	328	46.9	131	48	21	46.2	480	47.2
	With replacement	526	14.9	214	15.1	31	20.5	771	15.2
Machinery & Equipment	Sampling frame	1074		687		243		2004	
	Complete case	749	30.3	482	29.8	160	34.2	1,391	30.6
	With replacement	912	15.1	603	12.2	213	12.4	1728	13.8
Total	Sampling frame	3621		2271		957		6849	
	Complete case	2,247	38	1,537	32.3	664	30.6	4,448	35.1
	With replacement	2953	18.5	1968	13.3	829	13.4	5750	16.1

Notes:
[(a)] "Sampling frame" refers to the total number of observations (firms surveyed multiplied by the number of years of information).
[(b)] "Complete case" refers to the complete case on production function variables (sales, materials, capital and labor), missing values in other IC variables—different from production function— are not considered.
[(c)] "With replacement" refers to the sample after imputing IC variables according to the ICA Method; missing values in other IC variables—different from production function—are not considered. Notice that only observations with information available in at least one of sales, labor, labor cost, materials or capital, are imputed
[(d)] "Perc. lost" refer to the percentage of observations lost with respect to the sampling frame.
Source: Authors calculations with IC data.

Table 2.2: TURKEY, Percentage of observations lost due to missing values by industry and size

	Size	Small		Medium		Large		Total	
Industry		#Obs	%Lost[d]	#Obs	%Lost	#Obs	%Lost	#Obs	%Lost
Food and Beverages	Sampling frame[a]	192		170		202		564	
	Complete case[b]	56	70.8	57	66.5	82	59.4	195	65.4
	With replacement[c]	134	30.2	116	31.8	150	25.7	400	29.1
Textiles and Apparels	Sampling frame	110		230		398		738	
	Complete case	14	87.3	47	79.6	115	71.1	176	76.2
	With replacement	48	56.4	130	43.5	257	35.4	435	41.1
Chemicals	Sampling frame	118		98		136		352	
	Complete case	24	79.7	29	70.4	51	62.5	104	70.5
	With replacement	60	49.2	67	31.6	87	36.0	214	39.2
Non-metallic mineral products	Sampling frame	54		66		46		166	
	Complete case	15	72.2	20	69.7	19	58.7	54	67.5
	With replacement	46	14.8	51	22.7	30	34.8	127	23.5
Metal products (ex. M&E)	Sampling frame	94		98		92		284	
	Complete case	30	68.1	43	56.1	34	63.0	107	62.3
	With replacement	68	27.7	82	16.3	59	35.9	209	26.4
Machinery and Equipment	Sampling frame	98		78		80		256	
	Complete case	37	62.2	31	60.3	38	52.5	106	58.6
	With replacement	79	19.4	52	33.3	63	21.3	194	24.2
Electrical machinery	Sampling frame	58		40		36		134	
	Complete case	19	67.2	19	52.5	15	58.3	53	60.4
	With replacement	42	27.6	34	15.0	24	33.3	100	25.4
Transport equipment	Sampling frame	64		30		58		152	
	Complete case	31	51.6	18	40.0	15	74.1	64	57.9
	With replacement	54	15.6	25	16.7	46	20.7	125	17.8
Total	Sampling frame	788		810		1048		2646	
	Complete case	226	71.3	264	67.4	369	64.8	859	67.5
	With replacement	531	32.6	557	31.2	716	31.7	1804	31.8

Notes:
[a] "Sampling frame" refers to the total number of observations (firms surveyed multiplied by the number of years of information).
[b] "Complete case" refers to the complete case on production function variables (sales, materials, capital and labor), missing values in other IC variables—different from production function— are not considered.
[c] "With replacement" refers to the sample after imputing IC variables according to the ICA Method; missing values in other IC variables—different from production function—are not considered. Notice that only observations with information available in at least one of sales, labor, labor cost, materials or capital, are imputed
[d] "Perc. lost" refer to the percentage of observations lost with respect to the sampling frame.
Source: Authors calculations with IC data.

Table 2.3: SOUTH AFRICA, Percentage of observations lost due to missing values by industry and size

	Size	Small		Medium		Large		Total	
Industry		#Obs	%Lost[(d)]	#Obs	%Lost	#Obs	%Lost	#Obs	%Lost
Food & beverages	Sampling frame[(a)]	22		80		87		189	
	Complete case[(b)]	13	40.9	49	38.8	69	20.7	131	30.7
	With replacement[(c)]	14	36.4	66	17.5	82	5.7	162	14.3
Textiles & apparels	Sampling frame	12		43		120		175	
	Complete case	6	50	32	25.6	69	42.5	107	38.9
	With replacement	10	16.7	33	23.3	101	15.8	144	17.7
Chemicals, rubber & plastics	Sampling frame	42		119		118		279	
	Complete case	21	50	79	33.6	87	26.3	187	33
	With replacement	29	31	111	6.7	101	14.4	241	13.6
Paper, edition & publishing	Sampling frame	13		89		54		156	
	Complete case	10	23.1	65	27	45	16.7	120	23.1
	With replacement	10	23.1	78	12.4	49	9.3	137	12.2
Machinery & equipment	Sampling frame	47		252		256		555	
	Complete case	25	46.8	198	21.4	212	17.2	435	21.6
	With replacement	35	25.5	222	11.9	241	5.9	498	10.3
Wood & furniture	Sampling frame	13		74		58		145	
	Complete case	7	46.2	55	25.7	39	32.8	101	30.3
	With replacement	11	15.4	69	6.8	50	13.8	130	10.3
Non-metallic products	Sampling frame	13		23		30		66	
	Complete case	3	76.9	18	21.7	22	26.7	43	34.8
	With replacement	6	53.8	18	21.7	26	13.3	50	24.2
Other	Sampling frame	27		63		57		147	
	Complete case	19	29.6	38	39.7	47	17.5	104	29.3
	With replacement	25	7.4	50	20.6	51	10.5	126	14.3
Total	Sampling frame	189		743		780		1712	
	Complete case	104	45	534	28.1	590	24.4	1228	28.3
	With replacement	140	25.9	647	12.9	701	10.1	1488	13.1

Notes:

(a) "Sampling frame" refers to the total number of observations (firms surveyed multiplied by the number of years of information).

(b) "Complete case" refers to the complete case on production function variables (sales, materials, capital and labor), missing values in other IC variables—different from production function— are not considered.

(c) "With replacement" refers to the sample after imputing IC variables according to the ICA Method; missing values in other IC variables—different from production function—are not considered. Notice that only observations with information available in at least one of sales, labor, labor cost, materials or capital, are imputed

(d) "Perc. lost" refer to the percentage of observations lost with respect to the sampling frame.

Source: Authors calculations with IC data.

Table 2.4: TANZANIA, Percentage of observations lost due to missing values by industry and size

	Size	Small		Medium		Large		Total	
Industry		#Obs	%Lost[d]	#Obs	%Lost	#Obs	%Lost	#Obs	%Lost
Food & beverages	Sampling frame[a]	105		87		51		243	
	Complete case[b]	47	55.2	44	49.4	17	66.7	108	55.6
	With replacement[c]	82	21.9	57	34.5	31	39.2	170	30
Textiles & apparels	Sampling frame	33		41		19		93	
	Complete case	10	69.7	14	65.9	5	73.7	29	68.8
	With replacement	26	21.2	24	41.5	8	57.9	58	37.6
Chemicals, rubber & plastics	Sampling frame	23		55		24		102	
	Complete case	10	56.5	18	67.3	14	41.7	42	58.8
	With replacement	13	43.5	40	27.3	16	33.3	69	32.4
Paper, edition & publishing	Sampling frame	27		39		9		75	
	Complete case	8	70.4	19	51.3	6	33.3	33	56
	With replacement	16	40.7	30	23.1	9	0	55	26.7
Machinery & equipment	Sampling frame	49		29		9		87	
	Complete case	14	71.4	6	79.3	6	33.3	26	70.1
	With replacement	36	26.5	21	27.6	8	11.1	65	25.3
Wood & furniture	Sampling frame	133		53		9		195	
	Complete case	52	60.9	13	75.5	3	66.7	68	65.1
	With replacement	89	33.1	23	56.6	5	44.4	117	40
Non-metallic products	Sampling frame	11		16		6		33	
	Complete case	3	72.7	11	31.3	5	16.7	19	42.4
	With replacement	9	18.2	12	25	6	0	27	18.2
Total	Sampling frame	381		320		127		828	
	Complete case	144	62.2	125	60.9	56	55.9	325	60.7
	With replacement	271	28.9	207	35.3	83	34.6	561	32.2

Notes:

[a] "Sampling frame" refers to the total number of observations (firms surveyed multiplied by the number of years of information).

[b] "Complete case" refers to the complete case on production function variables (sales, materials, capital and labor), missing values in other IC variables—different from production function— are not considered.

[c] "With replacement" refers to the sample after imputing IC variables according to the ICA Method; missing values in other IC variables—different from production function—are not considered. Notice that only observations with information available in at least one of sales, labor, labor cost, materials or capital, are imputed

[d] "Perc. lost" refer to the percentage of observations lost with respect to the sampling frame.

Source: Authors calculations with IC data.

Table 3.1: INDIA, Representativity of sampling frame, complete case and sample with replacement

	Sampling frame[(a)]		Complete case[(b)]		With replacement[(c)]	
	# Obs	Perc over total	# Obs	Perc over total	# Obs	Perc over total
a) by Industry						
Food	597	8.7	307	6.9	454	7.9
Textiles & Leather	888	13	600	13.5	738	12.8
Apparels	825	12	634	14.3	715	12.4
Chemicals & Chemical prds	930	13.6	610	13.7	769	13.4
Plastics & Rubbers	480	7	316	7.1	411	7.1
Non-metallic products	216	3.2	110	2.5	164	2.9
Structural metal & metal prds	909	13.3	480	10.8	771	13.4
Machinery & Equipment	2,004	29.3	1,391	31.3	1,728	30.1
Total	6,849	100	4,448	100	5,750	100
b) by size						
Small	3,621	52.9	2,247	50.5	2,953	51.4
Medium	2,271	33.2	1,537	34.6	1,968	34.2
Large	957	14	664	14.9	829	14.4
Total	6,849	100	4,448	100	5,750	100

Notes:
[(a)] "Sampling frame" refers to the total number of observations (firms surveyed multiplied by the number of years of information).
[(b)] "Complete case" refers to the complete case on production function variables (sales, materials, capital and labor), missing values in other IC variables—different from production function— are not considered.
[(c)] "With replacement" refers to the sample after imputing IC variables according to the ICA Method; missing values in other IC variables—different from production function—are not considered. Notice that only observations with information available in at least one of sales, labor, labor cost, materials or capital, are imputed
Source: Authors calculations with ICSs data.

Table 3.2: TURKEY, Representativity of sampling frame, complete case and sample with replacement

	Sampling frame[(a)]		Complete case[(b)]		With replacement[(c)]	
	# Obs	Perc over total	# Obs	Perc over total	# Obs	Perc over total
a) by Industry						
Food and Bev.	564	21.3	195	22.7	400	22.2
Textiles and Apparels	738	27.9	176	20.5	435	24.1
Chemicals	352	13.3	104	12.1	214	11.9
Non-metallic mineral products	166	6.3	54	6.3	127	7.0
Metal products (ex. M&E)	284	10.7	107	12.5	209	11.6
Machinery and Equipment	256	9.7	106	12.3	194	10.8
Electrical machinery	134	5.1	53	6.2	100	5.5
Transport equipment	152	5.7	64	7.5	125	6.9
Total	2,646	100	859	100.0	1,804	100.0
b) by size						
Small	788	29.8	226	26.3	531	29.4
Medium	810	30.6	264	30.7	557	30.9
Large	1048	39.6	369	43.0	716	39.7
Total	2,646	100.0	859	100.0	1,804	100.0

Notes:
Same of Table 3.1.

Table 3.3: SOUTH AFRICA, Representativity of sampling frame, complete case and sample with replacement

	Sampling frame[(a)]		Complete case[(b)]		With replacement[(c)]	
	# Obs	Perc over total	# Obs	Perc over total	# Obs	Perc over total
a) by Industry						
Food & beverages	189	10.9	131	10.7	159	10.7
Texts & apparels	180	10.4	107	8.7	143	9.6
Chemicals rubber & plastics	285	16.4	187	15.2	241	16.2
Paper, edition & publishing	159	9.2	120	9.8	137	9.2
Machinery & equipment	561	32.3	435	35.4	497	33.4
Wood & furniture	147	8.5	102	8.3	131	8.8
Non-metallic products	66	3.8	43	3.5	49	3.3
Other	150	8.6	104	8.5	129	8.7
Total	1,737	100	1,229	100	1,486	100
b) by size						
Small	189	11	104	8.5	139	9.4
Medium	743	43.4	534	43.5	647	43.7
Large	780	45.6	590	48	696	47
Total	1,712	100	1,228	100	1,482	100

Notes:
Same of Table 3.1.

Table 3.4: TANZANIA, Representativity of sampling frame, complete case and sample with replacement

	Sampling frame[(a)]		Complete case[(b)]		With replacement[(c)]	
	# Obs	Perc over total	# Obs	Perc over total	# Obs	Perc over total
a) by Industry						
Food & beverages	243	29.3	108	33.2	170	30.3
Textiles & apparels	93	11.2	29	8.9	58	10.3
Chemicals, rubber & plastics	102	12.3	42	12.9	69	12.3
Paper, edition & publishing	75	9.1	33	10.2	55	9.8
Machinery & equipment/Metallic products	87	10.5	26	8	65	11.6
Wood & furniture	195	23.6	68	20.9	117	20.9
Non-metallic products	33	4	19	5.8	27	4.8
Total	828	100	325	100	561	100
b) by size						
Small	381	46	144	44.3	271	48.3
Medium	320	38.6	125	38.5	207	36.9
Large	127	15.3	56	17.2	83	14.8
Total	828	100	325	100	561	100

Notes:
Same of Table 3.1.

Table 4: Number of missing values in production function variables by size

		Small	Medium	Large
	a) INDIA			
Totals by size		3,621	2,271	957
Sales	Number of missing [(a)]	646	257	95
	Perc over totals by size [(b)]	17.8	11.3	9.9
Labor	Number of missing	0	0	0
	Perc over totals by size	0	0	0
Materials	Number of missing	688	278	101
	Perc over totals by size	19	12.2	10.6
Capital	Number of missing	1258	640	245
	Perc over totals by size	34.7	28.2	25.6
	b) TURKEY			
Totals by size		788	810	1048
Sales	Number of missing	335	365	449
	Perc over totals by size	42.5	45.1	42.8
Labor	Number of missing	34	37	46
	Perc over totals by size	4.3	4.6	4.4
Materials	Number of missing	346	396	521
	Perc over totals by size	43.9	48.9	49.7
Capital	Number of missing	462	388	507
	Perc over totals by size	58.6	47.9	48.4
	c) SOUTH AFRICA			
Totals by size		197	783	804
Sales	Number of missing	40	95	76
	Perc over totals by size	20.3	12.1	9.5
Labor	Number of missing	23	54	43
	Perc over totals by size	11.7	6.9	5.3
Materials	Number of missing	53	111	97
	Perc over totals by size	26.9	14.2	12.1
Capital	Number of missing	69	204	154
	Perc over totals by size	35	26.1	19.2
	d) TANZANIA			
Totals by size		361	302	127
Sales	Number of missing	129	121	40
	Perc over totals by size	35.7	40.1	31.5
Labor	Number of missing	28	21	11
	Perc over totals by size	7.8	7	8.7
Materials	Number of missing	114	87	38
	Perc over totals by size	31.6	28.8	29.9
Capital	Number of missing	53	111	97
	Perc over totals by size	14.7	36.8	76.4

Small: less than 20 employees; medium: in between 20 and 100 employees; large: more than 100 employees.
[(a)] Number of missing includes both missing values and outliers in the corresponding variables.
[(b)] Percentage over the total number of observations in each category of size.
Source: Authors calculations with IC data.

Table 5.1: INDIA, Proportion of observations with missing values in production function (PF) variables by key IC determinants

Key IC variables		Proportion of Establishments with: complete information on PF variables	at least one missing value in PF variables
Whole sample		67.2	32.8
1. Generator	Establishments not using own generator	68.6	31.4
	Establishments using own generator	66.3	33.7
2. Power outages	Establishments that do not suffer power outages	61	39
	Establishments suffering power outages	69.4	30.6
3. Water outages	Establishments that do not suffer water outages	66.9	33.1
	Establishments suffering water outages	71.5	28.5
4. E-mail	Establishments that do not use e-mail	62.2	37.8
	Establishments using e-mail	70.6	29.4
5. Web page	Establishments that do not use web page	66.8	33.2
	Establishments using web page	68.3	31.7
6. Informality (I)	Establishments reporting all sales to IRS authorities	76.4	23.6
	Establishments that hide some share of sales to IRS	63.5	36.5
7. Informality (II)	Establishments reporting all workforce to IRS authorities	78.1	21.9
	Establishments that hide some share of workforce to IRS	62	38
8. Corruption (I)	Establishments that do not pay bribes to deal with bureaucracy	63.4	36.6
	Establishments paying bribes to deal with bureaucracy	71.6	28.4
9. Corruption (II)	Establishments that do not pay bribes to obtain contracts with the gov.	64.6	35.4
	Establishments paying bribes to obtain contracts with the government	74.3	25.7
10. Crime	Establishments that do not suffer losses due to crime	67.7	32.3
	Establishments suffering losses due to crime	58.4	41.6
11. Security	Establishments without security expenses	67.1	32.9
	Establishments with security expenses	68.2	31.8
12. Loan	Establishments without access to a loan	67.5	32.5
	Establishments with access to a loan	67.2	32.8
13. Credit line	Establishments without access to a credit line	59.6	40.4
	Establishments with access to a credit line	73.8	26.2
14. Auditory	Establishments with annual statements reviewed by external auditory	49.8	50.2
	Establishments without annual statements reviewed by external auditory	70.4	29.6
15. Innovation (I)	Establishments without ISO certification	67	33
	Establishments with ISO certification	67.8	32.2
16. Innovation (II)	Establishments that do not introduce new products	66.4	33.6
	Establishments introducing new products	68.7	31.3
17. Innovation (III)	Establishments that do not introduce new technologies		
	Establishments introducing new technologies		
18. Training	Establishments that do not provide training	71.4	28.6
	Establishments providing training	65.1	34.9
19. Manager skills	Managers with less than university education	64.9	35.1
	Managers with more than university education	71.5	28.5
20. Exporting activity	Establishments that do not export	68.9	31.1
	Establishments exporting	81.8	18.2
21. FDI inflows	Establishments that do not receive FDI inflows	67.2	32.8
	Establishments receiving FDI inflows	60.7	39.3
22. Incorporated company	Establishments out of a incorporated company	66.8	33.2
	Establishments in a incorporated company	67.9	32.1
23. Holding	Establishments out of a holding		
	Establishments in a holding		
24. Capacity utilization	Establishments that do not use all its capacity	67.2	32.8
	Establishments using all its capacity	68.6	31.4

Within production function variables we include labor (labor cost), capital, sales and materials.
Source: Authors calculations with IC data.

Table 5.2: TURKEY, Proportion of observations with missing values in production function (PF) variables by key IC determinants

Key IC variables		Proportion of Establishments with: complete information on PF variables	at least one missing value in PF variables
Whole sample		52.4	47.6
1. Generator	Establishments not using own generator		
	Establishments using own generator		
2. Power outages	Establishments that do not suffer power outages	41.1	58.9
	Establishments suffering power outages	55.7	44.3
3. Water outages	Establishments that do not suffer water outages	53.7	46.3
	Establishments suffering water outages	44.7	55.3
4. E-mail	Establishments that do not use e-mail	56.0	44.0
	Establishments using e-mail	51.5	48.5
5. Web page	Establishments that do not use web page	51.8	48.2
	Establishments using web page	52.6	47.4
6. Informality (I)	Establishments reporting all sales to IRS authorities	47.1	52.9
	Establishments that hide some share of sales to IRS	55.2	44.8
7. Informality (II)	Establishments reporting all workforce to IRS authorities	47.6	52.4
	Establishments that hide some share of workforce to IRS	57.0	43.0
8. Corruption (I)	Establishments that do not pay bribes to deal with bureaucracy	48.0	52.0
	Establishments paying bribes to deal with bureaucracy	76.2	23.8
9. Corruption (II)	Establishments that do not pay bribes to obtain contracts with the gov	47.7	52.3
	Establishments paying bribes to obtain contracts with the government	63.9	36.1
10. Crime	Establishments that do not suffer losses due to crime	52.2	47.8
	Establishments suffering losses due to crime	54.4	45.6
11. Security	Establishments without security expenses	32.0	68.0
	Establishments with security expenses	93.0	7.0
12. Loan	Establishments without access to a loan	47.6	52.4
	Establishments with access to a loan	56.4	43.6
13. Credit line	Establishments without access to a credit line	45.5	54.5
	Establishments with access to a credit line	60.4	39.6
14. Auditory	Establishments with annual statements reviewed by external auditory	56.2	43.8
	Establishments without annual statements reviewed by external auditory	47.1	52.9
15. Innovation (I)	Establishments without ISO certification	51.0	49.0
	Establishments with ISO certification	54.4	45.6
16. Innovation (II)	Establishments that do not introduce new products	50.6	49.4
	Establishments introducing new products	55.5	44.5
17. Innovation (III)	Establishments that do not introduce new technologies	44.0	56.0
	Establishments introducing new technologies	64.0	36.0
18. Training	Establishments that do not provide training	47.5	52.5
	Establishments providing training	56.6	43.4
19. Manager skills	Managers with less than university education	52.0	48.0
	Managers with more than university education	53.9	46.1
20. Exporting activity	Establishments that do not export	54.3	45.7
	Establishments exporting	50.2	49.8
21. FDI inflows	Establishments that do not receive FDI inflows	52.8	47.2
	Establishments receiving FDI inflows	43.1	56.9
22. Incorporated company	Establishments out of a incorporated company	51.9	48.1
	Establishments in a incorporated company	62.1	37.9
23. Holding	Establishments out of a holding	53.1	46.9
	Establishments in a holding	42.5	57.5
24. Capacity utilization	Establishments that do not use all its capacity	55.5	44.5
	Establishments using all its capacity	38.0	62.0

Within production function variables we include labor (labor cost), capital, sales and materials.
Source: Authors calculations with IC data.

Table 5.3: SOUTH AFRICA, Proportion of observations with missing values in production function (PF) variables by key IC determinants

Key IC variables		Proportion of Establishments with: complete information on PF variables	at least one missing value in PF variables
Whole sample		72	28
1. Generator	Establishments not using own generator	71.8	28.2
	Establishments using own generator	73.3	26.7
2. Power outages	Establishments that do not suffer power outages	63.4	36.6
	Establishments suffering power outages	76.6	23.4
3. Water outages	Establishments that do not suffer water outages	64.9	35.1
	Establishments suffering water outages	89.4	10.6
4. E-mail	Establishments that do not use e-mail	33.3	66.7
	Establishments using e-mail	72.5	27.5
5. Web page	Establishments that do not use web page	71.9	28.1
	Establishments using web page	72.0	28.0
6. Informality (I)	Establishments reporting all sales to IRS authorities	59.3	40.7
	Establishments that hide some share of sales to IRS	74.3	25.7
7. Informality (II)	Establishments reporting all workforce to IRS authorities		
	Establishments that hide some share of workforce to IRS		
8. Corruption (I)	Establishments that do not pay bribes to deal with bureaucracy	73.4	26.6
	Establishments paying bribes to deal with bureaucracy	33.3	66.7
9. Corruption (II)	Establishments that do not pay bribes to obtain contracts with the gov	73.7	26.3
	Establishments paying bribes to obtain contracts with the government	40.0	60.0
10. Crime	Establishments that do not suffer losses due to crime	70.9	29.1
	Establishments suffering losses due to crime	72.9	27.1
11. Security	Establishments without security expenses	62.1	37.9
	Establishments with security expenses	74.5	25.5
12. Loan	Establishments without access to a loan	73.9	26.1
	Establishments with access to a loan	68.8	31.2
13. Credit line	Establishments without access to a credit line	72.6	27.4
	Establishments with access to a credit line	71.6	28.4
14. Auditory	Establishments with annual statements reviewed by external auditory	38.9	61.1
	Establishments without annual statements reviewed by external auditory	73.0	27.0
15. Innovation (I)	Establishments without ISO certification	70.9	29.1
	Establishments with ISO certification	73.6	26.4
16. Innovation (II)	Establishments that do not introduce new products	62.4	37.6
	Establishments introducing new products	76.4	23.6
17. Innovation (III)	Establishments that do not introduce new technologies	67.6	32.4
	Establishments introducing new technologies	74.9	25.1
18. Training	Establishments that do not provide training	73.5	26.5
	Establishments providing training	71.1	28.9
19. Manager skills	Managers with less than university education	63.7	36.3
	Managers with more than university education	75.3	24.7
20. Exporting activity	Establishments that do not export	69.8	30.2
	Establishments exporting	75.4	24.6
21. FDI inflows	Establishments that do not receive FDI inflows	71.9	28.1
	Establishments receiving FDI inflows	72.4	27.6
22. Incorporated company	Establishments out of a incorporated company	72.9	27.1
	Establishments in a incorporated company	51.4	48.6
23. Holding	Establishments out of a holding	72.4	27.6
	Establishments in a holding	69.0	31.0
24. Capacity utilization	Establishments that do not use all its capacity	72.8	27.2
	Establishments using all its capacity	66.7	33.3

Within production function variables we include labor (labor cost), capital, sales and materials.
Source: Authors calculations with IC data.

Table 5.4: TANZANIA, Proportion of observations with missing values in production function (PF) variables by key IC determinants

Key IC variables		Proportion of Establishments with: complete information on PF variables	at least one missing value in PF variables
Whole sample		44.8	55.2
1. Generator	Establishments not using own generator	44.9	55.1
	Establishments using own generator	45.1	54.9
2. Power outages	Establishments that do not suffer power outages	42.1	57.9
	Establishments suffering power outages	45.7	54.3
3. Water outages	Establishments that do not suffer water outages	42.5	57.5
	Establishments suffering water outages	50.4	49.6
4. E-mail	Establishments that do not use e-mail	43.2	56.8
	Establishments using e-mail	46.4	53.6
5. Web page	Establishments that do not use web page	43.4	56.6
	Establishments using web page	50.0	50.0
6. Informality (I)	Establishments reporting all sales to IRS authorities	45.6	54.4
	Establishments that hide some share of sales to IRS	44.3	55.7
7. Informality (II)	Establishments reporting all workforce to IRS authorities		
	Establishments that hide some share of workforce to IRS		
8. Corruption (I)	Establishments that do not pay bribes to deal with bureaucracy	41.3	58.7
	Establishments paying bribes to deal with bureaucracy	50.0	50.0
9. Corruption (II)	Establishments that do not pay bribes to obtain contracts with the gov	42.8	57.2
	Establishments paying bribes to obtain contracts with the government	54.2	45.8
10. Crime	Establishments that do not suffer losses due to crime	58.9	41.1
	Establishments suffering losses due to crime	0.0	0.0
11. Security	Establishments without security expenses	45.0	55.0
	Establishments with security expenses	47.6	52.4
12. Loan	Establishments without access to a loan	51.8	48.2
	Establishments with access to a loan	61.0	39.0
13. Credit line	Establishments without access to a credit line	42.1	57.9
	Establishments with access to a credit line	50.2	49.8
14. Auditory	Establishments with annual statements reviewed by external auditory	32.7	67.3
	Establishments without annual statements reviewed by external auditory	48.9	51.1
15. Innovation (I)	Establishments without ISO certification	43.4	56.6
	Establishments with ISO certification	57.6	42.4
16. Innovation (II)	Establishments that do not introduce new products	44.9	55.1
	Establishments introducing new products	47.0	53.0
17. Innovation (III)	Establishments that do not introduce new technologies	48.3	51.7
	Establishments introducing new technologies	39.9	60.1
18. Training	Establishments that do not provide training	44.5	55.5
	Establishments providing training	47.9	52.1
19. Manager skills	Managers with less than university education		
	Managers with more than university education		
20. Exporting activity	Establishments that do not export	44.6	55.4
	Establishments exporting	51.6	48.4
21. FDI inflows	Establishments that do not receive FDI inflows	43.9	56.1
	Establishments receiving FDI inflows	47.5	52.5
22. Incorporated company	Establishments out of a incorporated company	45.1	54.9
	Establishments in a incorporated company	38.1	61.9
23. Holding	Establishments out of a holding	46.4	53.6
	Establishments in a holding	33.3	66.7
24. Capacity utilization	Establishments that do not use all its capacity	45.5	54.5
	Establishments using all its capacity	36.1	63.9

Within production function variables we include labor (labor cost), capital, sales and materials.
Source: Authors calculations with IC data.

Table 6.1: INDIA, Linear probability models for the probability of observing TFP and sales

Dependent variables:	Missing on TFP (a)				Missing on sales (b)			
	[1]		[2]		[3]		[4]	
Explanatory variables:	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
Infrastructures:								
Longest days to clear customs to export (a)	-0.0279	[0.0112]**			-0.0108	[0.0082]		
Dummy for own generator	-0.0066	[0.0165]			0.0072	[0.0134]		
Water supply from public sources (b)	0.0001	[0.0002]			0.0000	[0.0002]		
Shipment losses in the domestic market (b)	-0.0044	[0.0015]***			-0.0028	[0.0014]**		
Dummy for own transport	-0.0083	[0.0208]			0.0122	[0.0199]		
Dummy for web page	0.0153	[0.0177]			0.0191	[0.0207]		
Losses due to power outages (b)			-0.0023	[0.0010]**			-0.0025	[0.0007]***
Dummy for e-mail (b)			0.0282	[0.0166]*			0.031	[0.0183]*
Shipment losses, domestic (b)			-0.0043	[0.0014]			-0.0028	[0.0011]**
Losses due to transport outages (b)			-0.0033	[0.0018]***			-0.0035	[0.0015]**
Red tape, corruption and crime:								
Dummy for security	0.0146	[0.0188]			0.0033	[0.0157]		
Sales reported to taxes (b)	0.0006	[0.0006]			0.0005	[0.0005]		
Workforce reported to taxes (b)	-0.0004	[0.0004]			-0.0001	[0.0004]		
Dummy for payments to speed up bureaucracy	0.0347	[0.0137]**			0.0359	[0.0122]***		
Dummy for interventionist labor regulation	-0.0327	[0.0180]*	-0.0379	[0.0187]**	-0.0383	[0.0185]**	-0.0409	[0.0189]**
Absenteeism (b)	-0.0165	[0.0074]**			-0.0122	[0.0057]**		
Dummy for payments to deal with bur. Issues (b)			0.0222	[0.0140]			0.0261	[0.0136]*
Finance:								
Dummy for external audit	0.0121	[0.0174]	0.0538	[0.0252]**	0.0086	[0.0140]	0.0423	[0.0161]***
Dummy for trade association	-0.0002	[0.0002]			0.0003	[0.0002]		
Working capital financed by domestic private banks (b)	0.0234	[0.0146]			0.0231	[0.0134]*		
Dummy for loan (b)	0.0337	[0.0209]			0.0319	[0.0159]**		
Largest shareholder (b)			-0.0003	[0.0002]			-0.0004	[0.0002]**
Dummy for loan with collateral (b)			-0.0802	[0.0318]**			-0.0573	[0.0252]**
Borrows denominated in foreign currency (b)			-0.0011	[0.0003]***			-0.0008	[0.0003]***
Quality, innovation and labor skills:								
Dummy for R&D (a)	0.0016	[0.1084]	0.0153	[0.0147]	-0.04	[0.0666]	0.0296	[0.0130]**
Dummy for product innovation	-0.0073	[0.0157]			-0.0099	[0.0133]		
Dummy for foreign license (b)	0.0481	[0.0314]			0.0572	[0.0297]*		
Dummy for internal training (b)	0.0025	[0.0197]			0.0001	[0.0186]		
Unskilled workforce (a)	0.0021	[0.0012]*			0.0017	[0.0011]		
Workforce with computer	0.0006	[0.0004]			0.0001	[0.0003]		
Dummy for ISO quality certification (b)			0.0148	[0.0173]			0.0325	[0.0156]***
Dummy for outsourcing (b)			0.0457	[0.0174]			0.0213	[0.0135]
Dummy for external training (b)			-0.0334	[0.0235]			-0.0256	[0.0164]
Other control variables:								
Dummy for incorporated company	0.0185	[0.0146]			0.0308	[0.0139]**		
Age	0.0077	[0.0103]			0.0097	[0.0095]		
Share of exports (b)	0.0002	[0.0002]			0.0002	[0.0002]		
Trade union (b)	0.0007	[0.0004]*	0.0008	[0.0003]	0.0006	[0.0003]*	0.0008	[0.0003]***
Strikes (b)	-0.0165	[0.0133]			-0.0037	[0.0158]		
Constant	*Yes*		*Yes*		*Yes*		*Yes*	
Industry/region/size dummies	*Yes*		*Yes*		*Yes*		*Yes*	
Observations	2048		2277		2048		2277	
R-squared	0.23		0.23		0.18		0.18	

(a) Missing in TFP takes value 1 if we observe all sales, materials, labor and capital, and 0 otherwise.
(b) Missing in TFP takes value 1 if we observe sales, and 0 otherwise.
[1] Model of the probability of observing a missing value in TFP on the IC and C variables significant in equation (1).
[2] Model of the probability of observing a missing value in TFP and the matrices IC* and C*, selected from the whole set of IC and C variables.
[1] Model of the probability of observing a missing value in sales on the IC and C variables significant in equation (1).
[2] Model of the probability of observing a missing value in sales and the matrices IC* and C*, selected from the whole set of IC and C variables.
Significance given by robust standard errors allowing for clustering by industry and region *** 1%, **5%, * 10%.
Source: Authors' estimations with ICSs data.

Table 6.2: TURKEY, Linear probability models for the probability of observing TFP and sales

Dependent variables:	Missing on TFP		Missing on sales	
	[1]	[2]	[3]	[4]
Explanatory variables:	Coeff. Std. Err.	Coeff. Std. Err.	Coeff. Std. Err.	Coeff. Std. Err.
Infrastructures:				
Days to clear customs to imports (a)	0.019 [0.0592]		0.0189 [0.0669]	
Losses due to power outages (b)				-0.0029 [0.0016]*
Losses due to water outages (b)				0.0035 [0.0010]***
Shipment losses (b)				-0.0038 [0.0017]**
Dummy foe e-mail (b)	0.021 [0.0341]		0.0811 [0.0378]**	0.1088 [0.0377]***
Electricity from generator (b)		0.0009 [0.0004]**		
Red tape, corruption and crime:				
Crime losses (b)		0.0024 [0.0005]***		0.0035 [0.0004]***
Security expenses (b)	0.1273 [0.0350]***		0.1322 [0.0403]***	
Manager's time spent in bur. Issues (b)		-0.003 [0.0009]***		-0.0025 [0.0012]**
Dummy for consultant to help deal with bur. Issues		-0.0693 [0.0175]***		-0.0713 [0.0270]**
Number of inspections (B)	-0.0036 [0.0022]	-0.0221 [0.0129]*	0.0003 [0.0002]	
Payments to deal with bureaucratic issues (a)	0.00001 [0.0002]	0.0013 [0.0004]***	0.0092 [0.0037]**	0.0019 [0.0004]***
Sales declared to taxes (a)	0.0087 [0.0035]**	-0.0011 [0.0004]**	-0.003 [0.0022]	-0.0013 [0.0004]***
Payments to obtain a contract with the government (b)	-0.0309 [0.0132]**	-0.0156 [0.0022]***	-0.0276 [0.0170]	-0.0136 [0.0028]***
Production lost due to absenteeism (b)	-0.0149 [0.0024]***		-0.0136 [0.0027]***	
Dummy for informal competition (b)	-0.0332 [0.0177]*		-0.0368 [0.0176]**	
Delay to obtain a water supply (a)	-0.0282 [0.0214]		-0.033 [0.0238]	
Dummy for lawsuit (b)		-0.0494 [0.0218]**		-0.0728 [0.0293]**
Finance:				
Dummy for credit line	-0.0763 [0.0243]***		-0.0908 [0.0247]***	-0.0778 [0.0232]***
Dummy for external auditory (a)	0.0443 [0.0194]**	-0.0548 [0.0234]**	0.0327 [0.0230]	
Borrows in foreign currency (b)	-0.0005 [0.0003]*		-0.0006 [0.0005]	
Dummy for new land purchased		-0.0528 [0.0313]*		
Dummy for loan denominated in Turkish Lira (b)		-0.1216 [0.0238]***		-0.1645 [0.0238]***
Dummy for loan denominated in foreign currency (b)		-0.1001 [0.0317]***		-0.1472 [0.0379]***
Dummy for long-term loan (b)				0.1261 [0.0356]***
Quality, innovation and labor skills:				
Dummy for ISO quality certification (b)		0.0869 [0.0192]***		0.0696 [0.0206]***
Dummy for new technology (b)		-0.1027 [0.0223]***		-0.0987 [0.0260]***
Dummy for foreign licensed technology (b)				0.0607 [0.0244]**
Staff with university education (b)	0.0001 [0.0010]	0.0016 [0.0007]**	0.001 [0.0010]	
Staff-part time workers	0.0018 [0.0007]**		0.0014 [0.0009]	0.0012 [0.0008]
Other control variables:				
Dummy for incorporated company		-0.092 [0.0557]		-0.0851 [0.0394]**
Age				-0.0457 [0.0220]**
Market share				0.0008 [0.0007]
Production lost due to strikes (b)	-0.0408 [0.0180]**		-0.0056 [0.0246]	
Dummy for recently privatized firm	0.0222 [0.0949]		-0.0344 [0.0877]	
Dummy for competition against imported products	-0.0472 [0.0441]		-0.0261 [0.0393]	
Constant	Yes	Yes	Yes	Yes
Industry/region/size dummies	Yes	Yes	Yes	Yes
Observations	1323	1323	1323	1323
R-squared	0.2	0.31	0.24	0.3

See footnotes in Table 6.1.
Source: Authors' estimations with ICSs data.

Table 6.3: SOUTH AFRICA, Linear probability models for the probability of observing TFP and sales

Dependent variables:	Missing on TFP	Missing on sales		
	[1]	[2]	[3]	[4]
Explanatory variables:	Coeff. Std. Err.	Coeff. Std. Err.	Coeff. Std. Err.	Coeff. Std. Err.
Infrastructures:				
Days to clear customs to import (a)	-0.018 [0.0587]		-0.0782 [0.0509]	
Sales lost due to power outages (b)	-0.0061 [0.0044]	-0.0068 [0.0036]*	-0.0059 [0.0026]**	-0.0051 [0.0022]**
Water outages (b)	0.0166 [0.0231]	0.016 [0.0032]***	0.0021 [0.0196]	
Average duration of transport failures (a)	-0.0206 [0.0467]		0.0064 [0.0445]	
Wait for electric supply (a)	0.0193 [0.0313]		0.0202 [0.0342]	
Dummy for email (b)		0.1795 [0.0686]**		
Dummy for internet				0.0356 [0.0138]**
Sales lost due to delivery delays (b)	0.0103 [0.0040]**	0.0115 [0.0034]***	0.003 [0.0028]	0.0039 [0.0027]
Red tape, corruption and crime:				
Manager's time spent in bur. issues (b)	0.0022 [0.0010]**		0.001 [0.0007]	
Payments to deal with bureaucratic issues (b)	-0.0011 [0.0007]	-0.0015 [0.0005]***	-0.0011 [0.0008]	
Sales declared to taxes (a)	0.0006 [0.0028]		-0.0022 [0.0029]	-0.0027 [0.0016]*
Payments to obtain a contract with the gov. (b)	0.0119 [0.0078]		0.0199 [0.0093]**	0.0199 [0.0079]**
Security expenses (a)	0.0033 [0.0102]	0.0078 [0.0024]***	0.0084 [0.0082]	
Crime losses (a)				0.0241 [0.0201]
Illegal payments in protection (b)	-0.0324 [0.0595]		-0.0003 [0.0424]	
Crime losses (a)	0.023 [0.0404]		0.0472 [0.0368]	
Finance:				
Percentage of credit unused (b)	0.0002 [0.0003]		0.0004 [0.0003]	0.0004 [0.0002]*
Dummy for loan	-0.0025 [0.0329]		0.0017 [0.0213]	
Dummy for credit line (b)				-0.0193 [0.0143]
Value of the collateral (b)	0.00001 [0.0002]		-0.0001 [0.0001]	
Borrows in foreign currency (b)	0.0002 [0.0008]		-0.0005 [0.0004]	-0.0006 [0.0003]*
Charge to clear a check (a)	-0.0094 [0.0279]		-0.037 [0.0252]	-0.0307 [0.0162]*
Largest shareholder	0.0002 [0.0004]		0.0003 [0.0004]	
Working capital finn. by foreign commercial banks (b)	0.003 [0.0026]		0.0046 [0.0026]*	0.0045 [0.0026]*
Working capital financed by informal sources (b)	0.0011 [0.0008]		0.0002 [0.0003]	
Dummy for external auditory (b)		-0.1669 [0.0911]*		-0.1817 [0.0812]**
Quality, innovation and labor skills:				
Dummy for ISO quality certification (b)	0.0375 [0.0258]		0.0304 [0.0175]*	0.036 [0.0180]*
Dummy for new product (b)	-0.0234 [0.0310]		0.007 [0.0205]	
Dummy for discontinued product line (b)	-0.0316 [0.0264]		-0.0185 [0.0143]	
Dummy for outsourcing (b)		-0.0421 [0.0192]**		-0.0267 [0.0138]*
Staff - management	0.0009 [0.0012]		0.0013 [0.0010]	
Staff - non-production workers	-0.0009 [0.0007]		-0.0008 [0.0006]	
Dummy for training (b)				-0.0231 [0.0146]
Training to unskilled workers (a)	0.0015 [0.0023]		0.00001 [0.0020]	
University staff (b)	-0.0007 [0.0007]		-0.0012 [0.0005]**	-0.0013 [0.0005]**
Manager's experience (b)	0.002 [0.0102]		-0.0063 [0.0073]	
Dummy for closed plant		-0.0463 [0.0210]**		
Other control variables:				
Age (b)	-0.0004 [0.0005]		-0.0002 [0.0003]	
Share of the local market (b)	0.0002 [0.0004]		0.0002 [0.0003]	
Capacity utilization (b)		-0.0018 [0.0009]**		
Constant	Yes	Yes	Yes	Yes
Industry/region/size dummies	Yes	Yes	Yes	Yes
Observations	586	594	586	594
R-squared	0.22	0.25	0.24	0.24

See footnotes in Table 6.1.
Source: Authors' estimations with ICSs data.

Table 6.4: TANZANIA, Linear probability models for the probability of observing TFP and sales

Dependent variables:	Missing on TFP		Missing on sales	
	[1]	[2]	[3]	[4]
Explanatory variables:	Coeff. Std. Err.	Coeff. Std. Err.	Coeff. Std. Err.	Coeff. Std. Err.
Infrastructures:				
Electricity from own generator (b)	-0.0007 [0.0014]		-0.0007 [0.0015]	
Losses due to power outages (b)	0.0035 [0.0050]	0.0049 [0.0023]**	0.0021 [0.0031]	
Losses due to water outages (b)				
Water from own well or water infrastructure (a)	0.00001 [0.0030]		0.001 [0.0023]	
Losses due to phone outages (a)	-0.0308 [0.0158]*		-0.0219 [0.0157]	
Transport outages (a)	-0.0125 [0.0349]		-0.0406 [0.0264]	
Losses due to transport delay (b)				-0.0067 [0.0020]***
Dummy for own roads (b)	-0.1213 [0.0768]		-0.0904 [0.0977]	
Dummy for webpage (b)	0.061 [0.0795]		0.0322 [0.0775]	
Wait for a water supply (a)	0.0192 [0.0249]		-0.0178 [0.0271]	
Low quality supplies (a)	-0.0035 [0.0109]		-0.0053 [0.0087]	-0.0025 [0.0013]*
Days of inventory of main supply				0.0358 [0.0175]**
Red tape, corruption and crime:				
Gift to obtain a operating license (b)	-0.0519 [0.0754]		-0.0152 [0.1104]	
Payments to deal with bureaucratic issues (b)	-0.0592 [0.0227]**	-0.0803 [0.0147]***	-0.045 [0.0267]	-0.0648 [0.0151]***
Days in inspections (b)	-0.0509 [0.0378]	-0.0788 [0.0403]*	-0.0241 [0.0387]	
Payments to obtain a contract with the gov. (b)	-0.0092 [0.0039]**	-0.0117 [0.0034]***	-0.0063 [0.0046]	-0.01 [0.0040]**
Security expenses (b)	-0.0023 [0.0026]		-0.0035 [0.0028]	
Illegal payments in protection (b)	-0.0075 [0.0224]		-0.0385 [0.0072]***	-0.0405 [0.0095]***
Finance:				
Dummy for credit line (b)				-0.1182 [0.0657]*
Interest rate of the loan (a)	0.0033 [0.0076]		-0.0017 [0.0061]	
Borrows denominated in foreign currency (b)				-0.0014 [0.0009]
Dummy for current or saving account (b)		0.1616 [0.0856]*		0.2347 [0.0706]***
Working capital financed by commercial banks (b)	-0.0009 [0.0007]		-0.0011 [0.0010]	
Working capital financed by leasing (b)	-0.0059 [0.0023]**		-0.0059 [0.0013]***	
Inputs bought on credit (b)		-0.0016 [0.0008]*		
Sales bought on credit (b)	0.0007 [0.0012]		0.0007 [0.0011]	
Delay to clear a domestic currency wire (a)	0.2385 [0.1403]*		0.196 [0.1479]	
Quality, innovation and labor skills:				
Dummy for new product (b)	0.0087 [0.0501]		0.002 [0.0462]	
Dummy for foreign license (b)				-0.2748 [0.0649]***
Dummy for upgraded product (b)				-0.1705 [0.0752]**
Dummy for new technology (b)		0.1973 [0.0631]***		0.3095 [0.0721]***
Dummy for joint venture (b)		-0.2179 [0.0796]**		
Dummy for outsourcing (b)		-0.2066 [0.0960]**		
Dummy for brought in house (b)		-0.2265 [0.0707]***		
Staff - skilled workers (b)	0.0007 [0.0004]*		0.0009 [0.0003]***	
Staff - professional workers (b)		-0.0055 [0.0033]		-0.0075 [0.0040]*
Workforce with computer (b)	0.003 [0.0017]*	0.0055 [0.0017]***	-0.0007 [0.0014]	0.0026 [0.0015]*
Dummy for training (b)				-0.0954 [0.0596]
Other control variables:				
Dummy for incorporated company (b)	0.012 [0.1990]		-0.075 [0.1534]	
Dummy for FDI (b)	0.1112 [0.0636]*	0.1255 [0.0549]**	0.1049 [0.0618]*	0.1717 [0.0586]***
Dummy for industrial zone (b)		0.121 [0.0737]		0.1274 [0.0668]*
Constant	Yes	Yes	Yes	Yes
Industry/region/size dummies	Yes	Yes	Yes	Yes
Observations	262	262	262	262
R-squared	0.18	0.22	0.16	0.3

See footnotes in Table 6.1.

Source: Authors' estimations with ICSs data.

Table 7: Linear probability models for the effect of TFP and sales on the probability of observing a missing value in t+1

A. Missing in TFP [1]				
Dependent variables: for each country a dummy taking value 1 if we observe all labor, materials, capital and sales				
Explanatory variables	India	Turkey	South Africa	Tanzania
log TFP (t+1)	0.0168*	0.0183**	0.0212	0.0281
	[0.0091]	[0.0084]	[0.0180]	[0.0250]
IC variables [3]	Yes	Yes	Yes	Yes
Constant	Yes	Yes	Yes	Yes
Industry/region/size dummies	Yes	Yes	Yes	Yes
Observations	1476	426	454	87
R-squared	0.27	0.07	0.19	0.32

B. Missing in sales [2]				
Dependent variables: for each country a dummy taking value 1 if we observe sales				
Explanatory variables	India	Turkey	South Africa	Tanzania
log sales (t+1)	0.0063*	0.0069	0.0079	0.0033
	[0.0033]	[0.0043]	[0.0083]	[0.0144]
IC variables [3]	Yes	Yes	Yes	Yes
Constant	Yes	Yes	Yes	Yes
Industry/region/size dummies	Yes	Yes	Yes	Yes
Observations	1894	677	564	155
R-squared	0.17	0.05	0.16	0.14

[1] Missing in TFP takes value 1 if we do observe all of sales, materials, labor and capital, and 0 otherwise.
[2] Missing in sales takes value 1 if we do observe sales and 0 otherwise.
[3] The set of IC variables of equation (1) is also included.
Both TFP and sales are used before imputing missing values.
Significance given by robust standard errors allowing for clustering by industry and region *** 1%, **5%, * 10%.
Source: Authors' estimations with ICSs data.

Table 8.1 INDIA, Descriptive of production function variables under different imputation mechanism

	Variable	#Obss. (#imputed)	Mean	Std. Dev.	Min	Max
Sales	Complete case	5841.00	12.08	2.30	1.30	22.79
	ICA method	5935 (94)	12.07	2.29	1.30	22.79
	Random ICA meth.	5935 (94)	12.13	2.32	1.30	22.79
	EM alg. [1]	6848 (1007)	12.02	2.19	1.30	22.79
	EM alg. [2]	5882 (41)	12.08	2.30	1.30	22.79
	EM alg. [3]	5882 (41)	12.08	2.30	1.30	22.79
Materials	Complete case	5597.00	11.44	2.30	2.94	22.20
	ICA method	5933 (336)	11.40	2.28	2.94	22.20
	Random ICA meth.	5933 (336)	11.57	2.35	2.94	22.20
	EM alg. [1]	6848 (1251)	11.35	2.17	2.94	22.20
	EM alg. [2]	5906 (309)	11.42	2.32	2.94	22.20
	EM alg. [3]	5906 (336)	11.42	2.32	2.94	22.20
Capital	Complete case	4555.00	10.31	2.11	1.85	20.73
	ICA method	5918 (1363)	10.28	2.10	1.85	20.73
	Random ICA meth.	5918 (1363)	11.20	2.47	1.85	20.73
	EM alg. [1]	6848 (2293)	10.26	1.89	1.85	20.73
	EM alg. [2]	5807 (1252)	10.25	2.04	1.85	20.73
	EM alg. [3]	5807 (1252)	10.23	2.02	1.85	20.73
Employment	Complete case	6164.00	10.82	1.33	6.54	16.16
	ICA method	6321 (157)	10.82	1.34	6.54	16.16
	Random ICA meth.	6321 (157)	10.84	1.34	6.54	16.16
	EM alg. [1]	6849 (687)	10.78	1.31	6.54	16.16
	EM alg. [2]	6164 (0)	10.82	1.33	6.54	16.16
	EM alg. [3]	6164 (0)	10.82	1.33	6.54	16.16

Source: Authors' estimations with ICSs data.

Table 8.2 TURKEY, Descriptive statistics of production function variables under different imputation mechanism

	Variable	#Obss. (#imputed)	Mean	Std. Dev.	Min	Max
Sales	Complete case	1497	14.24	2.10	7.78	19.40
	ICA method	1821 (324)	14.30	1.99	7.78	19.40
	Random ICA meth.	1821 (324)	14.44	1.97	7.78	19.40
	EM alg. [1]	2646 (1149)	14.27	1.78	7.78	19.40
	EM alg. [2]	1808 (311)	14.22	2.02	7.55	19.40
	EM alg. [3]	1808 (311)	14.22	2.01	7.78	19.40
Materials	Complete case	1293	13.19	2.31	4.33	18.65
	ICA method	1822 (529)	13.37	2.13	4.34	18.65
	Random ICA meth.	1822 (529)	13.59	2.12	4.34	18.65
	EM alg. [1]	2646 (1353)	13.31	1.86	4.33	18.65
	EM alg. [2]	1802 (509)	13.18	2.18	4.33	18.65
	EM alg. [3]	1802 (509)	13.15	2.18	4.33	18.65
Capital	Complete case	1289	11.39	2.26	0.63	19.65
	ICA method	1816 (527)	11.32	2.05	1.05	19.65
	Random ICA meth.	1816 (527)	11.86	2.05	1.05	19.65
	EM alg. [1]	2646 (1357)	11.22	1.79	0.63	19.65
	EM alg. [2]	1807 (518)	11.28	2.05	0.63	19.65
	EM alg. [3]	1807 (518)	11.30	2.04	0.63	19.65
Employment	Complete case	2529	11.63	1.45	7.64	15.42
	ICA method	2548 (19)	11.63	1.45	7.64	15.42
	Random ICA meth.	2548 (19)	11.63	1.44	7.64	15.42
	EM alg. [1]	2646 (117)	11.63	1.44	7.64	15.42
	EM alg. [2]	2539 (10)	11.63	1.45	7.64	15.42
	EM alg. [3]	2539 (10)	11.63	1.45	7.64	15.42

Source: Authors' estimations with ICSs data.

Table 8.3 SOUTH AFRICA, Descriptive statistics of production function variables under different imputation mechanism

	Variable	#Obss. (#imputed)	Mean	Std. Dev.	Min	Max
Sales	Complete case	1578	17.43	1.86	8.28	24.29
	ICA method	1587 (9)	17.44	1.87	8.28	24.29
	Random ICA meth.	1587 (9)	17.44	1.87	8.28	24.29
	EM alg. [1]	1789 (211)	17.42	1.81	8.28	24.29
	EM alg. [2]	1587 (9)	17.44	1.87	8.28	24.29
	EM alg. [3]	1587 (9)	17.44	1.87	8.28	24.29
Materials	Complete case	1508	16.59	2.03	3.56	24.21
	ICA method	1587 (79)	16.60	2.00	3.56	24.21
	Random ICA meth.	1587 (79)	16.66	2.01	3.56	24.21
	EM alg. [1]	1789 (281)	16.58	1.93	3.56	24.21
	EM alg. [2]	1586 (78)	16.59	2.08	3.56	24.21
	EM alg. [3]	1586 (78)	16.59	2.08	3.56	24.21
Capital	Complete case	1337	15.29	1.89	7.90	23.48
	ICA method	1586 (249)	15.25	1.86	7.90	23.48
	Random ICA meth.	1586 (249)	15.60	1.90	7.90	23.48
	EM alg. [1]	1786 (449)	15.24	1.75	7.90	23.48
	EM alg. [2]	1583 (246)	15.20	1.84	7.90	23.48
	EM alg. [3]	1580 (243)	15.22	1.87	7.90	23.48
Employment	Complete case	1664	12.12	1.40	5.19	17.47
	ICA method	1685 (21)	12.12	1.40	5.19	17.47
	Random ICA meth.	1685 (21)	12.13	1.40	5.19	17.47
	EM alg. [1]	1784 (120)	12.10	1.40	5.19	17.47
	EM alg. [2]	1680 (16)	12.13	1.40	5.19	17.47
	EM alg. [3]	1680 (16)	12.13	1.40	5.19	17.47

The null hypothesis of the one-sample Kolmogorov-Smirnov Test is that the cumulative distribution differs from the hypothesized theoretical normal distribution.
Source: Authors' estimations with ICSs data.

Table 8.4 TANZANIA, Descriptive statistics of production function variables under different imputation mechanism

	Variable	#Obss. (#imputed)	Mean	Std. Dev.	Min	Max
Sales	Complete case	511	14.52	2.43	7.54	20.73
	ICA method	667 (156)	14.60	2.30	7.54	20.73
	Random ICA meth.	667 (156)	14.85	2.25	7.54	20.73
	EM alg. [1]	801 (290)	14.51	2.18	7.54	20.73
	EM alg. [2]	647 (136)	14.48	2.42	7.54	20.73
	EM alg. [3]	647 (136)	14.48	2.41	7.54	20.73
Materials	Complete case	539	13.76	2.58	4.78	20.07
	ICA method	667 (128)	13.82	2.52	4.78	20.07
	Random ICA meth.	667 (128)	14.08	2.49	4.78	20.07
	EM alg. [1]	803 (264)	13.74	2.32	4.78	20.07
	EM alg. [2]	646 (107)	13.67	2.58	4.78	20.07
	EM alg. [3]	646 (107)	13.67	2.57	4.78	20.07
Capital	Complete case	529	13.59	2.69	6.86	19.54
	ICA method	664 (135)	13.54	2.57	6.86	19.54
	Random ICA meth.	664 (135)	13.91	2.51	6.86	19.54
	EM alg. [1]	806 (277)	13.46	2.40	6.86	19.54
	EM alg. [2]	654 (125)	13.26	2.74	6.86	19.54
	EM alg. [3]	654 (125)	13.26	2.81	5.47	19.54
Employment	Complete case	730	10.92	1.37	7.50	15.23
	ICA method	788 (58)	10.91	1.34	7.50	15.23
	Random ICA meth.	788 (58)	10.94	1.34	7.50	15.23
	EM alg. [1]	790 (60)	10.92	1.36	7.50	15.23
	EM alg. [2]	758 (28)	10.92	1.36	7.50	15.23
	EM alg. [3]	768 (38)	10.92	1.36	7.50	15.23

Source: Authors' estimations with ICSs data.

Table 9.1: INDIA, Extended production function and comparison of ICA method with EM algorithms

Dependent variable: Log of total sales		ICA Method[1]		EM Algorithms[2]		
				[1]	[2]	[3]
Category	**Variable**	**Coeff. std. err.**	**Boot. s.e**	**Coeff. std. err.**	**Coeff. std. err.**	**Coeff. std. err.**
PF variables	Log-employment	0.1027 [0.0341]***	(0.0306)***	0.0976 [0.0331]***	0.0516 [0.0250]**	0.0527 [0.0250]**
	Log-materials	0.7989 [0.0185]***	(0.0462)***	0.8362 [0.0186]***	0.8607 [0.0176]***	0.8628 [0.0177]***
	Log-capital	0.0676 [0.0239]***	(0.0153)***	0.0629 [0.0225]***	0.0537 [0.0146]***	0.0502 [0.0147]***
Infrastructure	Longest days to clear customs to export (a)	-0.0125 [0.0263]	(0.0376)	-0.0039 [0.0275]	-0.0158 [0.0209]	-0.0156 [0.0208]
	Dummy for own generator	0.0538 [0.0422]	(0.0424)	0.0378 [0.0396]	0.015 [0.0247]	0.0131 [0.0249]
	Water supply from public sources (b)	0.0014 [0.0005]***	(0.0008)*	0.0013 [0.0004]***	0.0009 [0.0003]***	0.0008 [0.0003]**
	Shipment losses in the domestic market (b)	-0.0047 [0.0039]	(0.0128)	-0.0023 [0.0035]	-0.0017 [0.0030]	-0.0016 [0.0030]
	Dummy for own transport	0.0238 [0.0475]	(0.0861)	-0.0084 [0.0464]	-0.003 [0.0340]	-0.0023 [0.0341]
	Dummy for web page	0.0402 [0.0394]	(0.0264)	0.0047 [0.0378]	0.0013 [0.0310]	0.0008 [0.0313]
	Dummy for security	0.0467 [0.0423]	(0.1407)	0.0426 [0.0403]	0.0497 [0.0285]*	0.0505 [0.0285]*
Red tape, corruption and crime	Sales reported to taxes (b)	0.0006 [0.0014]	(0.0052)	0.0009 [0.0013]	0.0008 [0.0010]	0.0009 [0.0010]
	Workforce reported to taxes (b)	-0.0015 [0.0012]	(0.0042)	-0.0015 [0.0010]	-0.0009 [0.0008]	-0.0009 [0.0008]
	Dummy for payments to speed up bureaucracy	-0.0464 [0.0336]	(0.0526)	-0.0443 [0.0292]	0.0041 [0.0255]	0.0083 [0.0259]
	Dummy for interventionist labor regulation	-0.036 [0.0361]	(0.0211)*	-0.0317 [0.0340]	-0.0259 [0.0330]	-0.028 [0.0331]
	Absenteeism (b)	-0.0299 [0.0222]	(0.0571)	-0.0204 [0.0195]	-0.0069 [0.0156]	-0.0071 [0.0160]
Finance and corporate governance	Dummy for trade association	0.0785 [0.0455]*	(0.0456)*	0.0756 [0.0408]*	0.024 [0.0297]	0.0194 [0.0300]
	Working capital financed by domestic private banks (b)	0.0002 [0.0007]	(0.0005)	-0.0002 [0.0007]	0.0003 [0.0006]	0.0003 [0.0006]
	Dummy for external audit	0.0691 [0.0395]*	(0.0452)	0.0662 [0.0362]*	0.0633 [0.0283]**	0.0655 [0.0282]**
	Dummy for loan (b)	0.1102 [0.0473]**	(0.0637)*	0.0892 [0.0464]*	0.0121 [0.0331]	0.006 [0.0327]
Quality, innovation and labor skills	Dummy for R&D (a)	0.1787 [0.2382]	(0.2347)	0.2041 [0.2534]	0.0702 [0.1322]	0.0638 [0.1320]
	Dummy for product innovation	-0.0073 [0.0360]	(0.0710)	-0.0153 [0.0332]	-0.025 [0.0244]	-0.0265 [0.0246]
	Dummy for foreign license (b)	0.204 [0.1053]*	(0.1302)	0.1425 [0.1033]	0.086 [0.0847]	0.0801 [0.0852]
	Dummy for internal training (b)	0.0579 [0.0533]	(0.0516)	0.0578 [0.0511]	0.0702 [0.0443]	0.0703 [0.0442]
	Unskilled workforce (a)	0.0013 [0.0036]	(0.0016)	0.0013 [0.0036]	-0.0034 [0.0030]	-0.0039 [0.0031]
	Workforce with computer	0.0017 [0.0011]	(0.0015)	0.0016 [0.0010]	0.0012 [0.0009]	0.0011 [0.0008]
Other control variables	Dummy for incorporated company	0.0265 [0.0396]	(0.0901)	0.0162 [0.0368]	0.0272 [0.0301]	0.0261 [0.0300]
	Age	0.0534 [0.0267]**	(0.0214)**	0.0438 [0.0251]*	0.0456 [0.0174]**	0.0487 [0.0174]***
	Share of exports (b)	0.001 [0.0009]	(0.0005)**	0.0006 [0.0009]	0.00004 [0.0006]	-0.0001 [0.0006]
	Trade union (b)	0.0008 [0.0012]	(0.0010)	0.0008 [0.0012]	0.0009 [0.0009]	0.0007 [0.0009]
	Strikes (b)	-0.0683 [0.0449]	(0.0821)	-0.0475 [0.0380]	-0.0112 [0.0307]	-0.0107 [0.0314]
	Constant	0.7377 [0.3449]**		0.4456 [0.3504]	1.0108 [0.2499]***	1.0335 [0.2492]***
	Industry/region/size/time dummies	Yes		Yes	Yes	Yes
	Observations	5211		5216	5175	5176
	R-squared	0.88		0.9	0.94	0.94

Estimating results of equation (1) under different imputation mechanisms for missing data. Those observations with missing values in all sales, labor (labor cost), materials and capital are excluded in all the regressions.

[1] ICA method is in section 3 of main text. Significance is given by clustered by industry and region Whit- robust standard errors in brackets; *** 1%, **5%, * 10%. In parentheses are bootstrap standard errors after 1000 replications (see section 5.2.2 on the motivation of using bootstrap standard errors). Correlation by clusters is also considered.

[2] EM algorithms are explained in section 5.1. EM alg [1] includes as covariates of the imputation mechanism industry/region/size/time (I/R/S/T) dummies (see section 5.1.1); EM alg [2] includes I/R/S/T dummies and production function variables (see section 5.1.2); Am alg [3] include also a set of IC variables (see section 5.1.3). Significance is given by cluster White-robust standard errors.

(a) IC variables instrumented with industry/region average variables. (b) missing values in IC variables replaced by means of *ICA method*.

Source: Authors' calculations with ICSs

Table 9.2: TURKEY, Extended production function and comparison of ICA method with EM algorithms

	Dependent variable: Log of total sales	ICA Method[1]		EM Algorithms[2] [1]	EM Algorithms[2] [2]	EM Algorithms[2] [3]
Category	**Variable**	**Coeff. std. err.**	**Boot. st. er.**	**Coeff. std. err.**	**Coeff. std. err.**	**Coeff. std. err.**
PF variables	Log-employment	0.416 [0.0492]***	(0.1088)***	0.3743 [0.0434]***	0.3421 [0.0459]***	0.3323 [0.0467]***
	Log-materials	0.4184 [0.0404]***	(0.0249)***	0.4829 [0.0429]***	0.6075 [0.0369]***	0.6052 [0.0370]***
	Log-capital	0.0371 [0.0165]**	(0.0428)	0.0548 [0.0199]***	0.0801 [0.0190]***	0.0783 [0.0184]***
Infrastructures	Days to clear customs to imports (a)	-0.0707 [0.0686]	(0.0688)	-0.1497 [0.0578]**	-0.1206 [0.0516]**	-0.1399 [0.0462]***
	Dummy for e-mail	0.2866 [0.0920]***	(0.1365)**	0.1659 [0.0789]**	0.1648 [0.0726]**	0.188 [0.0720]**
Red tape, corruption and crime	Security expenses (b)	-0.0246 [0.0828]	(0.0011)***	-0.0117 [0.0520]	-0.0504 [0.0456]	-0.0647 [0.0416]
	Payments to deal with bureaucratic issues (a)	-0.011 [0.0020]***	(0.0077)	-0.0092 [0.0013]***	-0.0065 [0.0011]***	-0.0072 [0.0012]***
	Sales declared to taxes (a)	-0.0226 [0.0057]***	(0.0045)***	-0.0234 [0.0046]***	-0.0148 [0.0042]***	-0.0177 [0.0040]***
	Number of inspections (b)	0.0046 [0.0044]	(0.0597)	-0.0002 [0.0026]	0.0001 [0.0026]	0.0007 [0.0023]
	Payments to obtain a contract with the government (b)	-0.0373 [0.0315]	(0.0058)***	-0.0524 [0.0217]**	-0.0274 [0.0175]	-0.0514 [0.0159]***
	Production lost due to absenteeism (b)	-0.0054 [0.0043]	(0.0367)	-0.0122 [0.0037]***	-0.0082 [0.0028]***	-0.0094 [0.0029]***
	Dummy for informal competition (b)	0.0044 [0.0295]	(0.1203)	-0.0055 [0.0236]	-0.0013 [0.0189]	-0.0059 [0.0196]
	Delay to obtain a water supply (a)	-0.1325 [0.0634]**	(0.0993)	-0.1388 [0.0565]**	-0.0746 [0.0559]	-0.0935 [0.0600]
Finance	Dummy for credit line	0.068 [0.0868]	(0.1383)	0.1157 [0.0702]	0.0744 [0.0660]	0.0778 [0.0674]
	Dummy for external auditory (a)	0.0863 [0.0753]	(0.1117)	0.0655 [0.0461]	0.0627 [0.0397]	0.0935 [0.0406]**
	Borrows in foreign currency (b)	0.0018 [0.0009]**	(0.0010)*	0.0013 [0.0005]**	0.0008 [0.0006]	0.0007 [0.0006]
Quality, innov. and labor skills	Staff with university education (b)	0.0095 [0.0026]***	(0.0018)***	0.0087 [0.0029]***	0.0064 [0.0029]**	0.0081 [0.0029]***
	Staff-part time workers	-0.008 [0.0030]**	(0.0222)	-0.0046 [0.0023]*	-0.0059 [0.0016]***	-0.0058 [0.0018]***
Other control variables	Production lost due to strikes (b)	-0.1689 [0.0634]**	(0.0351)***	-0.1596 [0.0435]***	-0.124 [0.0322]***	-0.1072 [0.0323]***
	Dummy for recently privatized firm	1.0606 [0.2812]***	(0.2511)***	0.8692 [0.2579]***	0.6825 [0.2478]***	0.6644 [0.2508]**
	Dummy for competition against imported products	0.2069 [0.0962]**	(0.2737)	0.1595 [0.0736]**	0.0951 [0.0603]	0.0755 [0.0607]
	Constant	3.5299 [0.7190]***		3.6661 [0.5851]***	1.6872 [0.3782]***	1.9648 [0.3791]***
	Industry/region/size/time dummies	Yes		Yes	Yes	Yes
	Observations	1684		1679	1733	1733
	R-squared	0.73		0.81	0.86	0.86

Estimating results of equation (1) under different imputation mechanisms for missing data. Those observations with missing values in all sales, labor (labor cost), materials and capital are excluded in all the regressions.

[1, 2] See footnotes in Table 9.1.

(a) IC variables instrumented with industry/region average variables. (b) missing values in IC variables replaced by means of *ICA method*.

Source: Authors' calculations with ICSs.

Table 9.3: SOUTH AFRICA, Extended production function and comparison of ICA method with EM algorithms

	Dependent variable: Log of total sales	ICA Method[1]		EM Algorithms[2]		
				[1]	[2]	[3]
Category	**Variable**	**Coeff. std. err.**	**Boot. st. er.**	**Coeff. std. err.**	**Coeff. std. err.**	**Coeff. std. err.**
PF variables	Log-employment	0.3226 [0.0711]***	(0.0365)***	0.3144 [0.0676]***	0.2285 [0.0667]***	0.2261 [0.0666]***
	Log-materials	0.5195 [0.1017]***	(0.0214)***	0.5355 [0.0942]***	0.5781 [0.0947]***	0.574 [0.0943]***
	Log-capital	0.1247 [0.0300]***	(0.0118)***	0.1287 [0.0370]***	0.123 [0.0373]***	0.1282 [0.0386]***
Infrastructure	Days to clear customs to import (a)	-0.1188 [0.1125]	(0.1233)	0.1291 [0.1320]	0.0193 [0.0935]	0.0322 [0.0975]
	Sales lost due to power outages (b)	-0.0171 [0.0114]	(0.0047)***	-0.0128 [0.0101]	-0.0112 [0.0077]	-0.0096 [0.0073]
	Water outages (b)	-0.1477 [0.0527]***	(0.0942)	-0.1287 [0.0438]***	-0.1482 [0.0533]***	-0.1611 [0.0562]***
	Average duration of transport failures (a)	-0.0439 [0.0806]	(0.0379)	0.06 [0.0893]	0.0021 [0.0628]	-0.0156 [0.0611]
	Wait for electric supply (a)	-0.0867 [0.0553]	(0.0173)***	-0.1368 [0.0337]***	-0.0921 [0.0272]***	-0.0863 [0.0258]***
	Sales lost due to delivery delays (b)	-0.0099 [0.0083]	(0.0073)	-0.0148 [0.0084]*	-0.0097 [0.0072]	-0.0077 [0.0065]
Red tape, corruption and crime	Manager's time spent in bur. issues (b)	0.007 [0.0051]	(0.0016)***	0.0077 [0.0050]	0.0077 [0.0057]	0.0084 [0.0058]
	Payments to deal with bureaucratic issues (b)	-0.0045 [0.0024]*	(0.3604)	-0.005 [0.0026]*	-0.0042 [0.0023]*	-0.0121 [0.0038]***
	Sales declared to taxes (a)	0.0056 [0.0046]	(0.0022)**	0.0056 [0.0042]	0.0059 [0.0025]**	0.0058 [0.0027]**
	Payments to obtain a contract with the government (b)	-0.0144 [0.0185]	(0.1975)	-0.0119 [0.0175]	-0.0161 [0.0146]	-0.015 [0.0144]
	Security expenses (a)	0.1407 [0.0511]**	(0.0069)***	-0.0023 [0.0148]	-0.0075 [0.0109]	-0.0056 [0.0113]
	Illegal payments in protection (b)	0.3969 [0.2428]	(0.1128)***	0.3754 [0.2492]	0.3882 [0.2202]*	0.4761 [0.2187]**
	Crime losses (a)	-0.0502 [0.0788]	(0.1374)	-0.0541 [0.0948]	0.0099 [0.0621]	0.0193 [0.0662]
Finance and corporate governance	Percentage of credit unused (b)	0.0014 [0.0010]	(0.0013)	0.0016 [0.0008]*	0.0019 [0.0010]*	0.002 [0.0010]*
	Dummy for loan	0.0715 [0.0492]	(0.0327)**	0.0841 [0.0479]*	0.0762 [0.0406]*	0.0761 [0.0407]*
	Value of the collateral (b)	-0.0008 [0.0002]***	(0.0009)	-0.0007 [0.0002]***	-0.0006 [0.0002]***	-0.0007 [0.0002]***
	Borrows in foreign currency (b)	0.0018 [0.0022]	(0.0024)	0.0007 [0.0018]	-0.0002 [0.0012]	0.0001 [0.0012]
	Charge to clear a check (a)	-0.1164 [0.0503]**	(0.0253)***	-0.0861 [0.0520]	-0.0995 [0.0387]**	-0.1068 [0.0384]***
	Largest shareholder	0.0006 [0.0010]	(0.0008)	0.0011 [0.0010]	0.001 [0.0007]	0.0011 [0.0007]
	Working capital financed by foreign commercial banks (b)	0.0106 [0.0083]	(0.0084)	0.0072 [0.0072]	0.0057 [0.0060]	0.0044 [0.0063]
	Working capital financed by informal sources (b)	-0.0022 [0.0023]	(0.0001)***	-0.0018 [0.0021]	-0.0027 [0.0018]	-0.0026 [0.0018]
Quality, innovation and labor skills	Dummy for ISO quality certification (b)	0.1603 [0.0766]**	(0.0365)***	0.1521 [0.0732]**	0.0838 [0.0404]**	0.0782 [0.0390]*
	Dummy for new product (b)	0.091 [0.0494]*	(0.0113)***	0.1083 [0.0530]**	0.1053 [0.0461]**	0.1001 [0.0460]**
	Dummy for discontinued product line (b)	-0.1007 [0.0610]	(0.0384)**	-0.1029 [0.0560]*	-0.0874 [0.0541]	-0.0805 [0.0534]
	Staff - management	0.004 [0.0028]	(0.0009)***	0.0036 [0.0028]	0.0034 [0.0030]	0.0032 [0.0030]
	Staff - non-production workers	-0.0034 [0.0022]	(0.0025)	-0.0032 [0.0022]	-0.0023 [0.0022]	-0.0024 [0.0022]
	Training to unskilled workers (a)	0.001 [0.0026]	(0.0030)	-0.0001 [0.0038]	0.0018 [0.0019]	0.0008 [0.0021]
	University staff (b)	0.0049 [0.0015]***	(0.0007)***	0.0055 [0.0016]***	0.0039 [0.0014]***	0.0038 [0.0014]**
	Manager's experience (b)	0.0391 [0.0249]	(0.0217)*	0.0369 [0.0222]	0.028 [0.0187]	0.0271 [0.0184]
Other control variables	Age (b)	0.0018 [0.0015]	(0.0016)	0.0014 [0.0013]	0.0023 [0.0013]*	0.0023 [0.0013]*
	Share of the local market (b)	0.0032 [0.0008]***	(0.0004)***	0.0035 [0.0009]***	0.0027 [0.0007]***	0.0028 [0.0007]***
	Constant	2.7174 [0.8932]***	(0.0365)***	2.0109 [0.8200]**	2.5368 [0.7330]***	2.5977 [0.7464]***
	Industry/region/size/time dummies	Yes		Yes	Yes	Yes
	Observations	1483		1528	1552	1550
	R-squared	0.89		0.89	0.9	0.9

Estimating results of equation (1) under different imputation mechanisms for missing data. Those observations with missing values in all sales, labor (labor cost), materials and capital are excluded in all the regressions. [1, 2] See footnotes in Table 9.1.
Source: Authors' calculations with ICSs.

Table 9.4: TANZANIA, Extended production function and comparison of ICA method with EM algorithms

	Dependent variable: Log of total sales	ICA Method[1]		EM Algorithms[2] [1]	EM Algorithms[2] [2]	EM Algorithms[2] [3]
Category	**Variable**	**Coeff. std. err.**	**Boot. st. er.**	**Coeff. std. err.**	**Coeff. std. err.**	**Coeff. std. err.**
PF variables	Log-employment	0.1655 [0.0853]*	(0.0512)***	0.1142 [0.0919]	0.0584 [0.0459]	0.0207 [0.0501]
	Log-materials	0.4252 [0.0581]***	(0.0340)***	0.4867 [0.0677]***	0.7201 [0.0435]***	0.724 [0.0401]***
	Log-capital	0.1589 [0.0323]***	(0.0208)***	0.1628 [0.0317]***	0.1326 [0.0286]***	0.1171 [0.0288]***
Infrastructure	Electricity from own generator (b)	0.0021 [0.0016]	(0.0053)	0.0035 [0.0016]**	0.0036 [0.0011]***	0.0027 [0.0011]**
	Losses due to water outages (b)	-0.0112 [0.0058]*	(0.0162)	-0.0172 [0.0049]***	-0.0087 [0.0033]**	-0.0082 [0.0034]**
	Water from own well or water infrastructure (a)	0.0001 [0.0051]	(0.0011)	-0.0044 [0.0044]	0.0013 [0.0031]	0.0029 [0.0035]
	Losses due to phone outages (a)	-0.0322 [0.0198]	(0.0071)***	-0.0066 [0.0268]	0.0115 [0.0232]	0.0159 [0.0260]
	Transport outages (a)	-0.0047 [0.0703]	(0.1168)	0.0366 [0.0623]	0.0069 [0.0295]	-0.0287 [0.0280]
	Dummy for own roads (b)	0.289 [0.1488]*	(0.0581)***	0.3789 [0.1279]***	0.2864 [0.1176]**	0.4766 [0.1189]***
	Dummy for webpage (b)	0.1578 [0.1212]	(0.1994)	0.0972 [0.1533]	0.1054 [0.1346]	0.2051 [0.1243]
	Wait for a water supply (a)	-0.1814 [0.0427]***	(0.0702)**	-0.1354 [0.0533]**	-0.093 [0.0262]***	-0.1649 [0.0235]***
	Low quality supplies (a)	-0.0163 [0.0128]	(0.0041)***	-0.0351 [0.0141]**	-0.0165 [0.0105]	-0.0202 [0.0112]*
Red tape, corruption and crime	Gift to obtain a operating license (b)	-0.4983 [0.1935]**	(0.1066)***	-0.3964 [0.1550]**	0.0537 [0.1051]	-0.0553 [0.0983]
	Payments to deal with bureaucratic issues (b)	0.0939 [0.0299]***	(0.0164)***	0.0808 [0.0272]***	0.0512 [0.0503]	0.085 [0.0396]**
	Days in inspections (b)	-0.1045 [0.0735]	(0.0494)**	-0.0735 [0.0703]	0.0027 [0.0379]	0.0005 [0.0362]
	Payments to obtain a contract with the government (b)	-0.0114 [0.0066]*	(0.0091)	-0.0026 [0.0059]	-0.0082 [0.0040]*	-0.0079 [0.0044]*
	Security expenses (b)	-0.0119 [0.0042]***	(0.0092)	-0.0081 [0.0040]*	-0.0023 [0.0031]	-0.0005 [0.0035]
	Illegal payments in protection (b)	-0.0827 [0.0170]***	(0.1019)	-0.0518 [0.0140]***	-0.031 [0.0144]**	-0.0489 [0.0206]**
Finance and corporate governance	Interest rate of the loan (a)	-0.0109 [0.0145]	(0.0099)	-0.0139 [0.0127]	0.0033 [0.0073]	0.0117 [0.0078]
	Working capital financed by commercial banks (b)	-0.0009 [0.0018]	(0.0012)	-0.0015 [0.0016]	-0.0016 [0.0011]	-0.001 [0.0011]
	Working capital financed by leasing (b)	-0.0794 [0.0282]***	(0.0054)***	-0.118 [0.0279]***	-0.015 [0.0038]***	-0.0893 [0.0428]**
	Sales bought on credit (b)	-0.0014 [0.0012]	(0.0011)	0 [0.0011]	0.0006 [0.0010]	0.0006 [0.0010]
	Delay to clear a domestic currency wire (a)	-0.3418 [0.3273]	(0.0935)***	-0.0439 [0.2600]	0.1691 [0.1544]	0.0498 [0.1606]
Quality, innovation and labor skills	Dummy for new product (b)	0.0429 [0.1063]	(0.2036)	-0.0053 [0.1090]	-0.0481 [0.0782]	-0.0897 [0.0632]
	Staff - skilled workers (b)	0.0026 [0.0023]	(0.0050)	0.0025 [0.0021]	0.0036 [0.0014]**	0.0038 [0.0014]**
	Workforce with computer (b)	0.0066 [0.0030]**	(0.0056)	0.0071 [0.0034]**	0.0001 [0.0049]	0.003 [0.0041]
Other control variables	Dummy for incorporated company (b)	0.2914 [0.2023]	(0.5683)	-0.0777 [0.4506]	0.1645 [0.1868]	0.0871 [0.2324]
	Dummy for FDI (b)	0.1397 [0.1445]	(0.2844)	0.0825 [0.1397]	-0.0662 [0.0792]	-0.0859 [0.0768]
	Constant	7.2978 [1.0168]***		6.3827 [0.8512]***	2.4414 [0.5932]***	3.296 [0.6161]***
	Industry/region/size/time dummies	Yes		Yes	Yes	Yes
	Observations	559		560	603	597
	R-squared	0.88		0.88	0.94	0.94

Estimating results of equation (1) under different imputation mechanisms for missing data. Those observations with missing values in all sales, labor (labor cost), materials and capital are excluded in all the regressions.

[1, 2,] See footnotes in Table 9.1.

Source: Authors' calculations with ICSs.

Table 10.1: INDIA, Extended production function and comparison of ICA method with extensions

Dependent variable: log of total sales		ICA method and extensions				Complete case[4]	Multiple imputation (Switching regr.)[5]
		Original ICA meth.[1]		Random ICA m.[2]	ICA m. on inputs[3]		
Category	Variable	Coeff. std. err.	Boot. s.e	Coeff. std. err.	Coeff. std. err.	Coeff. std. err.	Coeff. std. err.
PF variables	Log-employment	0.1027 [0.0341]***	(0.0306)***	0.1051 [0.0346]***	0.0922 [0.0343]***	0.1168 [0.0317]***	0.0659 [0.0245]***
	Log-materials	0.7989 [0.0185]***	(0.0462)***	0.8135 [0.0186]***	0.8054 [0.0192]***	0.7994 [0.0236]***	0.8560 [0.0169]***
	Log-capital	0.0676 [0.0239]***	(0.0153)***	0.0438 [0.0143]***	0.0722 [0.0248]***	0.0504 [0.0170]***	0.0452 [0.0128]***
Infrastructure	Longest days to clear customs to export (a)	-0.0125 [0.0263]	(0.0376)	-0.01 [0.0317]	-0.0167 [0.0266]	-0.0432 [0.0268]	-0.0155 [0.0213]
	Dummy for own generator	0.0538 [0.0422]	(0.0424)	-0.0083 [0.0453]	0.0516 [0.0431]	0.0424 [0.0293]	0.0198 [0.0254]
	Water supply from public sources (b)	0.0014 [0.0005]***	(0.0008)*	0.0009 [0.0006]	0.0014 [0.0005]***	0.0013 [0.0004]***	0.0008 [0.0003]**
	Shipment losses in the domestic market (b)	-0.0047 [0.0039]	(0.0128)	-0.0075 [0.0034]**	-0.0037 [0.0038]	-0.0023 [0.0054]	-0.0020 [0.0029]
	Dummy for own transport	0.0238 [0.0475]	(0.0861)	0.0013 [0.0459]	0.0334 [0.0482]	0.0465 [0.0369]	-0.0038 [0.0347]
	Dummy for web page	0.0402 [0.0394]	(0.0264)	0.0516 [0.0427]	0.0329 [0.0382]	0.0098 [0.0327]	0.0067 [0.0316]
	Dummy for security	0.0467 [0.0423]	(0.1407)	0.045 [0.0392]	0.0573 [0.0429]	0.0564 [0.0293]*	0.0582 [0.0293]**
Red tape, corruption and crime	Sales reported to taxes (b)	0.0006 [0.0014]	(0.0052)	0.002 [0.0012]*	0.0009 [0.0014]	0.0002 [0.0010]	0.0010 [0.0009]
	Workforce reported to taxes (b)	-0.0015 [0.0012]	(0.0042)	-0.0021 [0.0009]**	-0.0014 [0.0012]	0.0005 [0.0008]	-0.0010 [0.0007]
	Dummy for payments to speed up bureaucracy	-0.0464 [0.0336]	(0.0526)	-0.0148 [0.0265]	-0.0416 [0.0335]	0.0072 [0.0247]	0.0004 [0.0254]
	Dummy for interventionist labor regulation	-0.036 [0.0361]	(0.0211)*	-0.0372 [0.0369]	-0.0275 [0.0368]	-0.031 [0.0330]	-0.0303 [0.0322]
	Absenteeism (b)	-0.0299 [0.0222]	(0.0571)	-0.0233 [0.0256]	-0.0263 [0.0216]	-0.0011 [0.0193]	-0.0108 [0.0158]
Finance and corporate governance	Dummy for trade association	0.0785 [0.0455]*	(0.0456)*	0.094 [0.0480]*	0.0734 [0.0454]	0.022 [0.0388]	0.0263 [0.0302]
	Working capital financed by domestic private banks (b)	0.0002 [0.0007]	(0.0005)	0.0005 [0.0006]	0.0002 [0.0008]	0.0003 [0.0008]	0.0002 [0.0005]
	Dummy for external audit	0.0691 [0.0395]*	(0.0452)	0.0541 [0.0440]	0.0627 [0.0386]	0.0392 [0.0300]	0.0689 [0.0294]**
	Dummy for loan (b)	0.1102 [0.0473]**	(0.0637)*	0.0851 [0.0538]	0.1107 [0.0492]**	-0.0397 [0.0409]	0.0188 [0.0337]
Quality, innovation and labor skills	Dummy for R&D (a)	0.1787 [0.2382]	(0.2347)	0.0959 [0.1637]	0.1885 [0.2400]	0.0862 [0.1313]	0.1143 [0.1353]
	Dummy for product innovation	-0.0073 [0.0360]	(0.0710)	-0.0331 [0.0392]	-0.0079 [0.0366]	-0.0528 [0.0262]**	-0.0285 [0.0276]
	Dummy for foreign license (b)	0.204 [0.1053]*	(0.1302)	0.2384 [0.1181]**	0.1555 [0.1013]	0.1401 [0.0939]	0.1032 [0.0835]
	Dummy for internal training (b)	0.0579 [0.0533]	(0.0516)	0.0744 [0.0649]	0.0631 [0.0537]	0.0884 [0.0458]	0.0717 [0.0440]*
	Unskilled workforce (a)	0.0013 [0.0036]	(0.0016)	0.0038 [0.0042]	0.0003 [0.0037]	-0.001 [0.0033]	-0.0030 [0.0029]
	Workforce with computer	0.0017 [0.0011]	(0.0015)	0.0014 [0.0009]	0.0019 [0.0011]*	0.0007 [0.0007]	0.0012 [0.0008]
Other control variables	Dummy for incorporated company	0.0265 [0.0396]	(0.0901)	0.056 [0.0358]	0.0127 [0.0423]	0.0494 [0.0282]*	0.0280 [0.0311]
	Age	0.0534 [0.0267]**	(0.0214)**	0.0352 [0.0287]	0.0525 [0.0271]*	0.0322 [0.0208]	0.0392 [0.0182]**
	Share of exports (b)	0.001 [0.0009]	(0.0005)**	0.001 [0.0010]	0.001 [0.0009]	0.0002 [0.0005]	-0.0001 [0.0005]
	Trade union (b)	0.0008 [0.0012]	(0.0010)	0.0015 [0.0013]	0.001 [0.0013]	0.0001 [0.0008]	0.0007 [0.0008]
	Strikes (b)	-0.0683 [0.0449]	(0.0821)	-0.0557 [0.0470]	-0.0707 [0.0457]	0.0248 [0.0439]	-0.0112 [0.0321]
	Constant	0.7377 [0.3449]**		0.7174 [0.3636]*	0.7182 [0.3455]**	1.0943 [0.2692]***	0.9976 [0.2528]***
	Industry/region/size/time dummies	Yes		Yes	Yes	Yes	Yes
	Observations	5211		5063	5134	3943	5262
	R-squared	0.88		0.88	0.88	0.94	-

Estimating results of equation (1) under different imputation mechanisms for missing data. Those observations with missing values in all sales, labor (labor cost), materials and capital are excluded in all the regressions.

[1] See footnote 1 in Table 9.1. [2] Random ICA method is described in section 5.2.1. [3] ICA method on inputs is in section 5.2.3. [4] Complete case considers missingness in PF variables only, not in IC variables. [5] Multiple imputation via switching regression can be found in section 5.3.

In all the cases significance is given by clustered by industry and region Whit- robust standard errors in brackets; *** 1%, **5%, * 10%. In the case of the ICA method, in parentheses are bootstrap standard errors after 1000 replications (see section 5.2.2 on the motivation of using bootstrap standard errors). Correlation by cluster is also considered.

Source: Authors' calculations with ICSs.

Table 10.2: TURKEY, Extended production function and comparison of ICA method with extensions

Dependent variable: log of total sales		ICA method and extensions				Complete case[4]	Multiple imputation (Switching regr.)[5]
		Original ICA meth.[1]		Random ICA m.[2]	ICA m. on inputs[3]		
Category	**Variable**	**Coeff. std. err.**	**Boot. st. er.**	**Coeff. std. err.**	**Coeff. std. err.**	**Coeff. std. err.**	**Coeff. std. err.**
PF variables	Log-employment	0.416 [0.0492]***	(0.1088)***	0.3819 [0.0501]***	0.5106 [0.0558]***	0.4002 [0.0885]***	0.3446 [0.0524]***
	Log-materials	0.4184 [0.0404]***	(0.0249)***	0.4137 [0.0392]***	0.4615 [0.0484]***	0.5332 [0.0494]***	0.5779 [0.0316]***
	Log-capital	0.0371 [0.0165]**	(0.0428)	0.0193 [0.0198]	0.0686 [0.0232]***	0.0639 [0.0271]**	0.0603 [0.0246]**
Infrastructures	Days to clear customs to imports (a)	-0.0707 [0.0686]	(0.0688)	-0.1133 [0.0776]	-0.0711 [0.0705]	-0.1594 [0.0856]*	-0.1318 [0.0660]**
	Dummy for e-mail	0.2866 [0.0920]***	(0.1365)**	0.3833 [0.1048]***	0.3072 [0.1054]***	0.0317 [0.1295]	0.1729 [0.0754]**
Red tape, corruption and crime	Security expenses (b)	-0.0246 [0.0828]	(0.0011)***	0.0137 [0.0836]	-0.0861 [0.0919]	-0.0468 [0.0786]	-0.0215 [0.0587]
	Payments to deal with bureaucratic issues (a)	-0.011 [0.0020]***	(0.0077)	-0.0108 [0.0021]***	-0.0102 [0.0021]***	-0.0084 [0.0014]***	-0.0073 [0.0011]***
	Sales declared to taxes (a)	-0.0226 [0.0057]***	(0.0045)***	-0.0197 [0.0061]***	-0.0151 [0.0065]**	-0.0184 [0.0082]**	-0.0159 [0.0051]***
	Number of inspections (b)	0.0046 [0.0044]	(0.0597)	0.001 [0.0049]	0.005 [0.0044]	-0.0019 [0.0038]	0.0000 [0.0036]
	Payments to obtain a contract with the government (b)	-0.0373 [0.0315]	(0.0058)***	-0.0345 [0.0357]	-0.0217 [0.0368]	-0.0257 [0.0360]	-0.0354 [0.0236]
	Production lost due to absenteeism (b)	-0.0054 [0.0043]	(0.0367)	-0.0079 [0.0051]	-0.005 [0.0039]	-0.0107 [0.0054]*	-0.0110 [0.0036]***
	Dummy for informal competition (b)	0.0044 [0.0295]	(0.1203)	-0.0083 [0.0323]	0.0207 [0.0279]	-0.0015 [0.0315]	-0.0062 [0.0232]
	Delay to obtain a water supply (a)	-0.1325 [0.0634]**	(0.0993)	-0.1346 [0.0688]*	-0.1419 [0.0863]	-0.0825 [0.0785]	-0.0965 [0.0571]*
Finance	Dummy for credit line	0.068 [0.0868]	(0.1383)	0.0967 [0.0905]	0.0888 [0.1061]	0.0657 [0.0685]	0.0699 [0.0719]
	Dummy for external auditory (a)	0.0863 [0.0753]	(0.1117)	0.0992 [0.0739]	0.1012 [0.0791]	0.1385 [0.0709]*	0.0781 [0.0521]
	Borrows in foreign currency (b)	0.0018 [0.0009]**	(0.0010)*	0.0015 [0.0008]*	0.0018 [0.0010]*	0.0005 [0.0009]	0.0009 [0.0008]
Quality, innov. and labor skills	Staff with university education (b)	0.0095 [0.0026]***	(0.0018)***	0.0107 [0.0028]***	0.01 [0.0040]**	0.008 [0.0035]**	0.0060 [0.0032]*
	Staff-part time workers	-0.008 [0.0030]**	(0.0222)	-0.0077 [0.0032]**	-0.0102 [0.0029]***	-0.0069 [0.0027]**	-0.0067 [0.0019]***
Other control variables	Production lost due to strikes (b)	-0.1689 [0.0634]**	(0.0351)***	-0.1063 [0.0650]	-0.1538 [0.0671]**	-0.1765 [0.0521]***	-0.1092 [0.0564]*
	Dummy for recently privatized firm	1.0606 [0.2812]***	(0.2511)***	1.0239 [0.2791]***	1.0215 [0.3100]***	1.2627 [0.3162]***	0.8012 [0.2884]***
	Dummy for competition against imported products	0.2069 [0.0962]**	(0.2737)	0.2013 [0.0962]**	0.2096 [0.1041]*	0.0156 [0.0823]	0.1021 [0.0665]
	Constant	3.5299 [0.7190]***		4.6379 [0.7023]***	1.4306 [0.5738]**	2.6911 [0.7730]***	2.6126 [0.4577]***
	Industry/region/size/time dummies	Yes		Yes	Yes	Yes	Yes
	Observations	1684		1684	1360	792	1646
	R-squared	0.73		0.68	0.75	0.85	-

Notes of Table 10.1

Source: Authors' calculations with ICSs.

Table 10.3: SOUTH AFRICA, Extended production function and comparison of ICA method with extensions

	Dependent variable: log of total sales	ICA method and extensions				Complete case[4]	Multiple imputation (Switching regr.)[5]
		Original ICA meth.[1]		Random ICA meth.[2]	ICA met. on inputs[3]		
Category	**Variable**	**Coeff. std. err.**	**Boot. st. er.**	**Coeff. std. err.**	**Coeff. std. err.**	**Coeff. std. err.**	**Coeff. std. err.**
PF variables	Log-employment	0.3226 [0.0711]***	(0.0365)***	0.3822 [0.0776]***	0.3295 [0.0717]***	0.3428 [0.0541]***	0.2453 [0.0681]***
	Log-materials	0.5195 [0.1017]***	(0.0214)***	0.4914 [0.0877]***	0.5182 [0.1015]***	0.4877 [0.0961]***	0.5674 [0.0905]***
	Log-capital	0.1247 [0.0300]***	(0.0118)***	0.0791 [0.0264]***	0.124 [0.0302]***	0.1118 [0.0322]***	0.1180 [0.0345]***
Infrastructure	Days to clear customs to import (a)	-0.1188 [0.1125]	(0.1233)	-0.14 [0.1247]	-0.1407 [0.1176]	0.018 [0.1976]	0.0423 [0.1008]
	Sales lost due to power outages (b)	-0.0171 [0.0114]	(0.0047)***	-0.0194 [0.0127]	-0.0142 [0.0104]	-0.003 [0.0085]	-0.0107 [0.0080]
	Water outages (b)	-0.1477 [0.0527]***	(0.0942)	-0.1441 [0.0591]**	-0.1405 [0.0513]**	-0.1427 [0.0659]**	-0.1393 [0.0504]***
	Average duration of transport failures (a)	-0.0439 [0.0806]	(0.0379)	-0.0065 [0.0867]	-0.074 [0.0832]	0.1229 [0.1507]	-0.0022 [0.0762]
	Wait for electric supply (a)	-0.0867 [0.0553]	(0.0173)***	-0.1075 [0.0589]*	-0.0767 [0.0573]	-0.0629 [0.0558]	-0.1014 [0.0309]***
	Sales lost due to delivery delays (b)	-0.0099 [0.0083]	(0.0073)	-0.0111 [0.0092]	-0.0119 [0.0080]	-0.0074 [0.0081]	-0.0089 [0.0072]
Red tape, corruption and crime	Manager's time spent in bur. issues (b)	0.007 [0.0051]	(0.0016)***	0.0072 [0.0051]	0.0073 [0.0052]	0.0058 [0.0043]	0.0079 [0.0056]
	Payments to deal with bureaucratic issues (b)	-0.0045 [0.0024]*	(0.3604)	-0.0063 [0.0031]*	-0.0045 [0.0023]*	-0.0008 [0.0125]	-0.0044 [0.0024]*
	Sales declared to taxes (a)	0.0056 [0.0046]	(0.0022)**	0.0015 [0.0049]	0.0064 [0.0044]	0.0091 [0.0039]**	0.0058 [0.0031]*
	Payments to obtain a contract with the government (b)	-0.0144 [0.0185]	(0.1975)	-0.0218 [0.0201]	-0.017 [0.0208]	-0.0129 [0.0112]	-0.0180 [0.0162]
	Security expenses (a)	0.1407 [0.0511]**	(0.0069)***	0.1245 [0.0586]**	0.1159 [0.0477]**	0.0227 [0.0146]	-0.0075 [0.0123]
	Illegal payments in protection (b)	0.3969 [0.2428]	(0.1128)***	0.4048 [0.2751]	0.3997 [0.2428]	0.3265 [0.3225]	0.3606 [0.2254]*
	Crime losses (a)	-0.0502 [0.0788]	(0.1374)	0.0153 [0.0855]	-0.0679 [0.0786]	0.1115 [0.0871]	-0.0121 [0.0708]
Finance and corporate governance	Percentage of credit unused (b)	0.0014 [0.0010]	(0.0013)	0.0014 [0.0010]	0.0015 [0.0010]	0.0007 [0.0006]	0.0018 [0.0010]*
	Dummy for loan	0.0715 [0.0492]	(0.0327)**	0.0678 [0.0547]	0.072 [0.0493]	0.0602 [0.0421]	0.0814 [0.0443]*
	Value of the collateral (b)	-0.0008 [0.0002]***	(0.0009)	-0.0008 [0.0002]***	-0.0008 [0.0002]***	-0.0009 [0.0002]***	-0.0007 [0.0002]***
	Borrows in foreign currency (b)	0.0018 [0.0022]	(0.0024)	0.0024 [0.0023]	0.0016 [0.0021]	0.0012 [0.0011]	-0.0001 [0.0012]
	Charge to clear a check (a)	-0.1164 [0.0503]**	(0.0253)***	-0.1404 [0.0570]**	-0.1108 [0.0501]**	-0.1722 [0.0582]***	-0.0905 [0.0402]**
	Largest shareholder	0.0006 [0.0010]	(0.0008)	-0.0003 [0.0010]	0.0008 [0.0009]	0.0001 [0.0009]	0.0010 [0.0008]
	Working capital finn. by foreign commercial banks (b)	0.0106 [0.0083]	(0.0084)	0.0073 [0.0090]	0.0107 [0.0082]	0.0203 [0.0195]	0.0050 [0.0062]
	Working capital financed by informal sources (b)	-0.0022 [0.0023]	(0.0001)***	-0.0032 [0.0023]	-0.0021 [0.0023]	-0.0046 [0.0011]***	-0.0025 [0.0019]
Quality, innovation and labor skills	Dummy for ISO quality certification (b)	0.1603 [0.0766]**	(0.0365)***	0.1956 [0.0646]***	0.1578 [0.0764]**	0.121 [0.0670]*	0.1029 [0.0454]**
	Dummy for new product (b)	0.091 [0.0494]*	(0.0113)***	0.1233 [0.0587]**	0.0926 [0.0496]*	0.0461 [0.0393]	0.0948 [0.0475]**
	Dummy for discontinued product line (b)	-0.1007 [0.0610]	(0.0384)**	-0.1334 [0.0648]**	-0.099 [0.0597]	-0.0616 [0.0353]*	-0.0864 [0.0527]*
	Staff - management	0.004 [0.0028]	(0.0009)***	0.0049 [0.0027]*	0.0038 [0.0027]	0.0041 [0.0030]	0.0034 [0.0030]
	Staff - non-production workers	-0.0034 [0.0022]	(0.0025)	-0.0033 [0.0021]	-0.0033 [0.0022]	-0.0026 [0.0021]	-0.0024 [0.0021]
	Training to unskilled workers (a)	0.001 [0.0026]	(0.0030)	0.0023 [0.0028]	0 [0.0025]	-0.0047 [0.0045]	0.0011 [0.0027]
	University staff (b)	0.0049 [0.0015]***	(0.0007)***	0.0051 [0.0015]***	0.0049 [0.0014]***	0.0044 [0.0011]***	0.0043 [0.0014]***
	Manager's experience (b)	0.0391 [0.0249]	(0.0217)*	0.0412 [0.0271]	0.0387 [0.0249]	0.0325 [0.0254]	0.0292 [0.0196]
Other control variables	Age (b)	0.0018 [0.0015]	(0.0016)	0.0019 [0.0014]	0.0017 [0.0014]	0.0023 [0.0013]*	0.0021 [0.0013]*
	Share of the local market (b)	0.0032 [0.0008]***	(0.0004)***	0.0023 [0.0009]**	0.0032 [0.0008]***	0.0027 [0.0009]***	0.0029 [0.0007]***
	Constant	2.7174 [0.8932]***	(0.0365)***	3.5878 [0.8355]***	2.6721 [0.8751]***	2.6313 [0.9880]**	2.6249 [0.7400]***
	Industry/region/size/time dummies	Yes		Yes	Yes	Yes	Yes
	Observations	1483		1483	1474	1236	1483
	R-squared	0.89		0.87	0.89	0.91	

Notes of Table 10.1

Source: Authors' calculations with ICSs.

Table 10.4: TANZANIA, Extended production function and comparison of ICA method with extensions

	Dependent variable: log of total sales	ICA method and extensions				Complete case[4]	Multiple imputation (Swithching regression)[5]
		Original ICA meth.[1]		Random ICA met.[2]	ICA met. on inputs[3]		
Category	**Variable**	**Coeff. std. err.**	**Boot. st. er.**	**Coeff. std. err.**	**Coeff. std. err.**	**Coeff. std. err.**	**Coeff. std. err.**
PF variables	Log-employment	0.1655 [0.0853]*	(0.0512)***	0.2643 [0.1039]**	0.2339 [0.0603]***	0.1651 [0.0681]**	0.1217 (0.0625)**
	Log-materials	0.4252 [0.0581]***	(0.0340)***	0.4008 [0.0527]***	0.6087 [0.0406]***	0.6242 [0.0468]***	0.7170 (0.0390)***
	Log-capital	0.1589 [0.0323]***	(0.0208)***	0.0975 [0.0418]**	0.1302 [0.0280]***	0.1311 [0.0312]***	0.0977 (0.0294)***
Infrastructure	Electricity from own generator (b)	0.0021 [0.0016]	(0.0053]	0.0013 [0.0017]	0.0019 [0.0016]	-0.0002 [0.0022]	0.0039 (0.0016)**
	Losses due to water outages (b)	-0.0112 [0.0058]*	(0.0162]	-0.0132 [0.0081]	-0.0058 [0.0051]	-0.0107 [0.0062]*	-0.0094 (0.0046)**
	Water from own well or water infrastructure (a)	0.0001 [0.0051]	(0.0011]	-0.0094 [0.0060]	-0.0017 [0.0046]	0.0004 [0.0056]	-0.0003 (0.0038)
	Losses due to phone outages (a)	-0.0322 [0.0198]	(0.0071)***	-0.0453 [0.0237]*	0.0003 [0.0208]	0.0089 [0.0209]	0.0078 (0.0238)
	Transport outages (a)	-0.0047 [0.0703]	(0.1168]	0.0785 [0.0940]	0.0243 [0.0573]	-0.0859 [0.0567]	0.0054 (0.0322)
	Dummy for own roads (b)	0.289 [0.1488]*	(0.0581)***	0.1502 [0.1582]	0.4010 [0.1164]***	0.4073 [0.1249]***	0.3117 (0.1422)**
	Dummy for webpage (b)	0.1578 [0.1212]	(0.1994]	0.1453 [0.1280]	0.2560 [0.1038]**	0.3106 [0.1170]**	0.0977 (0.1635)
	Wait for a water supply (a)	-0.1814 [0.0427]***	(0.0702)**	-0.1769 [0.0531]***	-0.1388 [0.0411]***	-0.1252 [0.0326]***	-0.1036 (0.0356)***
	Low quality supplies (a)	-0.0163 [0.0128]	(0.0041)***	-0.0389 [0.0164]**	-0.0210 [0.0127]	-0.0285 [0.0142]*	-0.0183 (0.0120)
Red tape, corruption and crime	Gift to obtain a operating license (b)	-0.4983 [0.1935]**	(0.1066)***	-0.4607 [0.2385]*	-0.3262 [0.1439]**	-0.1671 [0.1562]	0.0694 (0.1218)
	Payments to deal with bureaucratic issues (b)	0.0939 [0.0299]***	(0.0164)***	0.0376 [0.0578]	0.1182 [0.0295]***	0.0767 [0.0192]***	0.0546 (0.0472)
	Days in inspections (b)	-0.1045 [0.0735]	(0.0494)**	-0.1172 [0.0984]	-0.0514 [0.0425]	-0.0524 [0.0643]	-0.0009 (0.0461)
	Payments to obtain a contract with the government (b)	-0.0114 [0.0066]*	(0.0091]	-0.0177 [0.0086]**	-0.0189 [0.0066]***	-0.0254 [0.0078]***	-0.0140 (0.0051)***
	Security expenses (b)	-0.0119 [0.0042]***	(0.0092]	-0.0151 [0.0055]**	-0.0072 [0.0034]**	0.008 [0.0193]	-0.0042 (0.0032)
	Illegal payments in protection (b)	-0.0827 [0.0170]***	(0.1019]	-0.081 [0.0329]**	-0.0774 [0.0179]***	-0.0603 [0.0251]**	-0.0392 (0.0131)***
Finance and corporate governance	Interest rate of the loan (a)	-0.0109 [0.0145]	(0.0099]	-0.0028 [0.0182]	-0.0038 [0.0094]	0.0111 [0.0113]	-0.0021 (0.0090)
	Working capital financed by commercial banks (b)	-0.0009 [0.0018]	(0.0012]	-0.0008 [0.0021]	-0.0013 [0.0014]	0.0007 [0.0013]	-0.0014 (0.0012)
	Working capital financed by leasing (b)	-0.0794 [0.0282]***	(0.0054)***	-0.1362 [0.0450]***	-0.0489 [0.0305]	-0.0304 [0.0329]	-0.0129 (0.0069)*
	Sales bought on credit (b)	-0.0014 [0.0012]	(0.0011]	-0.0036 [0.0017]**	-0.0003 [0.0011]	-0.0021 [0.0014]	-0.0002 (0.0014)
	Delay to clear a domestic currency wire (a)	-0.3418 [0.3273]	(0.0935)***	-0.0024 [0.3738]	0.1242 [0.2583]	0.3236 [0.2952]	0.2044 (0.1717)
Quality, innovation and labor skills	Dummy for new product (b)	0.0429 [0.1063]	(0.2036]	0.1217 [0.1118]	-0.0526 [0.0945]	-0.1533 [0.1066]	-0.1045 (0.0981)
	Staff - skilled workers (b)	0.0026 [0.0023]	(0.0050]	0.0053 [0.0028]*	0.0038 [0.0022]*	0.0054 [0.0021]**	0.0039 (0.0020)*
	Workforce with computer (b)	0.0066 [0.0030]**	(0.0056]	0.0079 [0.0038]**	0.0094 [0.0039]**	0.0154 [0.0055]***	0.0037 (0.0045)
Other control variables	Dummy for incorporated company (b)	0.2914 [0.2023]	(0.5683]	0.238 [0.2648]	0.2327 [0.1841]	-0.2476 [0.1896]	0.2544 (0.2270)
	Dummy for FDI (b)	0.1397 [0.1445]	(0.2844]	0.3044 [0.1888]	0.1788 [0.1225]	0.1061 [0.1123]	-0.0255 (0.1128)
	Constant	7.2978 [1.0168]***		7.2545 [1.3295]***	2.7433 [0.8631]***	3.1164 [0.8674]***	2.4194 [0.7159]
	Industry/region/size/time dummies	Yes		Yes	Yes	Yes	Yes
	Observations	559		557	442	291	570
	R-squared	0.88		0.81	0.9300	0.95	

Notes of Table 10.1
Source: Authors' calculations with ICSs.

Table 11.1: INDIA, Extended production function and comparison of ICA method with Heckman models

Dependent variable: log of total sales		ICA Method[1]		Heckman models[2]		
				Heckman on comp case	Heckman replacing inputs	
Category	**Variable**	**Coeff. std. err.**	**Boot. s.e**	**Coeff. std. err.**	**Coeff. std. err.**	**Boot. s.e**
PF variables	Log-employment	0.1027 [0.0341]***	(0.0306)***	0.1127 [0.0160]***	0.0806 [0.0184]***	(0.0452)***
	Log-materials	0.7989 [0.0185]***	(0.0462)***	0.7998 [0.0069]***	0.8121 [0.0070]***	(0.0567)***
	Log-capital	0.0676 [0.0239]***	(0.0153)***	0.0477 [0.0062]***	0.0578 [0.0070]***	(0.0168)***
Infrastructure	Longest days to clear customs to export (a)	-0.0125 [0.0263]	(0.0376)	-0.0451 [0.0155]***	-0.0077 [0.0150]	(0.1542)
	Dummy for own generator	0.0538 [0.0422]	(0.0424)	0.0466 [0.0229]**	0.0769 [0.0265]***	(0.0064)***
	Water supply from public sources (b)	0.0014 [0.0005]***	(0.0008)*	0.0014 [0.0003]***	0.0012 [0.0003]***	(0.0460)***
	Shipment losses in the domestic market (b)	-0.0047 [0.0039]	(0.0128)	-0.0029 [0.0033]	-0.0022 [0.0029]	(0.1197)
	Dummy for own transport	0.0238 [0.0475]	(0.0861)	0.0438 [0.0283]	-0.0063 [0.0336]	(0.0742)
	Dummy for web page	0.0402 [0.0394]	(0.0264)	0.0061 [0.0221]	0.0212 [0.0263]	(0.0051)**
	Dummy for security	0.0467 [0.0423]	(0.1407)	0.0487 [0.0200]**	0.018 [0.0240]	(0.0035)**
Red tape, corruption and crime	Sales reported to taxes (b)	0.0006 [0.0014]	(0.0052)	-0.0001 [0.0007]	0.0011 [0.0008]	(0.0073)
	Workforce reported to taxes (b)	-0.0015 [0.0012]	(0.0042)	0.0005 [0.0007]	-0.001 [0.0007]	(0.0049)
	Dummy for payments to speed up bureaucracy	-0.0464 [0.0336]	(0.0526)	0.0079 [0.0186]	-0.0259 [0.0226]	(0.0463)
	Dummy for interventionist labor regulation	-0.036 [0.0361]	(0.0211)*	-0.0407 [0.0226]*	-0.0334 [0.0272]	(0.0658)**
	Absenteeism (b)	-0.0299 [0.0222]	(0.0571)	0.0003 [0.0112]	-0.0147 [0.0129]	(0.1783)**
Finance and corporate governance	Dummy for trade association	0.0785 [0.0455]*	(0.0456)*	0.0339 [0.0241]	0.0143 [0.0274]	(0.0762)
	Working capital financed by domestic private banks (b)	0.0002 [0.0007]	(0.0005)	0.0004 [0.0004]	0.001 [0.0004]**	(0.0006)**
	Dummy for external audit	0.0691 [0.0395]*	(0.0452)	0.0419 [0.0204]**	0.0827 [0.0245]***	(0.0408)*
	Dummy for loan (b)	0.1102 [0.0473]**	(0.0637)*	-0.0395 [0.0301]	0.1181 [0.0340]***	(0.0002)***
Quality, innovation and labor skills	Dummy for R&D (a)	0.1787 [0.2382]	(0.2347)	0.0813 [0.0933]	0.2063 [0.1112]*	(0.0010)
	Dummy for product innovation	-0.0073 [0.0360]	(0.0710)	-0.0508 [0.0200]**	-0.0081 [0.0233]	(0.0352)***
	Dummy for foreign license (b)	0.204 [0.1053]*	(0.1302)	0.141 [0.0434]***	0.1478 [0.0499]***	(0.0006)
	Dummy for internal training (b)	0.0579 [0.0533]	(0.0516)	0.0794 [0.0290]***	0.0813 [0.0338]**	(0.0093)
	Unskilled workforce (a)	0.0013 [0.0036]	(0.0016)	-0.0016 [0.0017]	-0.004 [0.0019]**	(0.1225)
	Workforce with computer	0.0017 [0.0011]	(0.0015)	0.0006 [0.0005]	0.0015 [0.0006]***	(0.0498)***
Other control variables	Dummy for incorporated company	0.0265 [0.0396]	(0.0901)	0.0566 [0.0225]**	0.016 [0.0273]	(0.0398)**
	Age	0.0534 [0.0267]**	(0.0214)**	0.0363 [0.0146]**	0.0856 [0.0181]***	(0.0431)**
	Share of exports (b)	0.001 [0.0009]	(0.0005)**	0.0001 [0.0004]	0.0003 [0.0004]	(0.0020)**
	Trade union (b)	0.0008 [0.0012]	(0.0010)	-0.00004 [0.0005]	0.0002 [0.0005]	(0.0014)**
	Strikes (b)	-0.0683 [0.0449]	(0.0821)	0.0482 [0.0301]	-0.0213 [0.0317]	(0.0043)
	Constant	0.7377 [0.3449]**		1.1579 [0.1899]***	0.8508 [0.2174]***	
	Industry/region/size/time dummies	Yes		Yes	Yes	
	Observations	5211		4323 (Cens: 5515/ Unc: 3808	5407 (Censored: 515/ Uncens: 4982)	
	R-squared	0.88				
	Heckman's Lambda (Inverse of Mills ration)			0.0130 [0.0634]	0.1221 [0.0926]	.

Estimating results of equation (1) under different imputation mechanisms for missing data. Those observations with missing values in all sales, labor (labor cost), materials and capital are excluded in all the regressions.[1] See footnote in Table 8.1. [2] Heckman models are explained in section 5.4. Heckman model on complete case considers missingness only in PF variables, not in IC variables, see section 5.4.1. Heckman replacing inputs compute the model on the sample with replacement of missing values in inputs (labor, materials and capital), see section 5.4.2.
In all the cases significance is given by clustered by industry and region Whit- robust standard errors in brackets; *** 1%, **5%, * 10%. In the case of the ICA method and Heckmand replacing inputs, in parentheses are bootstrap standard errors after 1000 replications (see sections and 5.2.2 5.4.2). Correlation by cluster is also considered. Source: Authors' calculations with ICSs.

Table 11.2: TURKEY, Extended production function and comparison of ICA method with Heckman models

Dependent variable: log of total sales		ICA Method[1]		Heckman models[2]		
				Heckman on complete case	Heckman replacing inputs	
Category	**Variable**	**Coeff. std. err.**	**Boot. st. er.**	**Coeff. Std.Err**	**Coeff. std. err.**	**Boot. st. er.**
PF variables	Log-employment	0.416 [0.0492]***	(0.1088)***	0.4017 [0.0423]***	0.5104 [0.0427]***	(0.0376)***
	Log-materials	0.4184 [0.0404]***	(0.0249)***	0.5306 [0.0189]***	0.4585 [0.0187]***	(0.0310)***
	Log-capital	0.0371 [0.0165]**	(0.0428)	0.063 [0.0164]***	0.067 [0.0182]***	(0.0168)***
Infrastructures	Days to clear customs to imports (a)	-0.0707 [0.0686]	(0.0688)	-0.155 [0.0835]*	-0.0648 [0.0859]	(0.0556)***
	Dummy for e-mail	0.2866 [0.0920]***	(0.1365)**	0.0193 [0.0822]	0.3121 [0.0786]***	(0.0659)**
Red tape, corruption and crime	Security expenses (b)	-0.0246 [0.0828]	(0.0011)***	-0.0379 [0.0824]	-0.0658 [0.0831]	(0.0575)**
	Payments to deal with bureaucratic issues (a)	-0.011 [0.0020]***	(0.0077)	-0.0084 [0.0009]***	-0.0101 [0.0010]***	(0.0012)***
	Sales declared to taxes (a)	-0.0226 [0.0057]***	(0.0045)***	-0.0175 [0.0075]**	-0.0131 [0.0077]*	(0.0055)***
	Number of inspections (b)	0.0046 [0.0044]	(0.0597)	-0.0017 [0.0043]	0.0049 [0.0045]	(0.0028)
	Payments to obtain a contract with the government (b)	-0.0373 [0.0315]	(0.0058)***	-0.0371 [0.0323]	-0.0363 [0.0315]	(0.0256)**
	Production lost due to absenteeism (b)	-0.0054 [0.0043]	(0.0367)	-0.0138 [0.0073]*	-0.0102 [0.0074]	(0.0042)**
	Dummy for informal competition (b)	0.0044 [0.0295]	(0.1203)	-0.011 [0.0283]	0.0046 [0.0306]	(0.0194)
	Delay to obtain a water supply (a)	-0.1325 [0.0634]**	(0.0993)	-0.0926 [0.0588]	-0.165 [0.0603]***	(0.0467)***
Finance	Dummy for credit line	0.068 [0.0868]	(0.1383)	0.0473 [0.0621]	0.0493 [0.0644]	(0.0482)**
	Dummy for external auditory (a)	0.0863 [0.0753]	(0.1117)	0.1407 [0.0617]**	0.1075 [0.0641]*	(0.0448)***
	Borrows in foreign currency (b)	0.0018 [0.0009]**	(0.0010)*	0.0003 [0.0009]	0.0016 [0.0009]*	(0.0008)*
Quality, innov. and labor skills	Staff with university education (b)	0.0095 [0.0026]***	(0.0018)***	0.0083 [0.0023]***	0.0104 [0.0024]***	(0.0018)***
	Staff-part time workers	-0.008 [0.0030]**	(0.0222)	-0.0065 [0.0027]**	-0.0093 [0.0028]***	(0.0019)***
Other control variables	Production lost due to strikes (b)	-0.1689 [0.0634]**	(0.0351)***	-0.1805 [0.0593]***	-0.153 [0.0723]**	(0.0453)***
	Dummy for recently privatized firm	1.0606 [0.2812]***	(0.2511)***	1.3287 [0.3695]***	1.0391 [0.2582]***	(0.2653)***
	Dummy for competition against imported products	0.2069 [0.0962]**	(0.2737)	0.021 [0.0724]	0.2084 [0.0730]***	(0.0634)***
	Constant	3.5299 [0.7190]***		3.0323 [0.6775]***	1.7704 [0.7084]**	(0.0376)***
	Industry/region/size/time dummies	Yes		Yes	Yes	
	Observations	1684		1941 (Censored: 1149/ Uncensored: 792)	2509 (Censored: 1149/ Uncensored: 1360)	
	R-squared	0.73				
	Heckman's Lambda			-0.1531 [0.1188]	0.0639 (0.1332)	

Notes of Table 11.1.
Source: Authors' calculations with ICSs.

Table 11.3: SOUTH AFRICA, Extended production function and comparison of ICA method with Heckman models

	Dependent variable: log of total sales	**ICA Method[1]**		**Heckman models[2]**		
				Heckman on complete case	**Heckman replacing inputs**	
Category	**Variable**	**Coeff. std. err.**	**Boot. st. er.**	**Coeff. Std.Err**	**Coeff. std. err.**	**Boot. st. er.**
PF variables	Log-employment	0.3226 [0.0711]***	(0.0365)***	0.3427 [0.0261]***	0.3275 [0.0250]***	(0.0452)***
	Log-materials	0.5195 [0.1017]***	(0.0214)***	0.4871 [0.0121]***	0.5184 [0.0120]***	(0.0567)***
	Log-capital	0.1247 [0.0300]***	(0.0118)***	0.1117 [0.0123]***	0.1241 [0.0129]***	(0.0168)***
Infrastructure	Days to clear customs to import (a)	-0.1188 [0.1125]	(0.1233]	0.032 [0.1133]	-0.1728 [0.1286]	(0.1542)
	Sales lost due to power outages (b)	-0.0171 [0.0114]	(0.0047)***	-0.0059 [0.0062]	-0.0166 [0.0069]**	(0.0064)***
	Water outages (b)	-0.1477 [0.0527]***	(0.0942]	-0.1215 [0.0501]**	-0.1383 [0.0516]***	(0.0460)***
	Average duration of transport failures (a)	-0.0439 [0.0806]	(0.0379]	0.1092 [0.0936]	-0.0821 [0.0985]	(0.1197)
	Wait for electric supply (a)	-0.0867 [0.0553]	(0.0173]***	-0.0311 [0.0544]	-0.057 [0.0717]	(0.0742)
	Sales lost due to delivery delays (b)	-0.0099 [0.0083]	(0.0073]	-0.0069 [0.0054]	-0.0109 [0.0054]**	(0.0051)**
Red tape, corruption and crime	Manager's time spent in bur. issues (b)	0.007 [0.0051]	(0.0016]***	0.0065 [0.0016]***	0.0079 [0.0017]***	(0.0035)**
	Payments to deal with bureaucratic issues (b)	-0.0045 [0.0024]*	(0.3604]	-0.0028 [0.0101]	-0.0056 [0.0039]	(0.0073)
	Sales declared to taxes (a)	0.0056 [0.0046]	(0.0022]**	0.0079 [0.0041]*	0.0062 [0.0056]	(0.0049)
	Payments to obtain a contract with the government (b)	-0.0144 [0.0185]	(0.1975]	-0.0099 [0.0198]	-0.0134 [0.0228]	(0.0463)
	Security expenses (a)	0.1407 [0.0511]**	(0.0069]***	0.0308 [0.0152]**	0.1324 [0.0578]**	(0.0658)**
	Illegal payments in protection (b)	0.3969 [0.2428]	(0.1128]***	0.2767 [0.1745]	0.3686 [0.0888]***	(0.1783)**
	Crime losses (a)	-0.0502 [0.0788]	(0.1374]	0.1006 [0.0792]	-0.0561 [0.0817]	(0.0762)
Finance and corporate governance	Percentage of credit unused (b)	0.0014 [0.0010]	(0.0013]	0.0006 [0.0005]	0.0013 [0.0006]**	(0.0006)**
	Dummy for loan	0.0715 [0.0492]	(0.0327]**	0.0634 [0.0400]	0.0705 [0.0413]*	(0.0408)*
	Value of the collateral (b)	-0.0008 [0.0002]***	(0.0009]	-0.0009 [0.0002]***	-0.0008 [0.0002]***	(0.0002)***
	Borrows in foreign currency (b)	0.0018 [0.0022]	(0.0024]	0.0013 [0.0012]	0.0015 [0.0012]	(0.0010)
	Charge to clear a check (a)	-0.1164 [0.0503]**	(0.0253]***	-0.1773 [0.0324]***	-0.1239 [0.0340]***	(0.0352)***
	Largest shareholder	0.0006 [0.0010]	(0.0008]	0.0000 [0.0006]	0.0008 [0.0007]	(0.0006)
	Working capital financed by foreign commercial banks (b)	0.0106 [0.0083]	(0.0084]	0.0241 [0.0070]***	0.0134 [0.0045]***	(0.0093)
	Working capital financed by informal sources (b)	-0.0022 [0.0023]	(0.0001]***	-0.0044 [0.0031]	-0.002 [0.0036]	(0.1225)
Quality, innovation and labor skills	Dummy for ISO quality certification (b)	0.1603 [0.0766]**	(0.0365]***	0.1208 [0.0359]***	0.1599 [0.0389]***	(0.0498)***
	Dummy for new product (b)	0.091 [0.0494]*	(0.0113]***	0.0322 [0.0377]	0.0807 [0.0398]**	(0.0398)**
	Dummy for discontinued product line (b)	-0.1007 [0.0610]	(0.0384]**	-0.0565 [0.0333]*	-0.0865 [0.0375]**	(0.0431)**
	Staff - management	0.004 [0.0028]	(0.0009]***	0.0047 [0.0016]***	0.0041 [0.0015]***	(0.0020)**
	Staff - non-production workers	-0.0034 [0.0022]	(0.0025]	-0.0027 [0.0011]**	-0.0033 [0.0012]***	(0.0014)**
	Training to unskilled workers (a)	0.001 [0.0026]	(0.0030]	-0.0048 [0.0032]	0.0012 [0.0041]	(0.0043)
	University staff (b)	0.0049 [0.0015]***	(0.0007]***	0.0036 [0.0015]**	0.0044 [0.0014]***	(0.0012)***
	Manager's experience (b)	0.0391 [0.0249]	(0.0217]*	0.0336 [0.0142]**	0.0369 [0.0150]**	(0.0173)**
Other control variables	Age (b)	0.0018 [0.0015]	(0.0016]	0.0016 [0.0009]*	0.0012 [0.0010]	(0.0011)
	Share of the local market (b)	0.0032 [0.0008]***	(0.0004]***	0.0028 [0.0006]***	0.0031 [0.0006]***	(0.0007)***
	Constant	2.7174 [0.8932]***		2.7155 [0.5500]***	2.7170 [0.6986]***	
	Industry/region/size/time dummies	Yes		Yes	Yes	
	Observations	1483		1443 (Censored: 2007/ Uncens: 1236)	1657 (Censored: 183/ Uncens.: 1484)	
	R-squared	0.89				
	Heckman's Lambda			-0.2747 [0.1993]	-0.2471 [0.2303]	

Notes of Table 11.1.

Source: Authors' calculations with ICSs.

Table 11.4: TANZANIA, Extended production function and comparison of ICA method with Heckman models

	Dependent variable: log of total sales	ICA Method[1]		Heckman models[2]		
				Heckman on complete case	Heckman replacing inputs	
Category	Variable	Coeff. std. err.	Boot. st. er.	Coeff. std. err.	Coeff. std. err.	Boot. st. er.
PF variables	Log-employment	0.1655 [0.0853]*	(0.0512)***	0.1422 [0.0557]**	0.1742 [0.0669]***	(0.0677)**
	Log-materials	0.4252 [0.0581]***	(0.0340)***	0.6176 [0.0274]***	0.6099 [0.0317]***	(0.0439)***
	Log-capital	0.1589 [0.0323]***	(0.0208)***	0.1427 [0.0209]***	0.1417 [0.0265]***	(0.0235)***
Infrastructure	Electricity from own generator (b)	0.0021 [0.0016]	(0.0053)	-0.001 [0.0018]	0.0041 [0.0020]**	(0.0017)**
	Losses due to water outages (b)	-0.0112 [0.0058]*	(0.0162)	-0.0081 [0.0060]	-0.0029 [0.0063]	(0.0054)
	Water from own well or water infrastructure (a)	0.0001 [0.0051]	(0.0011)	0.001 [0.0031]	0.0044 [0.0036]	(0.0042)
	Losses due to phone outages (a)	-0.0322 [0.0198]	(0.0071)***	-0.0315 [0.0284]	-0.0226 [0.0321]	(0.0291)
	Transport outages (a)	-0.0047 [0.0703]	(0.1168)	-0.1172 [0.0503]**	-0.0214 [0.0583]	(0.0499)
	Dummy for own roads (b)	0.289 [0.1488]*	(0.0581)***	0.3742 [0.1143]***	0.3416 [0.1444]**	(0.1321)***
	Dummy for webpage (b)	0.1578 [0.1212]	(0.1994)	0.3178 [0.0972]***	0.1595 [0.1208]	(0.1468)
	Wait for a water supply (a)	-0.1814 [0.0427]***	(0.0702)**	-0.1214 [0.0415]***	-0.1888 [0.0551]***	(0.0466)***
	Low quality supplies (a)	-0.0163 [0.0128]	(0.0041)***	-0.0252 [0.0116]**	-0.0323 [0.0118]***	(0.0130)**
Red tape, corruption and crime	Gift to obtain a operating license (b)	-0.4983 [0.1935]**	(0.1066)***	-0.1757 [0.1281]	0.0688 [0.1482]	(0.1589)
	Payments to deal with bureaucratic issues (b)	0.0939 [0.0299]***	(0.0164)***	0.0365 [0.0420]	0.0245 [0.0446]	(0.0495)
	Days in inspections (b)	-0.1045 [0.0735]	(0.0494)**	-0.1106 [0.0525]**	-0.0246 [0.0585]	(0.0580)
	Payments to obtain a contract with the government (b)	-0.0114 [0.0066]*	(0.0091)	-0.0332 [0.0088]***	-0.0101 [0.0074]	(0.0066)
	Security expenses (b)	-0.0119 [0.0042]***	(0.0092)	0.0068 [0.0108]	-0.0051 [0.0058]	(0.0052)
	Illegal payments in protection (b)	-0.0827 [0.0170]***	(0.1019)	-0.1209 [0.0478]**	-0.026 [0.0467]	(0.0493)
Finance and corporate governance	Interest rate of the loan (a)	-0.0109 [0.0145]	(0.0099)	0.0036 [0.0098]	-0.0074 [0.0115]	(0.0127)
	Working capital financed by commercial banks (b)	-0.0009 [0.0018]	(0.0012)	0.0000 [0.0014]	-0.003 [0.0016]*	(0.0015)**
	Working capital financed by leasing (b)	-0.0794 [0.0282]***	(0.0054)***	-0.0234 [0.0408]	-0.0473 [0.0096]***	(0.0806)
	Sales bought on credit (b)	-0.0014 [0.0012]	(0.0011)	-0.0029 [0.0012]**	0.0038 [0.0014]***	(0.0014)***
	Delay to clear a domestic currency wire (a)	-0.3418 [0.3273]	(0.0935)***	0.4842 [0.1853]***	0.1533 [0.1876]	(0.1996)
Quality, innovation and labor skills	Dummy for new product (b)	0.0429 [0.1063]	(0.2036)	-0.1942 [0.0850]**	-0.003 [0.1014]	(0.0951)
	Staff - skilled workers (b)	0.0026 [0.0023]	(0.0050)	0.0074 [0.0020]***	0.0092 [0.0026]***	(0.0024)***
	Workforce with computer (b)	0.0066 [0.0030]**	(0.0056)	0.0183 [0.0037]***	-0.0084 [0.0032]***	(0.0070)
Other control variables	Dummy for incorporated company (b)	0.2914 [0.2023]	(0.5683)	-0.2149 [0.3207]	0.1701 [0.3050]	(0.1810)
	Dummy for FDI (b)	0.1397 [0.1445]	(0.2844)	0.1752 [0.1051]*	-0.0289 [0.1426]	(0.1326)
	Constant	7.2978 [1.0168]***		3.8725 [0.7997]***	3.1102 [0.9936]***	
	Industry/region/size/time dummies	Yes		Yes	Yes	
	Observations	559		581 (Censored: 290/ Uncens: 291)	771 (Censored: 317/ Uncens: 454)	
	R-squared	0.88				
	Heckman's Lambda			-0.2747[0.1993]	-0.2471[0.2303]	

Notes of Table 11.1
Source: Authors' calculations with ICSs.

Table 12: Summary of results from estimating equation (1) under different imputation methods with respect to the ICA method case

			Complete case	ICA method & variations				EM algorithms			Multiple imputation	Heckman models		
				ICA met.	ICA met. (boot. s. e.)	Random ICA met.	ICA met. on inputs	EM alg. [1]	EM alg. [2]	EM alg. [3]		Heckman on complete case	Heckman replacing inputs	Heckman (boot. s.e)
India: Tables 9.1, 10.1 & 11.1	Input-output elasticities	Significant change in estimated elasticity?[3]	No	**-**	-	No	No	No	Yes (L, M)	Yes (L, M)	Yes (L)	No	No	-
		Change in significance?[3]	No	**-**	No	No	No	No	No	No	No	No	No	No
	IC variables [27 vars.]	Significant variables[1]	4, (3)	**6**	6, (2)	6, (3)	4, (1)	5, (0)	4, (1)	4, (1)	5, (2)	11, (7)	11, (6)	15, (10)
		Non-significant variables[2]	23, (5)	**21**	21, (2)	21, (4)	23, (3)	22, (1)	23, (3)	23, (3)	22, (2)	16, (2)	16, (1)	12, (1)
		Change in the direction of the effect?[3]	No	**-**	-	No	No	No	No	No	No	No	No	No
	Number of observations		3943	**5211**	-	5063	5134	5216	5175	5176	5262	4233	5407	-
	Significant Heckman's Lambda?		-	**-**	-	-	-	-	-	-	-	No	No	-
Turkey: Table 9.2, 10.2 & 11.2	Input-output elasticities	Significant change in estimated elasticity?[3]	Yes (M)	**-**	-	No	Yes (L, M, K)	Yes (M, L)	Yes (L, M, K)	Yes (L, M, K)	Yes (L, M)	Yes (L. K)	Yes (L. K)	-
		Change in significance?[3]	No	**-**	Yes (L)	Yes (L)	No	No	No	No	No	No	No	No
	IC variables [18 vars.]	Significant variables[1]	9, (3)	**10**	8, (2)	9, (0)	9, (0)	13, (3)	9, (2)	11, (4)	10, (2)	9, (3)	11, (1)	16, (6)
		Non-significant variables[2]	9, (4)	**8**	10, (4)	9, (1)	9, (1)	5, (0)	9, (2)	7, (2)	8, (2)	9, (2)	7, (0)	2, (0)
		Change in the direction of the effect?[3]	No	**-**	-	No	No	No	No	No	No	No	No	No
	Number of observations		792	**1684**	-	1684	1360	1679	1733	1733	1646	1941	2509	-
	Significant Heckman's Lambda?		-	**-**	-	-	-	-	-	-	-	No	No	-
South Africa: Table 9.3, 10.3 & 11.3	Input-output elasticities	Significant change in estimated elasticity?[3]	No	**-**	-	Yes (K)	No	No	Yes (L)	Yes (L)	Yes (L)	No	No	-
		Change in significance?[3]	No	**-**	No	No	No	No	No	No	No	No	No	No
	IC variables [31 vars.]	Significant variables[1]	10, (3)	**9**	16, (10)	12, (3)	9, (0)	12, (5)	14, (6)	14, (5)	15, (7)	15, (8)	19, (11)	18, (10)
		Non-significant variables[2]	21, (2)	**22**	15, (3)	19, (0)	22, (0)	19, (2)	17, (1)	17, (1)	16, (1)	16, (2)	12, (1)	13, (1)
		Change in the direction of the effect?[3]	No	**-**	-	No	No	No	No	No	No	No	No	-
	Number of observations			**1483**				1528	1552	1550		1443	1657	-
	Significant Heckman's Lambda?		-	**-**	-	-	-	-	-	-	-	No	No	-
Tanzania: Table 9.4, 10.4 & 11.4	Input-output elasticities	Significant change in estimated elasticity?[3]	Yes (M)	**-**	-	Yes (L, K)	Yes (L, M)	Yes (L)	Yes (M, L)	Yes (M, L)	Yes (L, M, K)	Yes (M)	Yes (M)	-
		Change in significance?[3]	No	**-**	No	No	No	Yes (L)	Yes (L)	Yes (L)	No	No	No	No
	IC variables [25 vars.]	Significant variables[1]	10, (4)	**10**	9, (4)	11, (4)	10, (2)	11, (2)	8, (2)	10, (3)	8, (2)	14, (9)	9, (5)	7, (5)
		Non-significant variables[2]	15, (3)	**15**	16, (5)	14, (3)	15, (2)	14, (1)	17, (4)	15, (3)	17, (4)	11, (5)	16, (6)	18, (6)
		Change in the direction of the effect?[3]	No	**-**	-	No	No				No	No	No	-
	Number of observations		291	**559**	-	557	442	560	603	597	570	581	771	-
	Significant Heckman's Lambda?		-	**-**	-	-	-	-	-	-	-	No	No	-

[1] In parenthesis: variables non-significant in the ICA method that became significant under other imputation mechanisms.
[2] In parenthesis: variables significant in the ICA method and no longer significant under other imputation mechanisms.
[3] With respect to ICA method.
A more detailed description of the results is in tables 8.1 to 8.4.
Source: Authors' calculations with ICSs data.

Figure 1.1: INDIA, Patterns of missing values in PF variables

Sales	Materials	Capital	Labor	# of m.v	# of obs.	% of obs.
Yellow	Yellow	Yellow	Yellow	0	4631	67.6
Yellow	Yellow	White	Yellow	1	1113	16.3
White	White	White	Yellow	3	913	13.3
Yellow	White	White	Yellow	2	89	1.3
White	White	Yellow	Yellow	2	47	0.7
White	Yellow	White	Yellow	2	28	0.4
Yellow	White	Yellow	Yellow	1	18	0.3
White	Yellow	Yellow	Yellow	1	10	0.1

Figure 1.2: TURKEY, Patterns of missing values in PF variables

Sales	Materials	Capital	Labor	# of m.v	# of obs.	% of obs.
Yellow	Yellow	Yellow	Yellow	0	818	30.9
White	White	White	Yellow	3	737	27.9
Yellow	Yellow	White	Yellow	1	345	13.0
White	White	Yellow	Yellow	2	189	7.1
Yellow	White	Yellow	Yellow	1	185	7.0
Yellow	White	White	Yellow	2	133	5.0
White	White	White	White	4	96	3.6
White	Yellow	Yellow	Yellow	1	87	3.3
White	Yellow	White	Yellow	2	35	1.3
Yellow	White	White	White	3	6	0.2
Yellow	Yellow	White	White	2	5	0.2
White	White	Yellow	White	3	5	0.2
Yellow	Yellow	Yellow	White	1	3	0.1
Yellow	White	Yellow	White	2	2	0.1

Figure 1.3: SOUTH AFRICA, Patterns of missing values in PF variables

Sales	Materials	Capital	Labor	# of m.v	# of obs.	% of obs.
Yellow	Yellow	Yellow	Yellow	0	1265	69.9
Yellow	Yellow	White	Yellow	1	220	12.2
White	White	White	White	4	123	6.8
White	White	White	Yellow	3	99	5.5
Yellow	White	Yellow	Yellow	1	47	2.6
Yellow	White	White	Yellow	2	24	1.3
Yellow	Yellow	Yellow	White	1	17	0.9
White	White	Yellow	Yellow	2	7	0.4
Yellow	Yellow	White	White	2	4	0.2
White	Yellow	Yellow	Yellow	1	1	0.1
White	Yellow	White	Yellow	2	1	0.1
Yellow	White	White	White	3	1	0.1

Figure 1.4: TANZANIA, Patterns of missing values in PF variables

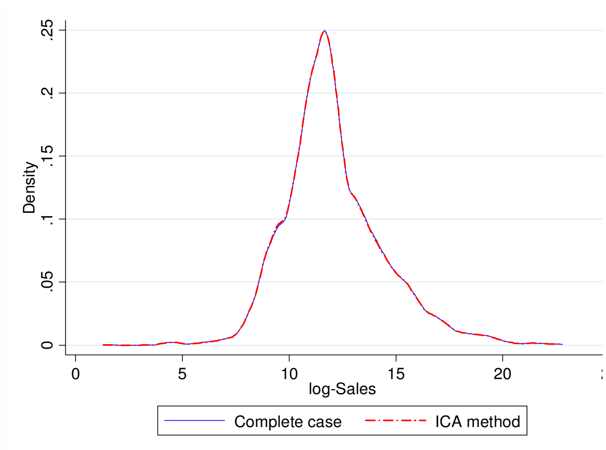
Sales	Materials	Capital	Labor	# of m.v	# of obs.	% of obs.
Yellow	Yellow	Yellow	Yellow	0	313	37.8
White	White	White	Yellow	3	130	15.7
White	Yellow	Yellow	Yellow	1	81	9.8
Yellow	Yellow	White	Yellow	1	74	8.9
Yellow	White	Yellow	Yellow	1	51	6.2
Yellow	Yellow	Yellow	White	1	38	4.6
White	White	White	White	4	37	4.5
White	White	Yellow	Yellow	2	30	3.6
Yellow	White	White	Yellow	2	26	3.1
White	Yellow	White	Yellow	2	25	3.0
White	White	Yellow	White	3	9	1.1
Yellow	White	Yellow	White	2	5	0.6
Yellow	Yellow	White	White	2	3	0.4
White	Yellow	White	White	3	3	0.4
White	Yellow	Yellow	White	2	2	0.2
Yellow	White	White	White	3	1	0.1

Notes:
Yellow means information available on the corresponding variable. White means information is missing.
Source: Authors' calculations with ICS data.

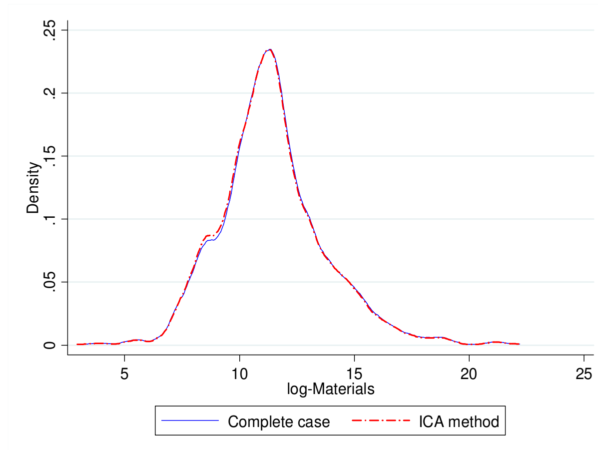
Figure 2.1: INDIA, evaluation of performance of ICA method

I. Kernel[1] estimates of output and inputs densities in the complete case and in the sample after imputing missing values by the ICA method

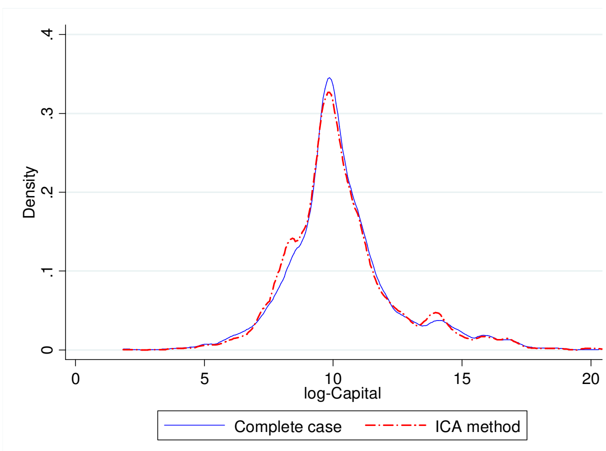
A. Sales (log)



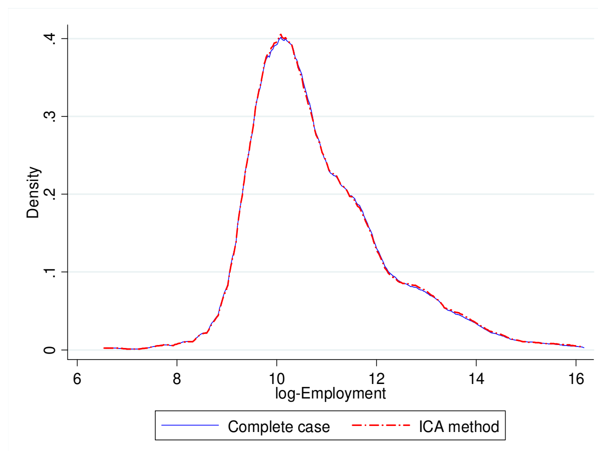
B. Materials (log)



C. Capital stock (log)



D. Employment (log)



II. Table of descriptive statistics and tests of equality of distributions of output and inputs in the complete case and in the sample with imputation by the ICA method

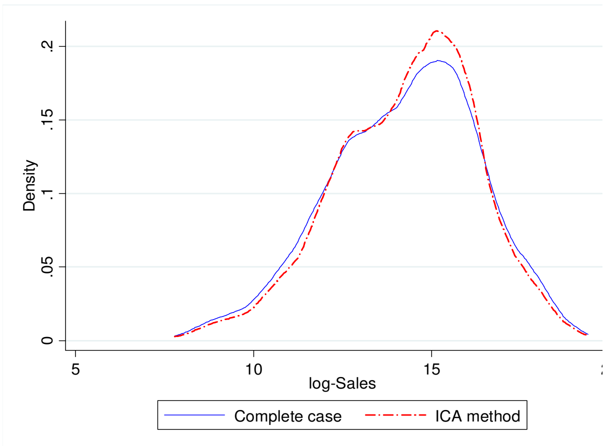
		# Obs. (# imputed)	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**One-sample K-S Test (p-value)**
Sales (log)	Complete case	5841	12.08	2.30	1.30	22.79	0.000
	ICA meth.	5935 (94)	12.07	2.29	1.30	22.79	0.000
Materials (log)	Complete case	5597	11.44	2.30	2.94	22.20	0.000
	ICA meth.	5933 (336)	11.40	2.28	2.94	22.20	0.000
Capital (log)	Complete case	4555	10.31	2.11	1.85	20.73	0.000
	ICA meth.	5918 (1363)	10.28	2.10	1.85	20.73	0.000
Empl (log)	Complete case	6164	10.82	1.33	6.54	16.16	0.000
	ICA meth.	6321 (157)	10.82	1.34	6.54	16.16	0.000

Notes:
1 Epanechnikov kernel. Each point estimated within a range of 300 values.
The null hypothesis of the one-sample Kolmogorov-Smirnov Test is that the cumulative distribution differs from the hypothesized theoretical normal distribution.
Source: Authors' estimations with ICSs data.

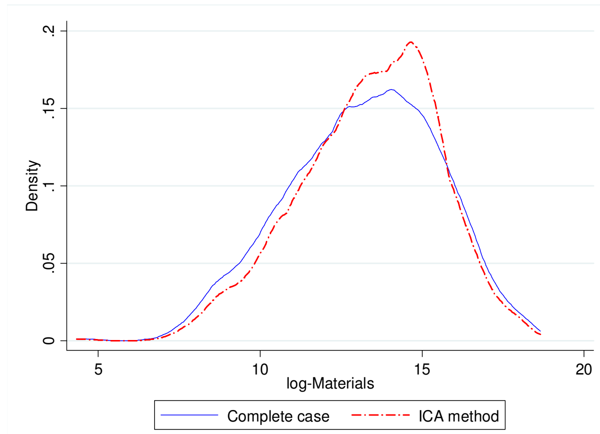
Figure 2.2: TURKEY, evaluation of performance of ICA method

I. Kernel[1] estimates of output and inputs densities in the complete case and in the sample after imputing missing values by the ICA method

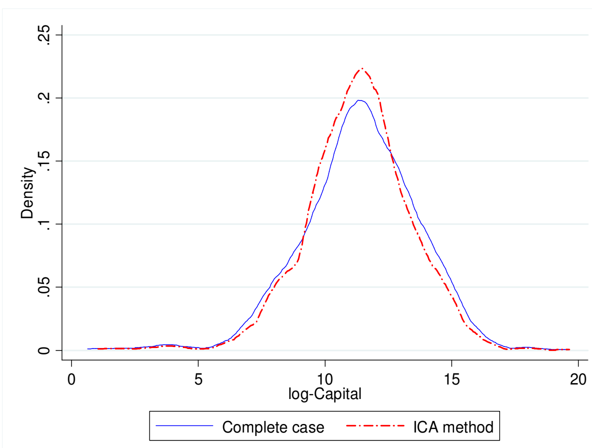
A. Sales (log)



B. Materials (log)



C. Capital stock (log)



D. Employment (log)

II. Table of descriptive statistics and tests of equality of distributions of output and inputs in the complete case and in the sample with imputation by the ICA method

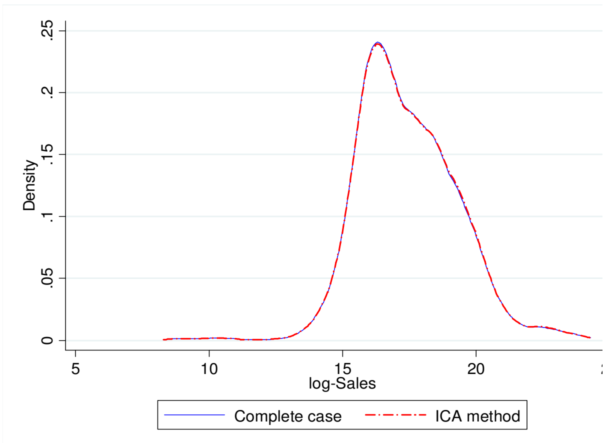
		# Obs. (# imputed)	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**One-sample K-S Test (p-value)**
Sales (log)	Complete case	1497	14.24	2.10	7.78	19.40	0.004
	ICA meth.	1821 (324)	14.30	1.99	7.78	19.40	0.000
Materials (log)	Complete case	1293	13.19	2.31	4.33	18.65	0.020
	ICA meth.	1822 (529)	13.37	2.13	4.34	18.65	0.000
Capital (log)	Complete case	1289	11.39	2.26	0.63	19.65	0.015
	ICA meth.	1816 (527)	11.32	2.05	1.05	19.65	0.004
Empl (log)	Complete case	2529	11.63	1.45	7.64	15.42	0.001
	ICA meth.	2548 (19)	11.63	1.45	7.64	15.42	0.001

Notes:
[1] Epanechnikov kernel. Each point estimated within a range of 300 values.
The null hypothesis of the one-sample Kolmogorov-Smirnov Test is that the cumulative distribution differs from the hypothesized theoretical normal distribution.
Source: Authors' estimations with ICSs data.

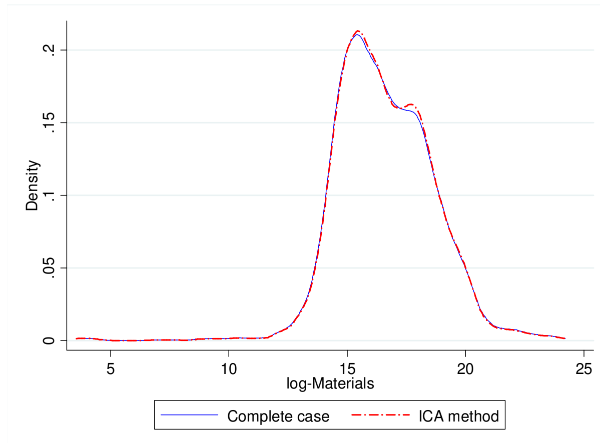
Figure 2.3: SOUTH AFRICA, evaluation of performance of ICA method

I. Kernel[1] estimates of output and inputs densities in the complete case and in the sample after imputing missing values by the ICA method

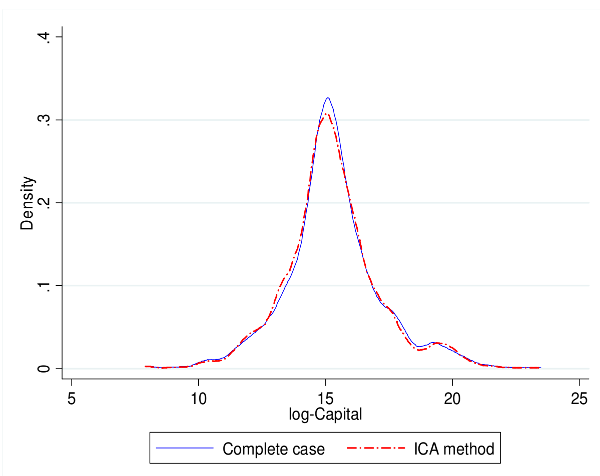
A. Sales (log)



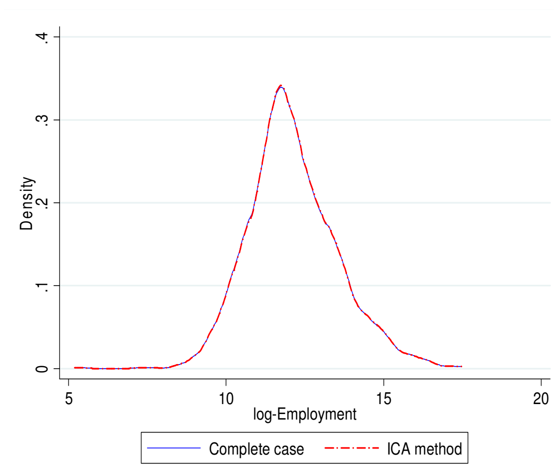
B. Materials (log)



C. Capital stock (log)



D. Employment (log)



II. Table of descriptive statistics and tests of equality of distributions of output and inputs in the complete case and in the sample with imputation by the ICA method

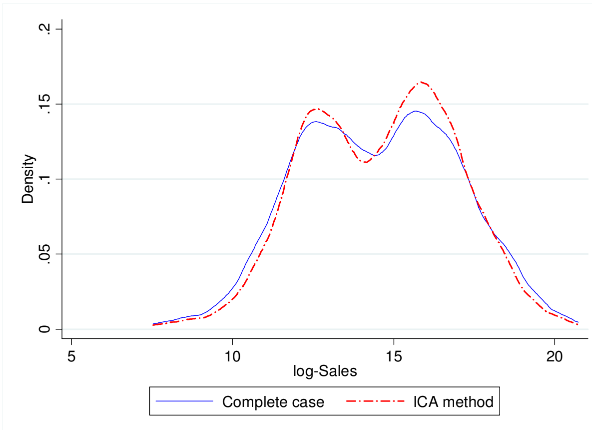
		# Obs. (# imputed)	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**One-sample K-S Test (p-value)**
Sales (log)	Complete case	1497	14.24	2.10	7.78	19.40	0.000
	ICA meth.	1821 (324)	14.30	1.99	7.78	19.40	0.000
Materials (log)	Complete case	1293	13.19	2.31	4.33	18.65	0.000
	ICA meth.	1822 (529)	13.37	2.13	4.34	18.65	0.000
Capital (log)	Complete case	1289	11.39	2.26	0.63	19.65	0.000
	ICA meth.	1816 (527)	11.32	2.05	1.05	19.65	0.000
Empl (log)	Complete case	2529	11.63	1.45	7.64	15.42	0.000
	ICA meth.	2548 (19)	11.63	1.45	7.64	15.42	0.000

Notes:
[1] Epanechnikov kernel. Each point estimated within a range of 300 values.
The null hypothesis of the one-sample Kolmogorov-Smirnov Test is that the cumulative distribution differs from the hypothesized theoretical normal distribution.
Source: Authors' estimations with ICSs data.

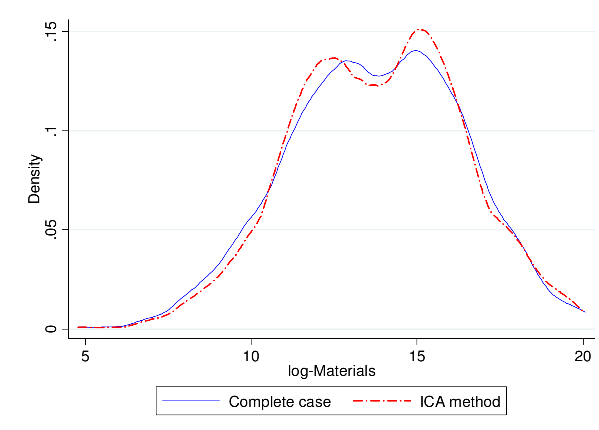
Figure 2.4: TANZANIA, evaluation of performance of ICA method

I. Kernel[1] estimates of output and inputs densities in the complete case and in the sample after imputing missing values by the ICA method

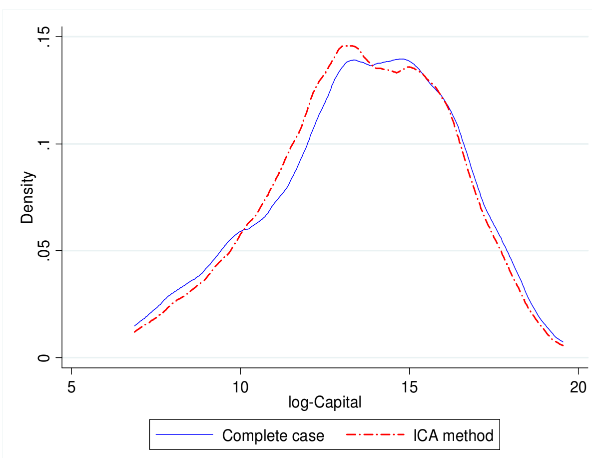
A. Sales (log)



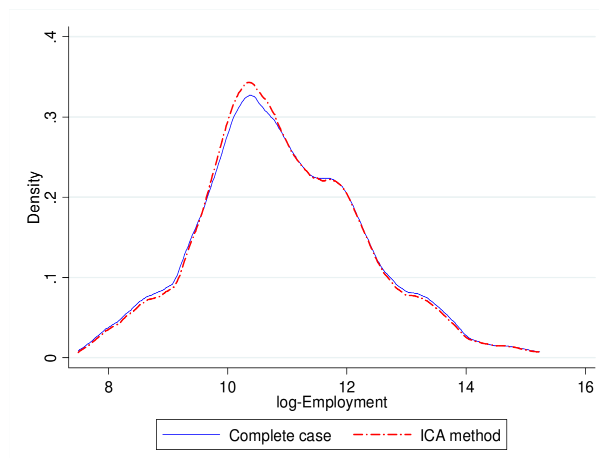
B. Materials (log)



C. Capital stock (log)



D. Employment (log)



II. Table of descriptive statistics and tests of equality of distributions of output and inputs in the complete case and in the sample with imputation by the ICA method

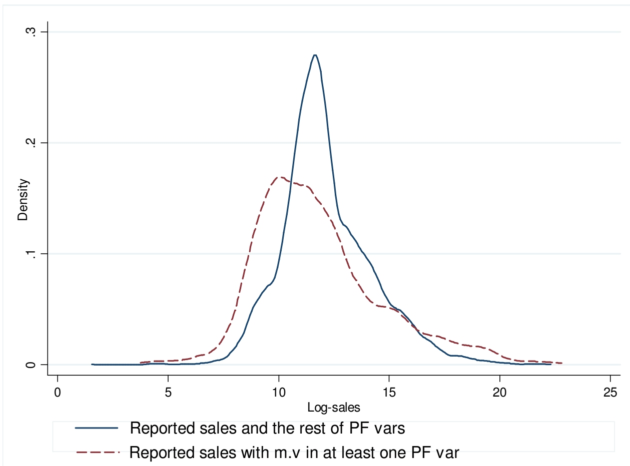
		# Obs. (# imputed)	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**One-sample K-S Test (p-value)**
Sales (log)	Complete case	1497	14.24	2.10	7.78	19.40	0.012
	ICA meth.	1821 (324)	14.30	1.99	7.78	19.40	0.001
Materials (log)	Complete case	1293	13.19	2.31	4.33	18.65	0.169
	ICA meth.	1822 (529)	13.37	2.13	4.34	18.65	0.093
Capital (log)	Complete case	1289	11.39	2.26	0.63	19.65	0.053
	ICA meth.	1816 (527)	11.32	2.05	1.05	19.65	0.027
Empl (log)	Complete case	2529	11.63	1.45	7.64	15.42	0.006
	ICA meth.	2548 (19)	11.63	1.45	7.64	15.42	0.002

Notes:
[1] Epanechnikov kernel. Each point estimated within a range of 300 values.
The null hypothesis of the one-sample Kolmogorov-Smirnov Test is that the cumulative distribution differs from the hypothesized theoretical normal distribution.
Source: Authors' estimations with ICSs data.

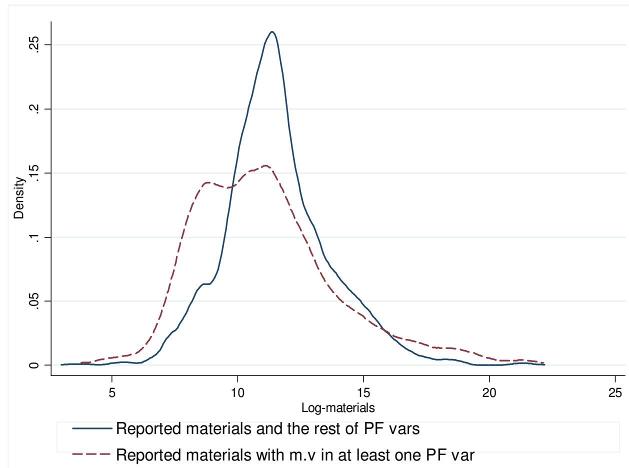
Figure 3.1: INDIA, Kernel density estimates of PF variables

(without M.V in PF variables and with M.V in any PF variable)

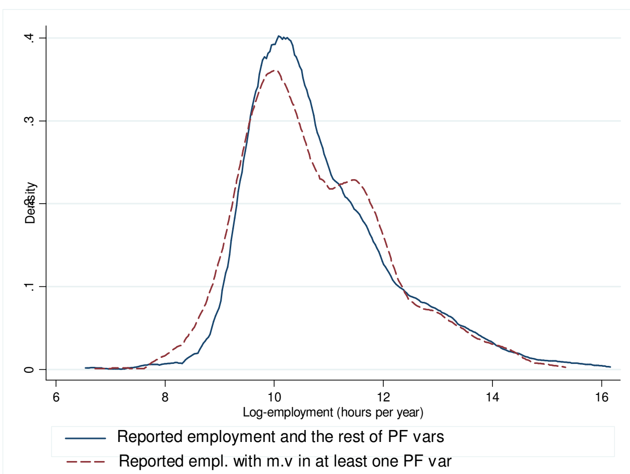
A. Sales



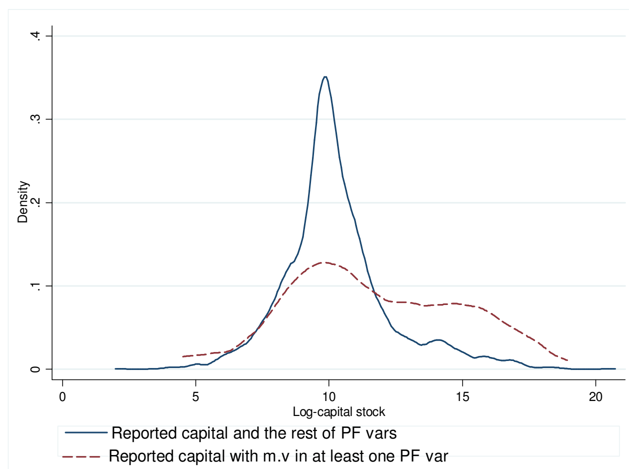
B. Materials



C. Labor



D. Capital

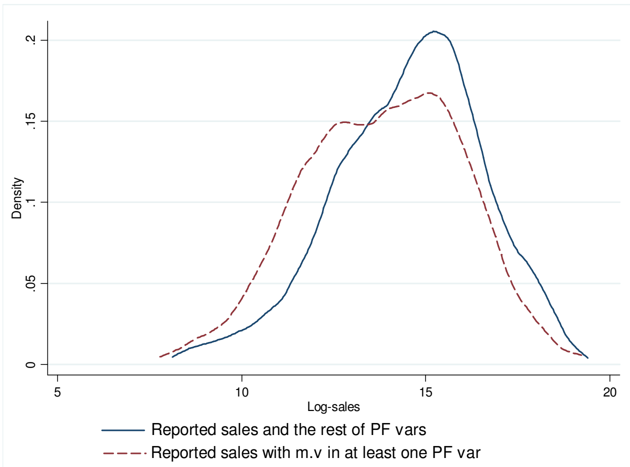


Notes:
Reported X and the rest of PF variables: is the distribution of those establishments reporting all PF variables
Reported X with m.v in at least one of the rest of P.F: is the distribution of those establishments reporting the corresponding PF variable and also reporting at least one missing value in the remaining PF variables
Epanechnikov kernel. Each point estimated within a range of 300 values.
Source: Authors' estimations with ICSs data.

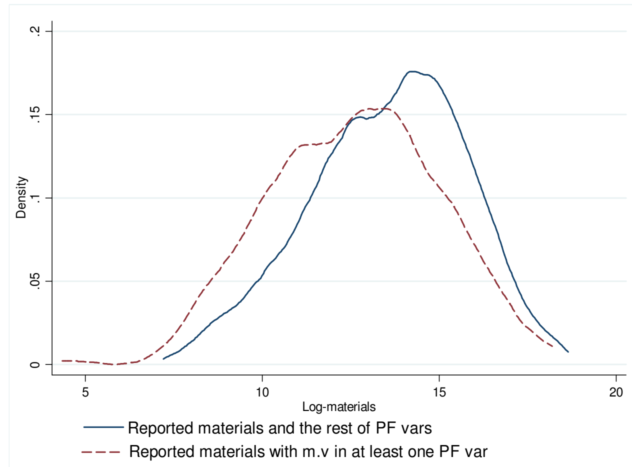
Figure 3.2: TURKEY, Kernel density estimates of PF variables

(without M.V in PF variables and with M.V in any PF variable)

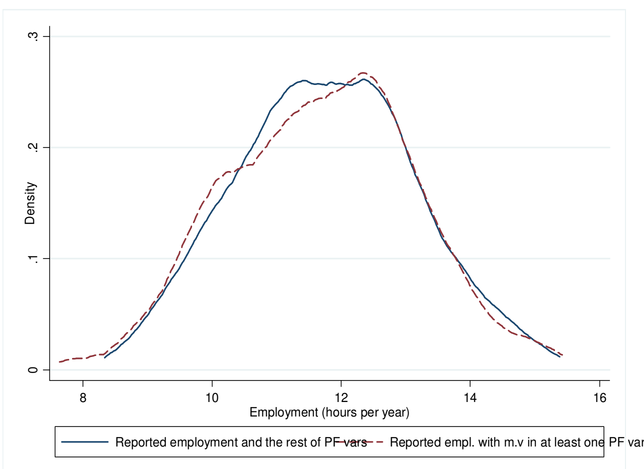
A. Sales



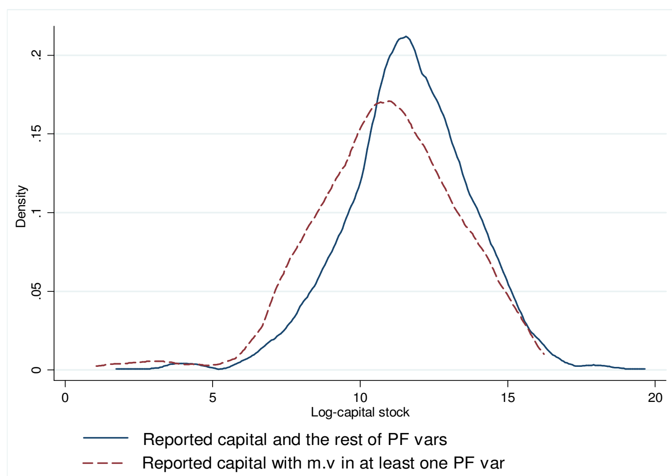
B. Materials



C. Labor



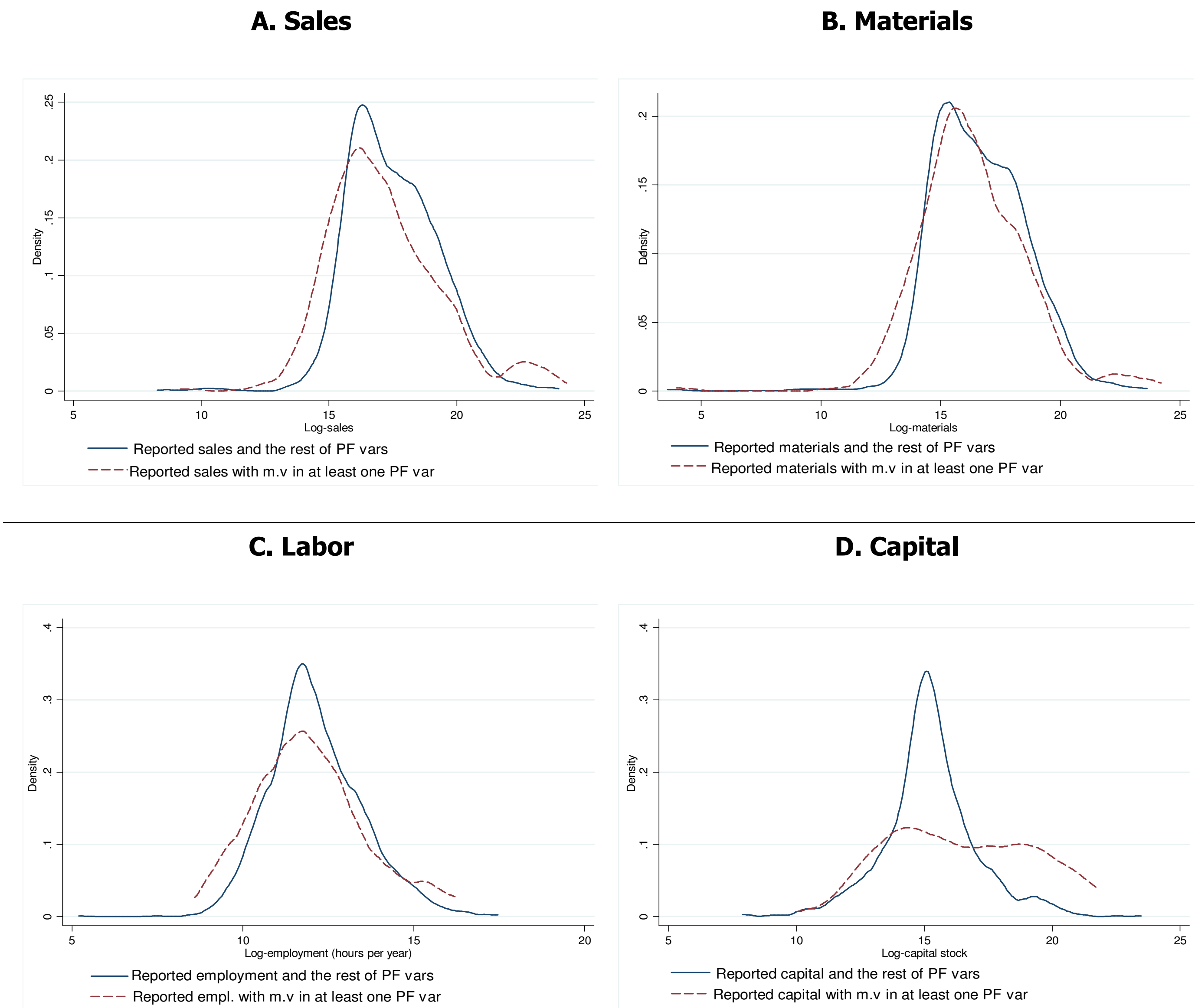
D. Capital



Notes:
Reported X and the rest of PF variables: is the distribution of those establishments reporting all PF variables
Reported X with m.v in at least one of the rest of P.F: is the distribution of those establishments reporting the corresponding PF variable and also reporting at least one missing value in the remaining PF variables
Epanechnikov kernel. Each point estimated within a range of 300 values.
Source: Authors' estimations with ICSs data.

Figure 3.3: SOUTH AFRICA, Kernel density estimates of PF variables

(without M.V in PF variables and with M.V in any PF variable)

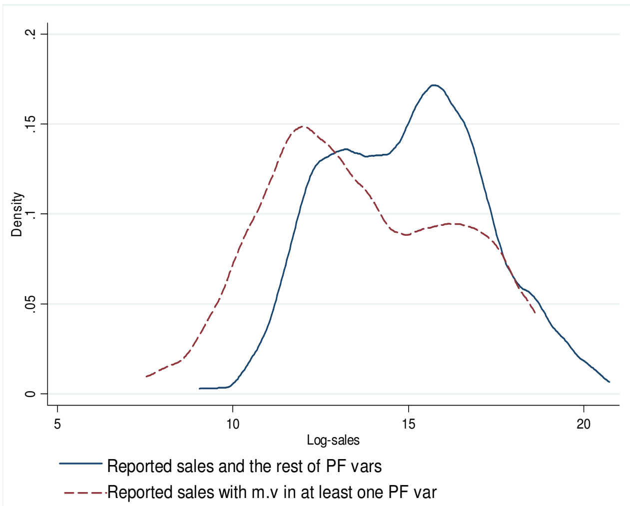


Notes:
Reported X and the rest of PF variables: is the distribution of those establishments reporting all PF variables
Reported X with m.v in at least one of the rest of P.F: is the distribution of those establishments reporting the corresponding PF variable and also reporting at least one missing value in the remaining PF variables
Epanechnikov kernel. Each point estimated within a range of 300 values.
Source: Authors' estimations with ICSs data.

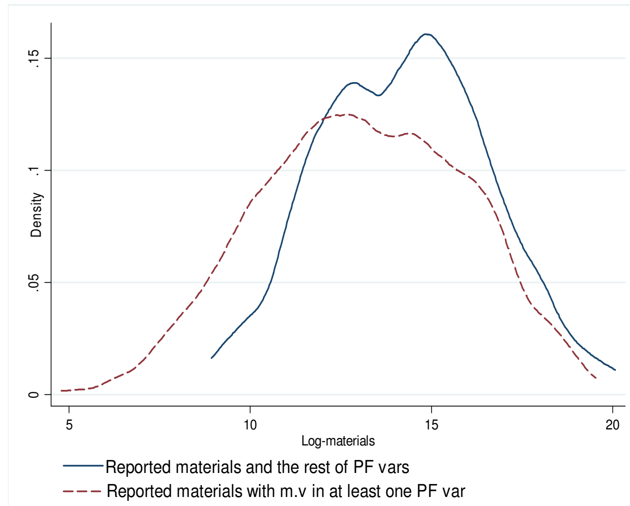
Figure 3.4: TANZANIA, Kernel density estimates of PF variables in Tanzania

(without M.V in PF variables and with M.V in any PF variable)

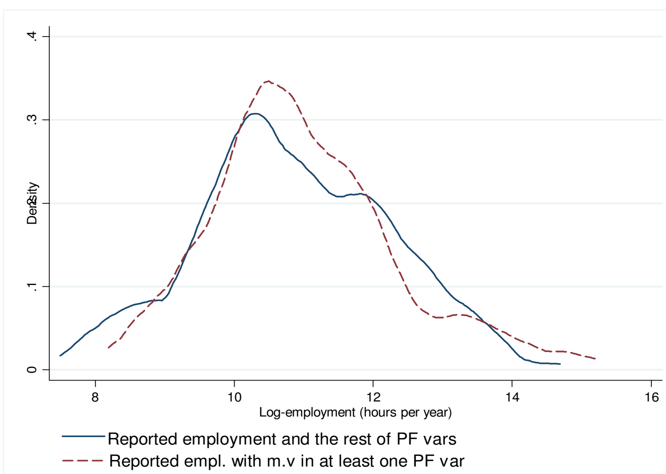
A. Sales



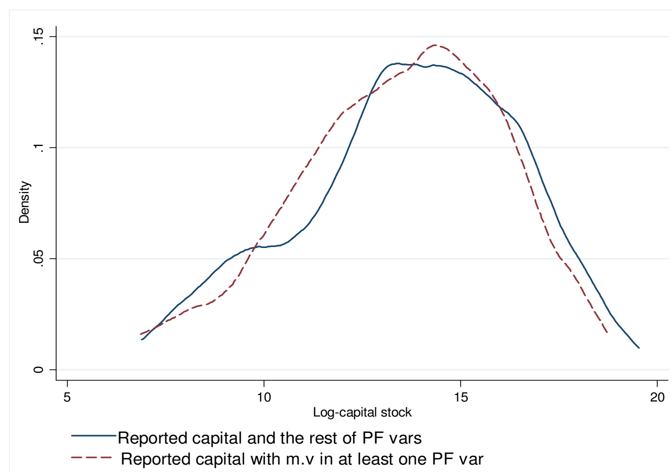
B. Materials



C. Labor



D. Capital

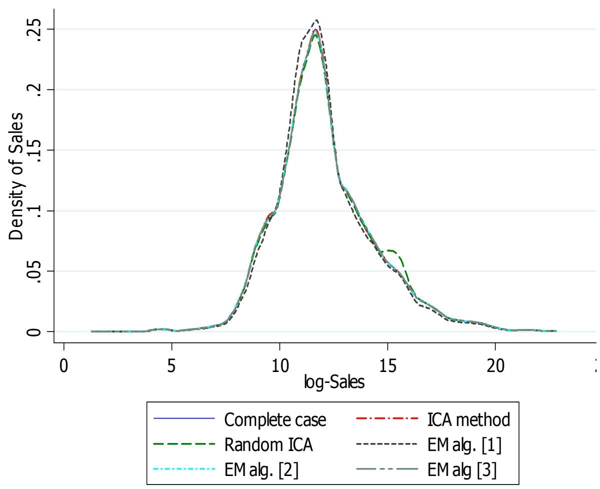


Notes:
Reported X and the rest of PF variables: is the distribution of those establishments reporting all PF variables
Reported X with m.v in at least one of the rest of P.F: is the distribution of those establishments reporting the corresponding PF variable and also reporting at least one missing value in the remaining PF variables
Epanechnikov kernel. Each point estimated within a range of 300 values.
Source: Authors' estimations with ICSs data.

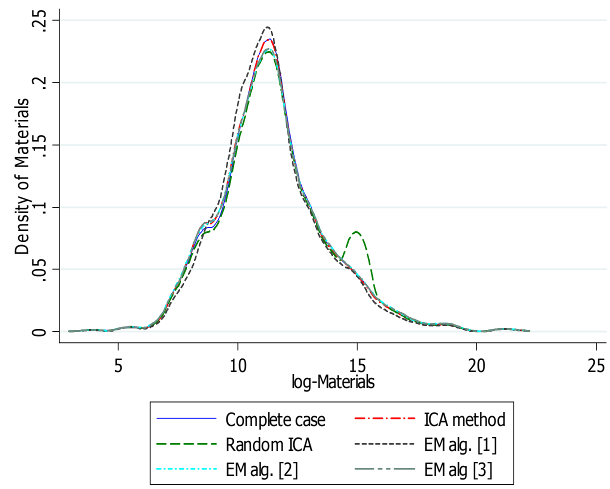
Figure 4.1: INDIA, comparison of of ICA method and other imputation mechanisms for PF variables

I. Kernel[1] estimates of output and inputs densities

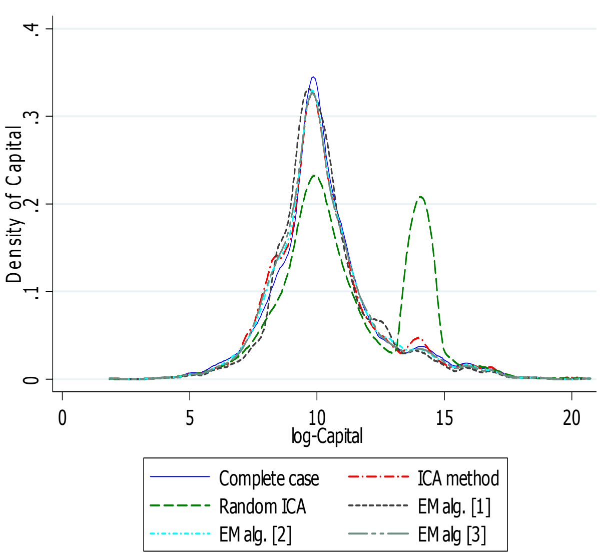
A. Sales (log)



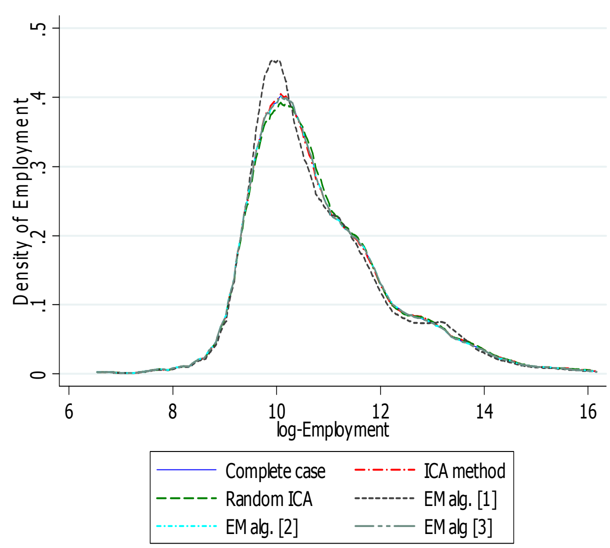
B. Materials (log)



C. Capital stock (log)



D. Employment (log)

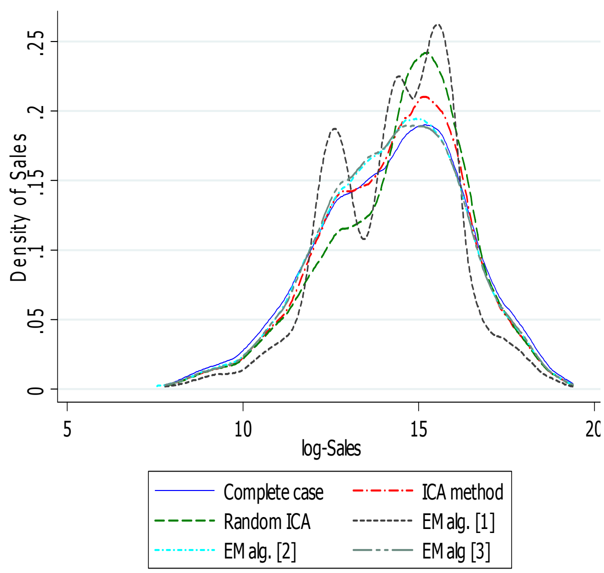


Notes:
1 Epanechnikov kernel. Each point estimated within a range of 300 values.
Source: Authors' estimations with ICSs data.

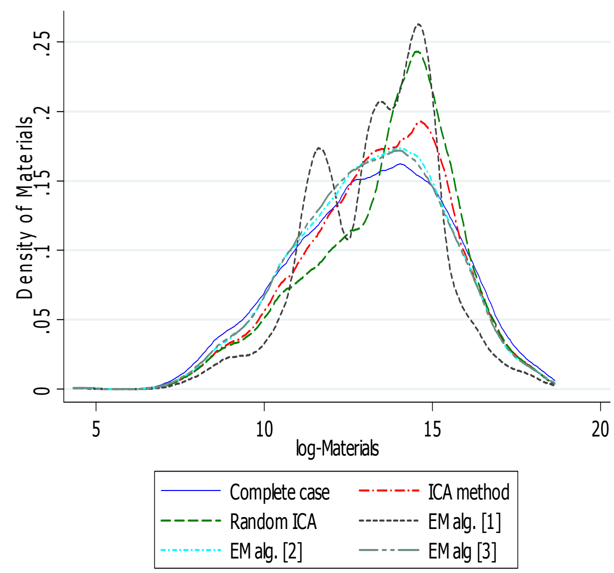
Figure 4.2: TURKEY, comparison of of ICA method and other imputation mechanisms for PF variables

I. Kernel[1] estimates of output and inputs densities

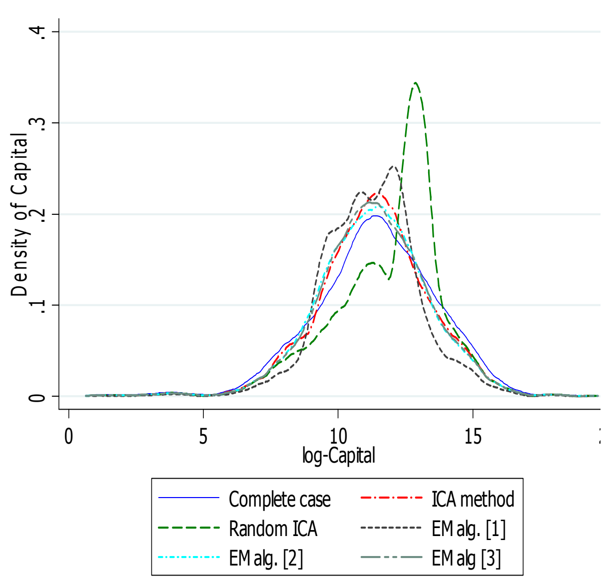
A. Sales (log)



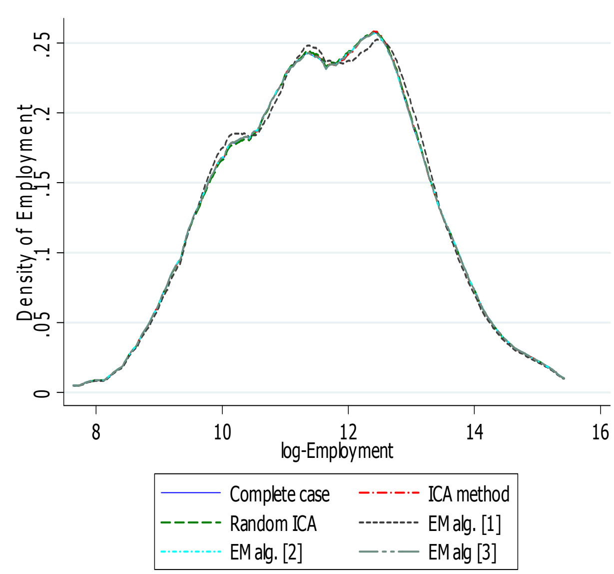
B. Materials (log)



C. Capital stock (log)



D. Employment (log)

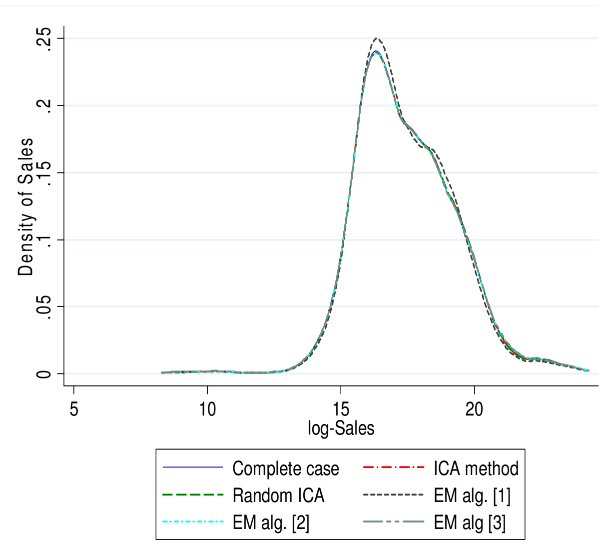


Notes:
[1] Epanechnikov kernel. Each point estimated within a range of 300 values.
Source: Authors' estimations with ICSs data.

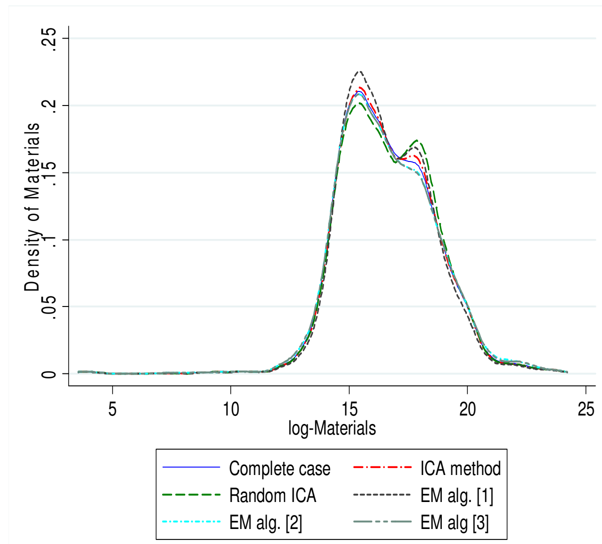
Figure 4.3: SOUTH AFRICA, comparison of of ICA method and other imputation mechanisms for PF variables

I. Kernel[1] estimates of output and inputs densities

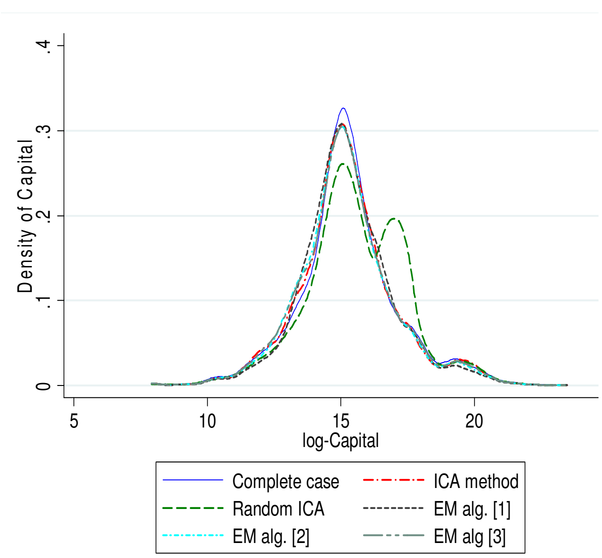
A. Sales (log)



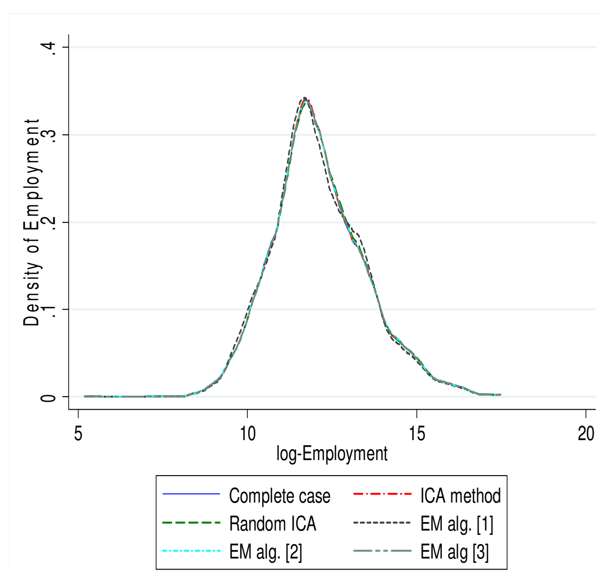
B. Materials (log)



C. Capital stock (log)



D. Employment (log)

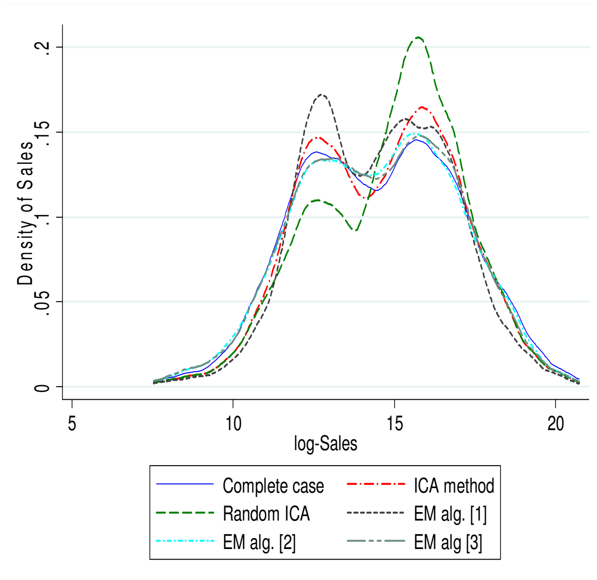


Notes:
[1] Epanechnikov kernel. Each point estimated within a range of 300 values.
Source: Authors' estimations with ICSs data.

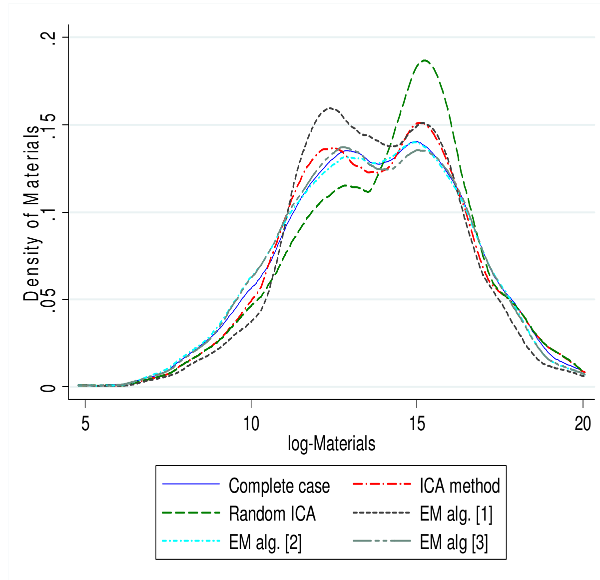
Figure 4.4: TANZANIA, comparison of of ICA method and other imputation mechanisms for PF variables

I. Kernel[1] estimates of output and inputs densities

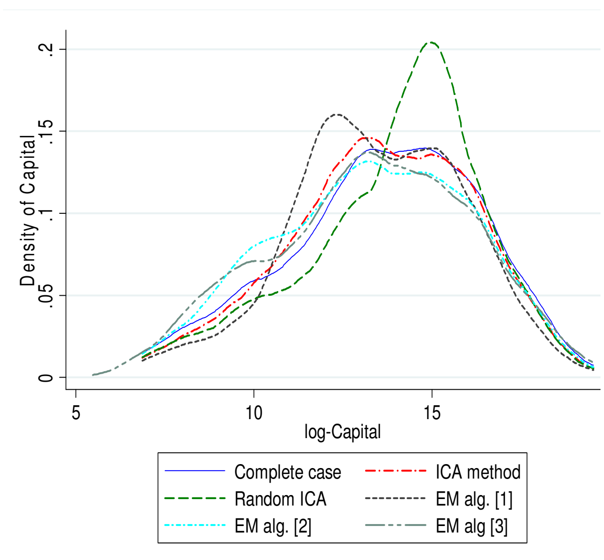
A. Sales (log)



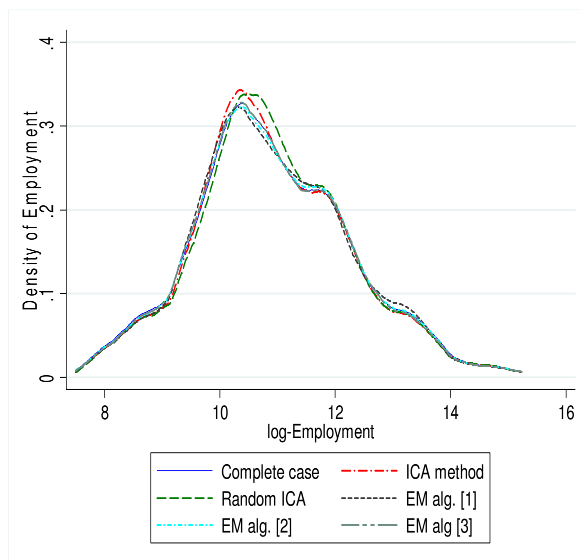
B. Materials (log)



C. Capital stock (log)



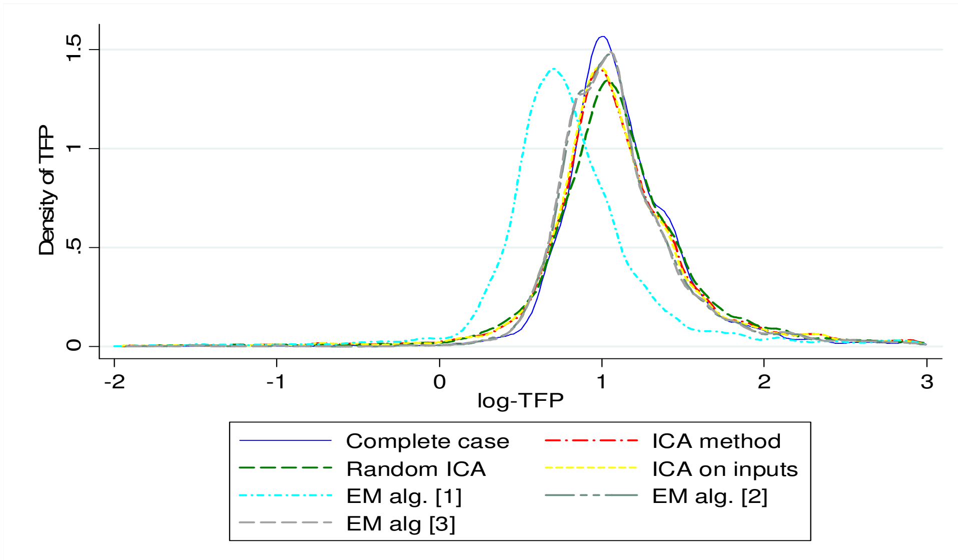
D. Employment (log)



Notes:
[1] Epanechnikov kernel. Each point estimated within a range of 300 values.
Source: Authors' estimations with ICSs data.

Figure 5.1: INDIA, evaluation of TFP measures under different imputation methods

I. Kernel[1] estimates of TFP densities



II. Table of descriptive statistics of TFP measures

	# Obs	Mean	Std. Dev.	Min	Max
Complete case	4327	1.15	0.68	-12.51	12.08
ICA meth.	5915	1.17	0.98	-12.53	12.19
Random ICA	5915	1.10	1.19	-12.51	12.15
ICA on inputs	5821	1.16	0.95	-12.55	12.25
Em alg [1]	6848	0.83	0.90	-12.96	12.09
Em alg [2]	5731	1.13	0.71	-12.66	12.44
Em alg [3]	5731	1.13	0.71	-12.67	12.43

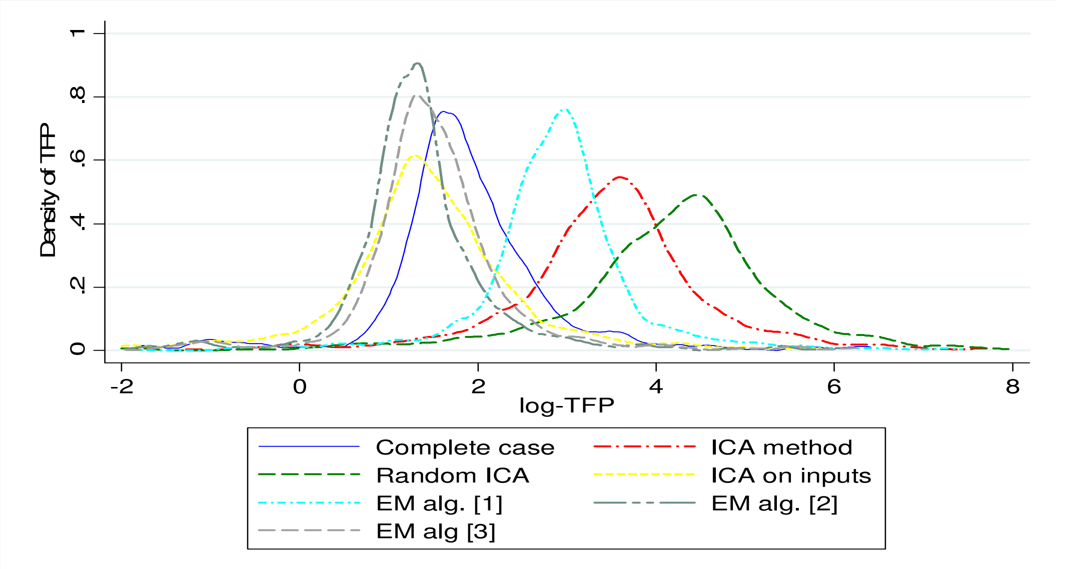
III. Correlation matrix between TFP measures

	Complete case	ICA meth.	Random ICA	ICA on inputs	Em alg [1]	Em alg [2]	Em alg [3]
Complete case	1.000						
ICA meth.	0.999	1.000					
Random ICA	1.000	0.999	1.000				
ICA on inputs	0.998	1.000	0.998	1.000			
Em alg [1]	0.993	0.995	0.994	0.996	1.000		
Em alg [2]	0.990	0.991	0.993	0.993	0.997	1.000	
Em alg [3]	0.990	0.991	0.993	0.993	0.997	1.000	1.000

Notes:
[1] Epanechnikov kernel. Each point estimated within a range of 300 values.
Complete case: TFP measure from the sample without replacement of missing values; likewise, input-output elasticities are obtained from estimating equation (1) in the complete case (see I-O elasticities in Table 9.1).
ICA method: TFP measure with inputs and output replaced by *ICA method* and input-output elasticities from Table 8.1.
Random ICA: TFP measure with inputs and output replaced by *random ICA method* and input output elasticities from Table 9.1.
ICA on inputs: TFP measure when only inputs are imputed by the ICA method (not sales), the I-O elasticities and semi-elasticities used are in Table 9.1.
Em alg. [1]: TFP measure obtained under imputation of inputs and output by the EM algorithm described in section 5.1.1. Likewise, the I-O elasticities are in Table 9.1.
Em alg. [2]: In this case the EM algorithm used is that described in section 5.1.2. The I-O elasticities are in Table 9.1.
Em alg. [3]: The description of the EM algorithm is in section 5.1.3, the I-O elasticities in Table 9.1.
Source: Authors' estimations with ICSs.

Figure 5.2: TURKEY, evaluation of TFP measures under different imputation methods

I. Kernel[1] estimates of TFP densities



II. Table of descriptive statistics of TFP measures

	# Obs	Mean	Std. Dev.	Min	Max
Complete case	818	1.84	1.01	-5.25	6.41
ICA meth.	1805	3.45	1.20	-3.36	7.85
Random ICA	1805	4.16	1.28	-2.67	8.91
ICA on inputs	1481	1.37	1.23	-5.64	5.87
Em alg [1]	2646	2.87	0.97	-4.05	7.44
Em alg [2]	1802	1.33	0.88	-5.84	6.13
Em alg [3]	1802	1.51	0.88	-5.65	6.31

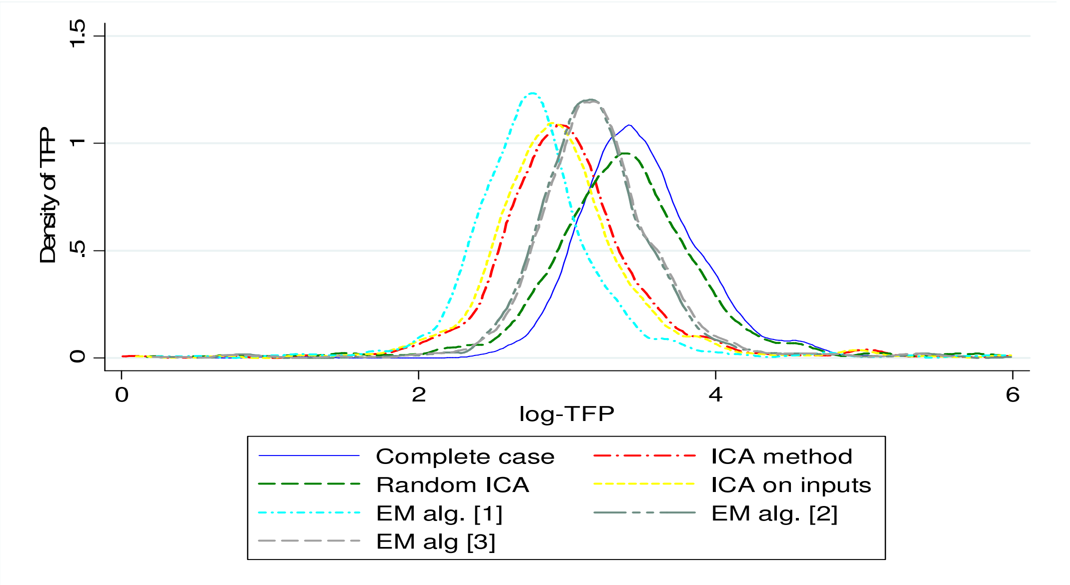
III. Correlation matrix between TFP measures

	Complete case	ICA meth.	Random ICA	ICA on inputs	Em alg [1]	Em alg [2]	Em alg [3]
Complete case	1.000						
ICA meth.	0.969	1.000					
Random ICA	0.954	0.998	1.000				
ICA on inputs	0.992	0.974	0.956	1.000			
Em alg [1]	0.990	0.993	0.986	0.985	1.000		
Em alg [2]	0.990	0.927	0.908	0.969	0.964	1.000	
Em alg [3]	0.991	0.932	0.914	0.969	0.968	1.000	1.000

Notes:
[1] Epanechnikov kernel. Each point estimated within a range of 300 values.
Complete case: TFP measure from the sample without replacement of missing values; likewise, input-output elasticities are obtained from estimating equation (1) in the complete case (see I-O elasticities in Table 9.2).
ICA method: TFP measure with inputs and output replaced by *ICA method* and input-output elasticities from Table 8.2.
Random ICA: TFP measure with inputs and output replaced by *random ICA method* and input output elasticities from Table 9.2.
ICA on inputs: TFP measure when only inputs are imputed by the ICA method (not sales), the I-O elasticities and semi-elasticities used are in Table 9.2.
Em alg. [1]: TFP measure obtained under imputation of inputs and output by the EM algorithm described in section 5.1.1. Likewise, the I-O elasticities are in Table 9.2.
Em alg. [2]: In this case the EM algorithm used is that described in section 5.1.2. The I-O elasticities are in Table 9.2.
Em alg. [3]: The description of the EM algorithm is in section 5.1.3, the I-O elasticities in Table 9.2.
Source: Authors' estimations with ICSs.

Figure 5.3: SOUTH AFRICA, evaluation of TFP measures under different imputation methods

I. Kernel[1] estimates of TFP densities



II. Table of descriptive statistics of TFP measures

	# Obs	Mean	Std. Dev.	Min	Max
Complete case	1265	3.50	0.70	-3.74	10.34
ICA meth.	1585	2.99	0.84	-4.34	10.28
Random ICA	1585	3.38	0.90	-4.97	10.31
ICA on inputs	1576	2.94	0.84	-4.39	10.21
Em alg [1]	1784	2.78	0.80	-4.47	10.26
Em alg [2]	1581	3.21	0.72	-4.01	11.21
Em alg [3]	1578	3.22	0.72	-4.00	11.18

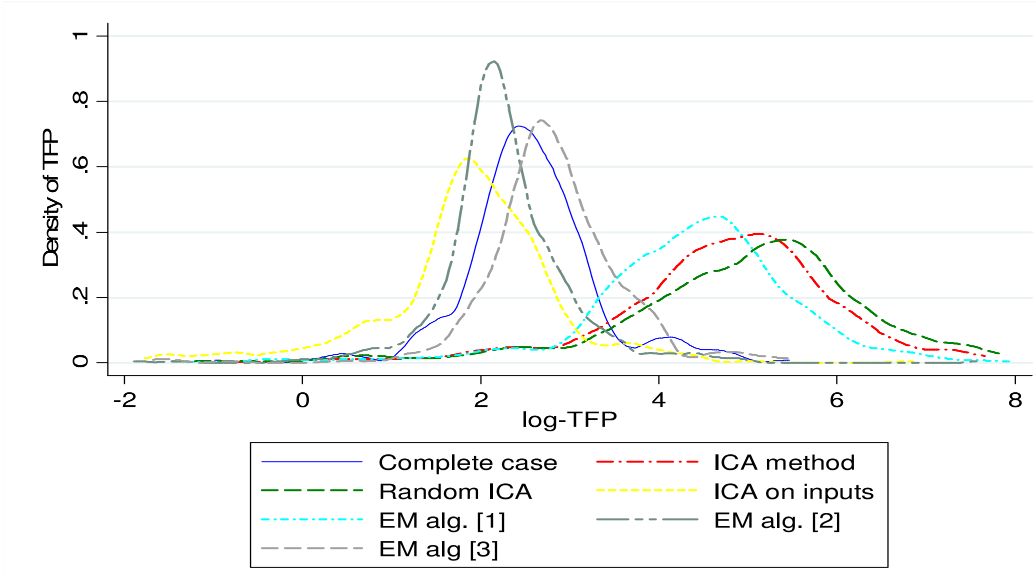
III. Correlation matrix between TFP measures

	Complete case	ICA meth.	Random ICA	ICA on inputs	Em alg [1]	Em alg [2]	Em alg [3]
Complete case	1.000						
ICA meth.	0.996	1.000					
Random ICA	0.998	0.993	1.000				
ICA on inputs	0.996	1.000	0.993	1.000			
Em alg [1]	0.992	0.999	0.988	0.999	1.000		
Em alg [2]	0.982	0.991	0.975	0.990	0.992	1.000	
Em alg [3]	0.982	0.991	0.975	0.990	0.993	1.000	1.000

Notes:
[1] Epanechnikov kernel. Each point estimated within a range of 300 values.
Complete case: TFP measure from the sample without replacement of missing values; likewise, input-output elasticities are obtained from estimating equation (1) in the complete case (see I-O elasticities in Table 9.3).
ICA method: TFP measure with inputs and output replaced by *ICA method* and input-output elasticities from Table 8.3.
Random ICA: TFP measure with inputs and output replaced by *random ICA method* and input output elasticities from Table 9.3.
ICA on inputs: TFP measure when only inputs are imputed by the ICA method (not sales), the I-O elasticities and semi-elasticities used are in Table 9.3.
Em alg. [1]: TFP measure obtained under imputation of inputs and output by the EM algorithm described in section 5.1.1. Likewise, the I-O elasticities are in Table 9.3.
Em alg. [2]: In this case the EM algorithm used is that described in section 5.1.2. The I-O elasticities are in Table 9.3.
Em alg. [3]: The description of the EM algorithm is in section 5.1.3, the I-O elasticities in Table 9.3.
Source: Authors' estimations with ICSs.

Figure 5.4: TANZANIA, evaluation of TFP measures under different imputation methods

I. Kernel[1] estimates of TFP densities



II. Table of descriptive statistics of TFP measures

	# Obs	Mean	Std. Dev.	Min	Max
Complete case	313	2.53	0.87	-3.21	5.47
ICA meth.	661	4.79	1.30	-0.75	9.72
Random ICA	661	4.98	1.50	-1.47	10.03
ICA on inputs	505	1.81	1.14	-3.68	6.85
Em alg [1]	790	4.39	1.18	-1.30	8.92
Em alg [2]	628	2.25	0.80	-3.82	7.64
Em alg [3]	638	2.81	0.86	-3.21	8.35

III. Correlation matrix between TFP measures

	Complete case	ICA meth.	Random ICA	ICA on inputs	Em alg [1]	Em alg [2]	Em alg [3]
Complete case	1.000						
ICA meth.	0.913	1.000					
Random ICA	0.877	0.991	1.000				
ICA on inputs	0.997	0.904	0.869	1.000			
Em alg [1]	0.948	0.994	0.975	0.937	1.000		
Em alg [2]	0.981	0.829	0.779	0.971	0.884	1.000	
Em alg [3]	0.979	0.849	0.804	0.963	0.901	0.996	1.000

Notes:

[1] Epanechnikov kernel. Each point estimated within a range of 300 values.

Complete case: TFP measure from the sample without replacement of missing values; likewise, input-output elasticities are obtained from estimating equation (1) in the complete case (see I-O elasticities in Table 9.4).

ICA method: TFP measure with inputs and output replaced by *ICA method* and input-output elasticities from Table 8.4.

Random ICA: TFP measure with inputs and output replaced by *random ICA method* and input output elasticities from Table 9.4.

ICA on inputs: TFP measure when only inputs are imputed by the ICA method (not sales), the I-O elasticities and semi-elasticities used are in Table 9.4.

Em alg. [1]: TFP measure obtained under imputation of inputs and output by the EM algorithm described in section 5.1.1. Likewise, the I-O elasticities are in Table 9.4.

Em alg. [2]: In this case the EM algorithm used is that described in section 5.1.2. The I-O elasticities are in Table 9.4.

Em alg. [3]: The description of the EM algorithm is in section 5.1.3, the I-O elasticities in Table 9.4.

Source: Authors' estimations with ICSs.