Université Catholique de Louvain (UCLouvain)
Faculté des Sciences Economique, Sociale, Politique et de Communication
Economic School of Louvain (ESL)
Institut de Recherches Economiques et Sociales (IRES)

Informal Economy: Development and Distributional Impacts

Douglas AMULI Ibale

Thèse Présentée en vue de l'obtention du grade de docteur en sciences économiques et de gestion

Composition du jury :

Prof. Dejemeppe Muriel (Promotrice), UCLouvain - Belgique
Prof. Docquier Frédéric (Promoteur), LISER - Luxembourg
Prof. Bucekuderhwa Bashige Célestin, UCB - DR Congo
Prof. Parienté William, UCLouvain - Belgique
Prof. Tenikue Michel, LISER - Luxembourg

Louvain-La-Neuve, Belgique, 27 Octobre 2021

To Leslie,
My blessing,
Whose presence continuously makes me better.

To my sons Keynes, Karl and Lévi,
I wish you to do more better than this.

Acknowledgements

Four years ago, when I started the Ph.D., I looked forward to the day I would put into writing these lines of acknowledgement with enthusiasm, because it would mean that I managed to complete the last mile of my Ph.D. journey. These years of work, though stressful in terms of effort required and challenges to overcome, were enriching in terms of people met, ideas shared and experience gained. Being at the end, I want to thank those who stood by me throughout this journey, helping me directly or indirectly to carry out my research.

As the thesis deals with the Labor force, my first thanks go to the almighty God for his abundant grace to me and to My Parents; Pacifique A. and Françoise M. for their hard work and affection that have considerably shaped who I am. All my achievements are also theirs.

Next, I would like to express my deepest gratitude to my supervisors Frédéric Docquier and Muriel Dejemeppe. More than supervisors, I consider Frédéric and Muriel as mentors from whom I have learnt a great deal, not only as a researcher but also as a person. I owe you a lot. You gave me the space to explore my own research interests and I could always count on your inputs, guidance, mentoring and encouragement. I consider myself extremely lucky to have such amazing and supportive supervisors. I am also grateful to the members of my doctoral jury Professors Célestin Bucekuderhwa Bashige, William Parienté and Michel Tenikue for their excellent comments and suggestions which were particularly helpful in finalizing the last version of this thesis.

Pursuing postgraduate studies in economics was not my initial choice. However, the switch to this field has always brought me happiness and satisfaction until now. This would not be possible without the precious influence incitement and help of Professors Augustin Mutabazi, Célestin Bucekuderhwa, Déogratias Bugandwa, Eddy Balemba, Jean-Baptiste Ntagoma and Albert Lukuitshi.

I would not feel confident to be granted a doctorate in economics without the various knowledge gained from the teaching opportunities that I have been offered by the Université Catholique de Bukavu (UCB). I would like to thank UCB authorities for the position and for giving feedback on my performance. I thank the economics faculty members for the stimulating discussions and valuable help for my research. On the other hand, without the work experience at the "Cellule d'Analyse des Indicateurs de Développement (CAID)", I would feel less comfortable being called an expert in my research field. Therefore, I am grateful to Mr. Gregoire MWEPU, the coordinator, for the repeated work opportunity at CAID and his comprehensive attitude. Many thanks to all my Colleagues at CAID, especially those from the South-Kivu. Working for the country with you has been enriching for my Ph.D. journey.

I am immensely grateful to the UCLouvain community, and especially those at IRES, including but not exclusively Riccardo Turati, Keiti Kondi, Fabio Blasutto, Adam Levai, François Courtoy, Guillermo Santos, Daniele Verdini, Ehui Adovor and so many others for whom I am forever grateful. I have spent wonderful moments, learning and engaging with kind, intelligent and compassionate young economists and friends.

I would like to thank my lovely family. My sisters and brothers: Esther, Pamela, Idris, Akochris, Solide, Amina, and Meilleur for their guiding words and unconditional love that have always steadied my pace. I also thanks my in-laws brothers and sisters-in-law for our frequent conversations that soothed my emotions and sparked my intuitions.

To have endured this endeavor and dared to enjoy it, I cannot thank my friends enough. The Ph.D. journey is, to a large extent, an emotional journey, a wobbly roller coaster that friends can help steady. For this, I am eternally grateful to all my friends whose company and heart-felt conversations had soothed my soul and sparked my intrigue. It was also the love, friendship and jokes of you Fabrice K., Patrick K., Charles K., Pépin M. and Arsène M. that have pulled me out of my bubble so many times.

Finally, and I save the best for last, I would not have endured this journey or reached the end of it without the unrelenting support of my wife, Leslie Foussiakda, whose faith in me had never waned, and whose partnership brought me courage and relief more times than I can count. It is her presence in my life that makes this thesis taste so sweet. During my PhD. Journey and while she was working, she managed to take care, to love and to support four men at home; I and and our three sons Keynes, Karl and Lévi. She is for me the illustration of strength, courage, patience, kindness and love. Whatever claim I have to this thesis is also hers.

Contents

List of Figures

List of Tables

General Introduction

Due to an increasing role of the government by the late 1970s, the level of taxation had risen sharply and regulations had proliferated in many countries creating strong incentives for individuals and firms to work informal. Governments became convinced that many normal economic activities, that should be measured and taxed, were presumably escaping the government regulations. Therefore, the interest in the informal economy has gradually moved from the market regulators' reports into the pages of academic reviews. Reflecting this scholarly interest, several books have been published and conferences have been organized. The reason for studying the informal economy is self-evident because overlooking informality might result in distortion biases in the level of national output, employment rate and other macro indicators, which in turn misleads pictures of wealth, income disparities and labor market, creating bias in international comparisons. Although informality is of a relative small share in advanced economies, it is a widespread phenomenon in developing countries. It accounts for up to half of economic activity and provides livelihood for billions of people. Yet its role in economic development remains controversial.

This PhD dissertation contributes to the literature on the determinants of the size of the informal economy, and on its development and distributional consequences. The thesis consists of three chapters, two of which being single-authored. The thesis first provides a cross-country study providing new estimates of the evolution of the informal sector, and analyze how informality affects convergence between countries and changes in the world distribution of income. Second, the thesis uses empirical methods to characterize the composition of the formal and informal sectors, and to analyze how the earnings structure is affected by informal activities. Third, the thesis proposes a structural, two-sector model with labor market frictions. The model is calibrated on each province of the DRC and used to analyze the effectiveness of different types of development policies in the presence of a large informal sector. These three chapters are summarized as follow :

Chapter 1 aims to re-estimate the size of the informal economy worldwide and study its trends and patterns. Thereafter, it examines whether accounting for the informal output affects the cross-country convergence and the evolution of global inequality. Using the

Multiple Indicators Multiple Causes model, the empirical procedure generates estimates of the informal economy that exhibits a declining trend over time in relative terms. As informal activities keep declining overtime, the gap between the trend in official and actual/total output is decreasing as well. It is then shown that, accounting for the informal output revises downward the cross-country income differentials and the level of inequality without, however, significantly affecting the mobility of economies within the world distribution of income.

Chapter 2 uses a unique broad individual, household and expenditure survey data on the DRC. The initial descriptive statistics of the labor market highlight five different sectors with two "higher-paid" that are completely formal and two "lower-paid" that are largely informal. Based on a linear regression result, the study reports a significant heterogeneity across them when it comes to earnings. With an unconditional quantile regression methodology corrected for selectivity bias the results show that, though the effect of education on earnings provides a clear support to the human capital theory, basic education has no significant impact on earnings in higher-paid sectors. Likewise, tertiary education matters for earnings in lower-paid sectors as well. Then the study decompose the earning gap across sectors and show that workers in the lower-paid sectors earn less not only because they are less skill endowed but also because they earn lower returns on such skills. However, when higher-paid and lower paid sectors are concerned, the coefficient effect at the upper end of the distribution is negative. Implying that the labor market provides an "informal employment earning premium" to some workers of the lower-paid sectors who, given their characteristics, would not earn more in the higher-paid sectors.

Chapter 3, coauthored with Zainab Iftikhar and Frédéric Docquier, builds a two-sector model with labor market frictions to explain income disparities between provinces, sectors (formal vs. informal) and skill groups in the Democratic Republic of Congo. The model is parameterized to exactly match the observed labor allocation of workers across sectors and income distribution. The calibration reveals large differences across provinces, both in observed characteristics and identified parameters. Using the model, a set of counterfactual "policy" experiments is conducted to analyze the role of technologies in the formal and informal sectors, human capital, public infrastructure and labor market frictions in explaining spatial and within-province inequalities. The study first quantify the high level of complementarity between policies, identify o-ring patterns of spatial inequality, and shed light on the role of labor market frictions. Second, It shows that spatial inequalities are mostly determined by technological disparities, reflecting both endowment in mineral resources, geographic position and institutional quality. Third, the study finds that a development policy that disregards the situation of the informal sector has low or even detrimental effects on inequality and extreme poverty. In particular, policies targeting education, labor market frictions, or public infrastructure in isolation have little effects as they mostly impact productivity in the formal sector, and reduce the skill ratio and productivity in informality, where many unskilled workers are trapped.

Overall, the thesis provides novel evidence on the development and distributional impacts of the informal economy. In combining empirical and theoretical approaches, it contributes to the description of a worldwide phenomenon that varies between 40 to 70 percent of the official GDP and provides insights on how it shapes the labor market in a context of developing countries where many unskilled workers are trapped in informality.

Chapter 1

Informality and the World Distribution of Income[1]

[1]We would like to thank Dejemeppe Muriel, Frédéric Docquier, Célestin Bucekuderhwa Bashige, William Parienté and Michel Tenikue for their useful comments. We are grateful towards UCLouvain for supporting this research through the "Fond de coopération au développement"

1.1 Introduction

Although informality is of a relative small share in advanced economies, it is a widespread phenomenon in developing countries. It includes unreported incomes from the production of legal goods and services in the informal sector and represents a sizeable part of the economic activity that goes unrecorded in official statistics. As a result, it is often overlooked in the growing literature on the world distribution of income, usually based solely on official statistics. In doing so, the literature excludes a worldwide phenomenon that varies between 40 to 70 percent of the official GDP in developing countries (Schneider and Enste, 2000). Though some of these activities may be illegal, others are legal and socially valuable (Méon et al., 2011). They should therefore be taken into account when measuring a countries' output (Gutiérrez_Romero et al., 2006). This daunting figure calls into question the policy implication of empirical studies that ignore this phenomenon, especially because developing countries are also known to be more diverging and leading the cross-country poverty and inequality worldwide. Thus, overlooking informal output might result in downward biases in the level of national output and other macro indicators, which in turn misleads pictures of wealth and income disparities, creating bias in international comparisons (Tanzi, 1999).

In this paper, we use the Multiple Indicators, Multiple Causes (MIMIC) model to estimate the size of the informal economy over the period 1995-2013 for 120 countries. Thereafter, we account for the informal output to revisit the literature on income convergence and global inequality. To take into account the debate related to the endogeneity of the GDP in the MIMIC model, we develop different specifications of the model. The benchmark specification model completely excludes the GDP as a cause or as an indicator. The other specifications use the GDP or its proxy, the nighttime light intensity, as a measure of global activity. Based on these different specifications, we estimate the size of the informal economy and examine the trend and patterns, emphasizing the differences in distributions between the benchmark specification and the other specifications. Thereafter, we construct a Total Output indicator that accounts for the official and informal output. We then use it to examine whether accounting for the informal output (i) affects the income differentiation across countries, (ii) affects the mobility of the different economies within the distribution of world income, and (iii) changes the cross-country trend and patterns of inequality, as measured by the Theil index.

The increase in informal activities has, as reported by empirical studies, cast doubt on the accuracy of official statistics (Tanzi, 1999). This concern has raised the importance to explain the origin, the behavior over time and the size of informal activities for which, the larger it is, the more distorted might be the estimates of the official economy. Estimating the informal part of the economy however, still a difficult task because individuals who are engaged in such activities avoid to be identified. Nevertheless, the extent and the developments of the informal economy over time and its political and economic relevance have motivated the development of different tools aiming to estimate this unknown part of the economy. This includes direct approaches such as survey methods as well as indirect ones such as the transaction approach. Recently, the MIMIC model, applied to time series data, was derived as a special type of structural equation model with latent variables. The model analyzes covariance structures between observable cause and indicator variables to derive information about the informal economy taken as a latent variable. The model however, relies on the GDP variable and this has exposed it to critics. On the one hand, using the GDP as cause and indicator is said to distort the correlation structure of the model (Breusch, 2005). On the other hand, the link between formal and informal economy has proven to be ambiguous (Schneider and Enste, 2000). Based on this, Breusch (2005) castigates the fact that all of the connections that the indicator variables have with the causal variables have to be carried through the latent variable. Dell'Anno and Schneider (2006) prove that the connection does not affect the general appropriateness of MIMIC method. In examining the different approaches aiming to measure the informal economy, Dell'Anno (2003) and Trebicka (2014) show that, among others, the MIMIC approach could be considered as a relatively robust methodology in that field.

Our empirical investigation adds to the body of literature that has reviewed existed measures to estimate the informal economy or to assess the world distribution of income. For example, with the aim to estimate the size of the informal sector, Medina and Schneider (2018) address the critics of using the "GDP" in the MIMIC model. In so doing, they rely on the nighttime light intensity as a proxy of global activities. Their findings showed that, informal economy and global activity are negatively correlated. However, we know from Henderson et al. (2012) that, nighttime light is a measure of GDP. Dell'Anno (2008) transforms Schneider (2005) and Schneider (2007)'s estimates of the informal economy as percentage of official GDP into unofficial GDP per capita measured on purchasing power parity (PPP). The aim is to investigate the relationship between unofficial and official output. In so doing, he reports a positive and quantitatively important effect of the unofficial economy on the official economy and vice versa. Gennaioli et al. (2014) gather satellite data on nighttime lights as a proxy for real GDP on 1,528 regions from 83 countries to compare the speed of the per-capita income convergence within and across countries. Sala-i Martin (2006) estimates the world distribution of income by combining national accounts GDP per capita to anchor the mean with survey data in order to pin down the dispersion and after to assess the world distribution of income. However, most of these studies use formal output figures basically to assess the world distribution of income. nonetheless, the total economic activity, including formal and informal production of goods and services is important in the design of economic policies that respond to fluctuations and economic development over time and across space. This is what this study is all about.

The contribution of this study is twofold, First it contributes to the debate on the andogeneity issue of using the GDP in the MIMIC model. It emphasizes the trend of the distributions over time for different specifications of the model when GDP is or is not accounted for. As far as we know, this fact has not been previously noted in the literature. we develop different specifications of the MIMIC model. In some, we disregard the GDP, in others; we use it as a cause or as an indicator variable there after we test for the difference in the trend of their distribution. Secondly, using the benchmark specification, we account for both the formal and the informal output in a total output indicator. Thereafter, we revisit the convergence and inequality across countries using both to the official (formal) and the total output indicators (in per-capita term). As far as countries have different levels of informality, accounting for informal output can revise the level of cross-country income disparities and can affect the redistribution of economies within the distribution of world income.

With the help of different specifications, our empirical procedure predicts different series of estimates of informal economy that are, however, highly correlated. Furthermore, thought without the use of the GDP, the estimates of our benchmark specification keep the same distribution as that of GDP-based specifications. We then show that, the size of the informal economy varies significantly across countries where, high-income countries tend to be relatively of small size. Worldwide, the informal economy exhibits a declining trend over time and the larger is the initial size for a country, the higher is the declining rate. In comparison with the official output, accounting for the informal output revises downward the level of cross-country income disparities. Focusing on the beta-convergence, we notice that, although bottom countries seem to diverge at a slower pace and top countries to converge at a raising pace, accounting for the informal output does not significantly affect the mobility of economies within the distribution of the world income. Within the framework of total output indicator, the level of global inequality across countries is lower worldwide. However, as the informal output declines over time, the gap between the trend in official and total output is also declining.

The rest of the paper is organized as follows. In Section 1.2, we introduce the empirical model used to measure informal economy with a focus on MIMIC approach, we explain the choice of variables and we describe the different variants of the model. In section 2.4, we present our empirical results and our estimates of informality. Section 1.4 discusses the implications for income convergence and the world distribution of income. Finally, Section 1.5 concludes.

1.2 Measuring informality

This section presents different techniques used to estimate the informal economy but first clarifies what the study means by the informal sector. Any attempt to measure the informal economy first faces the problem of defining what informality means. A revision of the extensive literature leads to several adjectives that have been used to refer to informal activities: among others, we have terms like "hidden", "irregular", "no visible", "shadow", "undeclared", "underground" and "unregulated". For sure, a relative consensus exists on the fact that the informal sector should cover all remunerated activities that are not declared to the state for tax, social security and labor law purposes, but are legal in all other respects (European_Commission, 1998; Williams and Round, 2008). However, this consensus definition involves blurred edges as, in a context where income is unreported, illegal activities such as smuggling, prostitution, drugs trafficking and fraud, which require money transaction, go undetected. For the purposes of our study, and given the variables we used to capture the informal economy, informal output comprises all unregistered economic activities which entail monetary transaction and would potentially generate income taxation were they reported to the tax authorities (Medina and Schneider, 2018; Schneider et al., 2010).

In the literature, the main methodologies to estimate the size of the shadow economy can be attributed to the categories of direct, indirect and model approach. The direct approaches (Isachsen and Strøm, 1985; Mogensen et al., 1995) are micro approaches, based on contacts with or observations of persons and/or firms to collect direct information on undeclared income. Here, two techniques are being used; (1) the auditing of tax returns and, (2) the questionnaire survey. The main advantage of this method lies on the detailed information that can be obtained concerning the structure of the informal economy. However, it is criticized for the results are said to be highly sensitive owing to the wording of the survey questions (Schneider et al., 2015) and it fails to cover the groups of hidden population truly acting in the informal (Pocius, 2015).

The indirect approaches, also called indicator approaches, are mostly of a macroeconomic nature (Del Boca, 1981; Park, 1979; Tanz, 1983). They determine the size of the informal economy by measuring the scent that it leaves out in official statistics. These approaches rely on different sets of indicators: (i) Discrepancy between national expenditure and income statistics; (ii) The discrepancy between the official and actual statistics of labor force; (iii) The transaction approach; (iv) The currency demand (or cash-deposit ratio) approach; (v) The physical input (e.g. electricity) method. These methods of indirect approach are criticized for data disparities and discrepancies (Schneider et al., 2015; Williams, 2010, 2012) and a comparatively narrow scale of applicability (Schneider et al., 2015).

Finally, the model approach also called the MIMIC method (Giles and Lindsay, 2012; Medina and Schneider, 2018; Schneider, 2017; Schneider and Enste, 2000) is mainly based on the statistical theory of latent variables, and explicitly considers the multiple causes as well as the multiple effects of the informal economy. The structural equation in the MIMIC allows to model the causal relationship between observable causes and the latent variables, while the measurement equations link the latent variable with their indicators. Inclusion of a comparatively wide dataset is considered to be one of the key advantages of the methods of multiple-causes approach (Bose et al., 2012; Elgin and Schneider, 2016; Zhou and Oostendorp, 2014). This allows the model to capture several phenomena that are related in one way or another to one or another variable used in the estimation. Among other phenomena, we aim to capture the hidden income, which is, on the one hand, caused by taxation and, on the other hand, correlated with the amount of money in circulation in the economy.

The MIMIC approach can be considered as most reliable because it relies on a variety of measures rather than on a single measure. It allows to identify "the slope coefficients between the size of the informal economy and its cause variables without directly observing the latent variable" (Bose et al. (2012), p.22). This way, the changes in causal variables help to predict the changes in the total size of the informal economy. For these reasons, we select the MIMIC model for our empirical research. Below, the MIMIC model is explained in detail.

1.2.1 The MIMIC model

The MIMIC model assumes that the informal economy cannot be measured directly, but rather through multiple determinants and multiple effects that lead to its existence and its growth over time. Thus, the approach models the informal economy as a latent variable, and estimates it in a set of structural equations. The aim is to examine the relationships between the latent variable "size of informal economy" and observable variables through information on covariance of observable variables. The model consists of two parts. The first part concerns the structural equation (Equation (1.1)) and examines the relationship between latent variable (η) and the causes (X_q). The second part concerns the measurement equation (Equation (1.2)) that links indicators (Y_p) and the informal economy variable (η). The latent variable is then linearly determined, subject to a disturbance ξ

by a set of observable exogenous causes as:

$$\eta = \gamma_1 x_1 + \gamma_2 x_2 + ...\gamma_q x_q + \xi \tag{1.1}$$

Where $X_q = (x_1, x_2, ..., x_q)$ is a set of observable exogenous causes, and $\gamma_q = (\gamma_1, \gamma_2, ..., \gamma_q)$ is a set of structural parameters in the structural equation. Latent variable η linearly determines, subject to a disturbance $\varepsilon_1, \varepsilon_2, ..., \varepsilon_p$, the set of observable indicators $y_1, y_2, \ldots, y_p$ as :

$$\begin{cases} y_1 = \lambda_1 \eta + \varepsilon_1 \\ y_2 = \lambda_2 \eta + \varepsilon_2 \\ y_p = \lambda_p \eta + \varepsilon_p \end{cases} \tag{1.2}$$

where: $Y_p = (y_1, y_2, \ldots, y_p)$ is a set of observable endogenous indicators, $\lambda_P = (\lambda_1, \lambda_2, ...\lambda_p)$ is a set of structural parameters in the measurement model, and $\varepsilon_1, \varepsilon_2, ...\varepsilon_p$ are the measurement errors [2]. More precisely, ε refers to a vector of measurement error variables associated with the indicators, while ξ is a structural error that captures unmodeled variables affecting η and measurement error associated with it. The measurement equations relate each indicator variable to the latent informal economy and a random measurement error term. It is assumed that ξ and all the elements of ε are mutually uncorrelated. [3]. Further, var$(\xi) = \Psi$ and the variance covariance matrix of $\varepsilon = \Theta_\varepsilon = E(\varepsilon\varepsilon')$. The vector of parameters λ is also known as factor-loadings that need to be estimated along with the γ parameters. Substituting equation (1.2) into (1.1), the MIMIC model can be conceived as a p-equation multivariate regression model which is no longer a function of the latent variable and takes the standard reduced form below:

$$Y = \Pi x + z \tag{1.3}$$

where $\Pi = \lambda'\gamma$ is a reduced form coefficient matrix and $z = \lambda\xi + \varepsilon$ is a reduced form vector of a linear transformation of disturbances. The variance-covariance matrix of z can be written as: Var(z)= $\Omega = E(zz') = E[(\lambda\xi + \varepsilon) + (\lambda\xi + \varepsilon)'] = \lambda\lambda'\Psi + \Theta_\varepsilon$. Using the standard normalization, $\Psi = E(\xi\xi') = 1$ we get that $\Omega = \lambda\lambda' + \Theta_\varepsilon$. Based on this,

[2] The model consisting of (1) and (2) cannot determine the scale of all of the parameters, so a normalization condition is required. There are many possibilities, but the standard way adopts the convention of setting the first element of λ to be unity, as $\lambda_1 = 1$

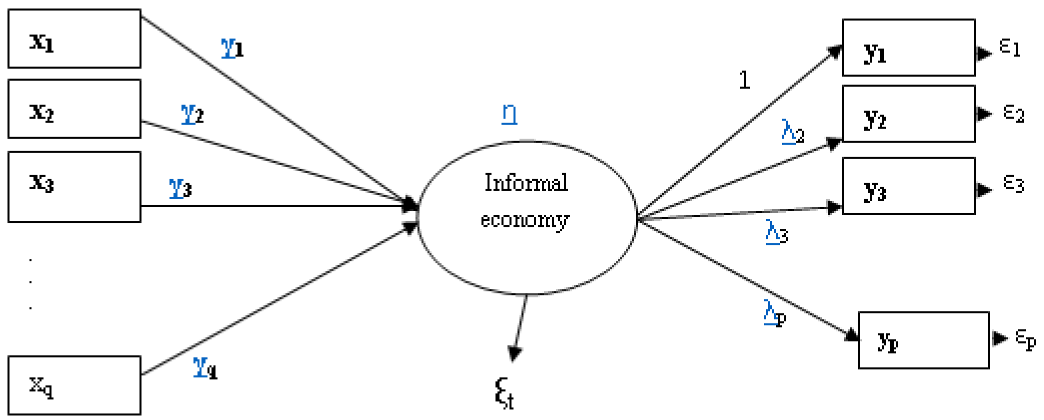
[3] However, this standard restriction could be relaxed. (see Stapleton (1978) and Dell'Anno (2003)). When this assumption is relaxed the changes in the estimates of structural coefficients are insignificant (Dell'Anno and Schneider, 2006)

Joreskog and Goldberger (1975) show that the estimator of η is given by:

$$\eta = (1 - \lambda'\Omega^{-1}\lambda)^{-1}(\gamma' x + \lambda'\Theta_{\varepsilon}^{-1} y) \tag{1.4}$$

This shows that the MIMIC latent factor estimator is a sum of two terms: The "causes" term (function of x) and the "indicators" term which is the factor scores of the measurement model. Identification of the MIMIC model requires that p (the number of y variables) be two or more and q (the number of x variables) be one or more when η is a scalar (Dell'Anno, 2003). The model is fitted by minimizing the discrepancy function where the discrepancy is defined as the difference between the sample and model implied covariance. The closer this difference is to zero, the better is the fit (Schermelleh-Engel et al., 2003). Following this MIMIC procedure, one obtains the latent variable scores $\eta_{t,i}$, called the index of the latent variables, for each year t and each country i. An additional procedure, benchmarking or calibration procedure is needed in order to compute absolute values. Thus, by using information regarding the specific value of informal activity for countries in a given point of time, obtained from some other source, the within-sample predictions for η can be converted into absolute series. The latter represents an estimate of the informal economy as percentage of GDP. The interaction over a given period of time between the causes $X_q(q = 1, 2, 3, ..., q)$, and the size of the informal economy and its indicators $Y_p(p = 1, 2, ...p)$ is shown in the following Figure 1.1.

Figure 1.1: Theoretical specification of a MIMIC model



Even though the standard MIMIC model has been widely used by the literature for many years, it has also been subject of criticism. Doubt is cast on its debatable reliability and comparatively narrow applicability of the Model (i.e., the procedure of index construction is not the only one which can be considered as appropriate, the intertemporal and the cross-country stability of parameters are questioned, etc). The critics of the model disagree on the confirmatory rather than exploratory nature of this approach (i.e., the approach is more likely to determine whether a certain model is valid than to "find" a suitable model (Schneider and Buehn, 2016)). Furthermore, as illegal economic activities (e.g. drug trafficking, cigarette smuggling, etc.) are hard to dissociate from the analysis of the general informal economy, the risk of data duplication may emerge. Looking at the coefficients, Helberger and Knepel (1988) note that the MIMIC model generates unstable coefficients due to the changes in a sample size and alternative specifications. Notwithstanding the different criticisms, the MIMIC model is treated as a valuable tool for identification of the informal economy's causes and indicators. In this connection, the above-explicated critical arguments should serve as an incentive for further research. Indeed, Dell'Anno (2003) and Trebicka (2014) have proven that, among the econometric alternatives to measure the informal economy, the MIMIC approach remains a relatively robust methodology in this field. Although the model is still in progress and supplementary improvement is called for, Thomas (1992) and Giles and Lindsay (2012) have indicated that, a good choice of variables could significantly reduce the limits of this approach. Bellow, we shall theoretically support the variables we used in the econometric framework.

1.2.2 Selection of variables in this paper

The size of informal economy has different roots in various countries. The literature reports a large body of possible causes and indicators of the existence and growth of informal economy. As regards this study, causes and indicators variables consist of annual observations from 1995 to 2013 collected on the 120 countries worldwide. Below, we present the causes and indicators variables used and describe their sources and unit of measurement.

Causes of informality

The causes of informal economy are grouped into three mains classes; the burden of taxation, the quality of institutions that includes the regulatory burden and the state of the economy

Tax Burden. - The tax burden is the key cause that brings into being and rise of the informal economy (Schneider et al., 2015). A rising burden of taxation, which widens the gap between the before and after-tax earnings, provides a strong incentive to work underground to evade taxation. On the one hand, some authors, like Allingham and Sandmo (1972), have provided theoretical foundations showing that informal economy increases with tax burden. On the other hand, empirical analysis have reported also a nexus between an increasing tax burden and the size of the informal economy. For example, in their study, Johnson and Zoido-Lobatón (1998) point out that, the tax burden is one of the main causes of the informal economy. Schneider and Enste (2000) argue that, the taxes and the state regulatory activities are the main determinants leading the dynamics of the informal economy. The tax burden is often measured, using the total share of direct and indirect taxes and social contribution as a percentage of GDP or using the government spending. In this study, we use the tax revenue per worker, measured by the means of the total tax (direct and indirect) per worker. This variable is obtained using the expression below:

$$\frac{\text{Tax revenue}}{\text{Working age population}} = \frac{\text{Tax revenue}}{\text{GDP}} . \frac{\text{GDP}}{\text{Population}} . \left(\frac{\text{Working age population}}{\text{Population}}\right)^{-1} \tag{1.5}$$

The tax revenue includes the revenues collected from taxes on income and profits, social security contributions, taxes levied on goods and services, payroll taxes, taxes on the ownership and transfer of property, and other taxes. It is taken from the IMF-GFS database as a percentage of GDP in order to get the share of a country's output that is collected by the government through taxes.

Quality of institutions. - The quality of institutions appears to be among the key determinants of the informal economy (Schneider, 2007). Elgin and Oztunali (2014) provide a theoretical support on the link between informal economy and institutions. They show that, low quality of institutions proxied by the level of law enforcement is positively associated with a larger informal sector size. Singh et al. (2012) insist on the central role of institutions in preventing the growth of informal economy and Dell'Anno (2010) shows that institutional quality is one of the key indicators of informality in Latin American economies. Several indexes that aim to measure the quality of institution have been used in the literature. In our study, we focus on **the rule of law through secured property rights.** This variable is an index ranging from 10 to 100 and based on the data from the Heritage Foundation. Its component includes an assessment of the ability of individuals to accumulate private property, secured by clear laws that are fully enforced by the state. Indeed, Singh et al. (2012) note that, among others factors, a strong culture of the rule of law is a critical priority to deal with informal economy given that it is the basis for good institutions. Thus, economic progress matching with improvement in institutional quality would reduce incentives of firms and households to go informal.

State of the official economy. - The state of the official economy also plays a crucial role in people's decision to informally work or not (Feld and Schneider, 2010). In this paper, we use four variables to capture the state of the economy. First, **the poverty rate**, taken from the World Bank Indicators; it represents the proportion of population living below income poverty line, PPP \$1.90 a day (%). In economies facing poverty, more people working in the formal sector try to supplement their low income from the official economy through additional informal activities. Using a cross-section data for Chile, Amuedo-Dorantes (2004) concludes that household poverty increases the likelihood of employment in the informal sector. Second, **the human capital index**, an index of human capital per person estimated by the Penn Word Table. It is based on years of schooling and returns to education. Human capital appears to be one of the most striking differences between formal and informal firms. Many people, manly low skilled ones, excluded from the formal sector, develop subsistence activities in the informal sector. These people are typically uneducated and they run small businesses producing low-quality products supplied to low-income buyers. La Porta and Shleifer (2014a) argue that: "a shortage of educated entrepreneurs might be the most important constraint on transition to formality." Third, **inflation**. It is represented by the implicit price deflator and taken from the World Bank Indicator. It is a measure of the level of prices of all new, domestically produced, final goods and services in an economy. Following Giles (1999) and Vuletin (2008), accounting for the inflation rate allows for the upward "creep"of tax brackets, and the associated incentive for taxpayers to engage in informal activities. A more pervasive effect of inflation is that, as it tends to be uneven across sectors, it alters the income distribution, which in turn may induce disrespect for tax law. Therefore, the higher the inflation, the larger the expected size of the informal economy. Lastly, **the openness:** that is an index of trade freedom from Heritage Foundation. It is a composite measure of the absence of tariff and non-tariff barriers that affect imports and exports of goods and services. Trade influences market size and a country's openness to the world. In Schneider (2017)'s formulation, as economies grow, it is likely to be more difficult to move economic activity from the formal to the informal sector. Furthermore, as international trade increases, it would be harder to hide trade from the authorities. In using a dynamic industry model with firm heterogeneity, Alemán-Castilla (2006) shows that, openness through import tariff reduction could reduce the incidence of informality by making it more profitable for some firms to enter the formal sector. This could force the less productive informal firms to exit the industry, and induce the most productive formal firms to engage in trade. Empirically, he shows that reductions in the Mexican import tariffs are positively related to reductions in the likelihood of informality in the tradable industries.

Indicators of informality

The literature reports a set of indices where a change in the size of the informal economy may be reflected. Among others, three mains variables reported by Schneider and Enste (2000) are considered in this study.

Monetary indicators (Currency outside of banks) - People engaged in informal activities avoid leaving traces of their transactions. For this reason, they primarily use cash (Schneider, 2017). Hence, most informal economy activities are reflected in an additional use of cash (or currency). In line with empirical studies, we integrate this "currency" variable in our analysis. We use the ratio of monetary base (typically includes only the most liquid instruments, such as coins and notes in circulation) to broad money (includes narrow money along with other assets that can be easily converted into cash to buy goods and services) taken from the IMF/IFS database.

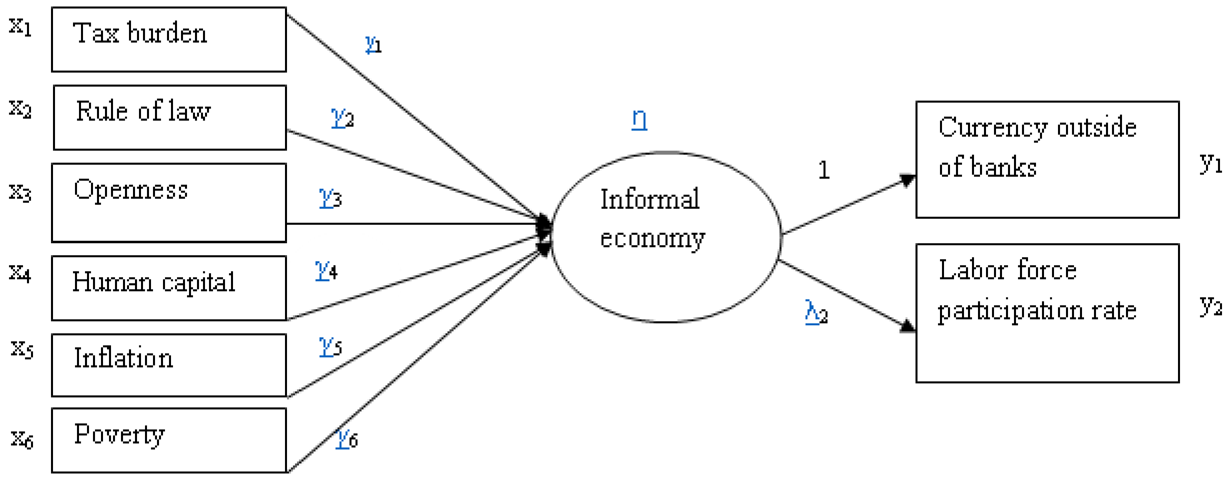
Labor market indicators Informal activities are also reflected in labor market indicators as a decline in official labor force participation that could signal some giving upon searching for work in the formal sector. These workers would no longer show up as part of the labor force in national surveys (Schneider, 2017). In this study, we use ILO estimates of the labor force participation that are harmonized to account for differences in national data and scope of coverage. It refers to the sum of all persons of working age who are employed or in search of a job.

The gross domestic product (official output) - A priori, it is not possible to determine what is the effect of the informal economy on the formal economy. Indeed, the GDP can be both a determinant and an outcome of informal activities. This is because a downturn in the official activities may lead to a loss of jobs and thus, drive more individuals into the informal activities. Alternatively, a contraction in the official activities may reduce the demand for informal products and thus offset the first effect. The informal economy represents a "life jacket" for firms and individuals in financial troubles and for that reason, it increases when the GDP decreases, or rather more growth means more opportunity to evade. Leaving out of this puzzle is a very hard task (Dell'Anno, 2003). Empirically, the link between the unofficial and official economy is proven to be ambiguous (Schneider and Enste, 2000). Different perspectives on informality imply a specific macroeconomic relationship between GDP and informal economy. For example, the tax-avoidance and poverty views for which informal economy is a parasitic organizational form that hinders economic growth advocates for a negative relationship (Docquier et al., 2017; Levy, 2008). The romantic view sees informal firms as an untapped tank of entrepreneurial energy, which, if unleashed, would fuel growth and development (De Soto, 1989, 2000). Finally, the dual perspective, from which formal and informal firms are fundamentally different and independent from each other, pleads for a zero effect (Lewis, 1954; Rauch, 1991).

In the literature, the GDP variable is used in different ways in the MIMIC model. Buehn and Schneider (2008) and Dell'Anno (2003) use the GDP as an indicator. Medina and Schneider (2018) use the per capita GDP as a cause while keeping the growth rate of GDP as an indicator. Giles and Lindsay (2012), Dell'Anno and Schneider (2003) and Bajada and Schneider (2005) use the GDP as a cause and as an indicator jointly. These different specifications have raised strong criticisms against the model. Breusch (2005) claims that, in some of these studies, the dynamics of the informal economy estimates might be strongly attributed to the change in the GDP. Furthermore, he argues that the causes and the indicator variables should be related to each other only through the size of informal economy to prevent biased estimators and ill-behaved test of overall model. Thus, he castigates the use of GDP as a cause and an indicator. However, Dell'Anno and Schneider (2006) have proven that, the connection between causes and indicators variable does not affect the general appropriateness of MIMIC method. In this paper, we will not take position concerning the use of GDP. However, we develop different specifications of the MIMIC model to account for all views presented above. To eliminate the non-stationary of the time series, we transformed the variables in their first differences. In so doing, we develop a first specification, the benchmark specification, which excludes completely the GDP variable neither as a cause nor as an indicator. Our second specification includes the per-capita GDP (taken in PPP from the World Bank Indicator database) as a cause while the third specification includes the GDP (taken in PPP from the Penn World Table 9.0 portal) as an indicator. The last specification use the nighttime light intensity (by Proville et al. (2017)) instead of the GDP as a proxy of global activities.

As far as informality is concerned, the economic literature considers the cause variables enumerated above as exogenous (See, Dell'Anno (2003); Dell'Anno and Schneider (2003); Giles and Lindsay (2012); Medina and Schneider (2018); Schneider (2007); Schneider and Enste (2000)). So it is for this study. However, we are aware of the fact that, the exogeneity of some of our causes variables is debatable because, for example, it can argue that the decision to invest in human capital may depend on the size of the informal sector, poverty or other factors. In any cases, any empirical analysis that uses the estimates of informal economy in its framework is necessarily subject to the same limits. Among these, even putting aside measurement errors and the effect of omitted variables, the most relevant is probably this issue of endogeneity of variables (Dell'Anno, 2010). For this, we carried out robustness checks by modifying the structure of the model. We have tested variants of the reference model in which we hypothesize that variables initially taken as causes (taxation, human capital, poverty and inflation) fall into the category of indicators of informality. We thus check how this affects the size and the trend of the informal economy obtained from our benchmark specification. What is pointed out in this study on informality and the world income distribution contributes to the ongoing debate, confirming previous empirical results and offering new insights. Using the MIMIC procedure, we obtain estimates of informal economy for the different specifications and we study theirs trends and patterns. The following figure 1.2 gives the hypothesized MIMIC path for our benchmark specification to estimate the informal economy.

Figure 1.2: Benchmark specification of our MIMIC model



1.3 Results

In this section, we present our empirical results on the size of informal economy as a percentage of the GDP under different specifications. Focusing on the benchmark specification, we construct a global output indicator that accounts for both the official output (GDP) and the informal output and we analyze its implication on cross-country convergence and inequality.

1.3.1 Estimation results and robustness

We first conduct our empirical procedure on the benchmark specification. Thereafter, we apply the procedure on three alternatives specifications that are a modified version of the benchmark model as previously described. Table 1.1 presents the results of our benchmark specification and three alternatives. From the Table 1.1 and from the benchmark specification, we notice that, except the poverty variable that is not significant, all others determine significantly the informal economy with the expected sign. The tax burden and inflation determine with positive sign while rule of law, openness and human capital have negative sign. In the first alternative, the per-capita GDP appears to be a significant cause with negative sign. In the second alternative, the GDP is a significant indicator of informal economy with negative sign. In the third alternative, the nighttime light is a significant indicator with negative sign.

Table 1.1: Estimation results on the full sample

	Benchmark	Alternative_1	Alternative_2	Alternative_3
Cause variables				
Constant	-1.498***	-1.498***	-1.220***	-1.178***
	(0.011)	(0.011)	(0.011)	(0.000)
Tax burden	0.073*	0.099**	0.095	0.106*
	(0.042)	(0.044)	(0.059)	(0.058)
Rule of law	-0.152***	-0.153***	-0.159***	0.043
	(0.025)	(0.025)	(0.033)	(0.055)
Openness	-0.153***	-0.156***	-0.207***	-0.126***
	(0.029)	(0.029)	(0.039)	(0.038)
Human capital	-0.028**	-0.028**	-0.027*	0.967
	(0.012)	(0.012)	(0.015)	(1.321)
Inflation	0.169***	0.168***	0.202**	0.150*
	(0.063)	(0.063)	(0.083)	(0.086)
Poverty	0.011	-	-	-
	(0.008)	-	-	-
Per-capita GDP	-	-0.253*	-0.266	-0.010
	-	- (0.152)	(0.200)	(0.011)
Indicator variables				
Currency outside of banks	1	1	1	1
Labor force participation	-0.024***	-0.024***	-	-
	(0.007)	(0.007)	-	-
GDP	-	-	-0.049***	-
	-	-	(0.010)	-
Nighttime light intensity	-	-	-	-0.0234**
	-	-	-	(0.010)
Likelihood ratio test	0.000	0.000	0.000	0.002
Countries	120	120	120	120

Notes: Author's computation. Std errors are in parenthesis; $*p < 0.10$ $**p < 0.05$ $***p < 0.01$..

For robustness check, we run our benchmark specification on different subsamples, as some of the potential causes of informality can be different according to countries' level of development (Docquier et al., 2017). For this purpose, we use the GDP in order to group countries in quartiles[4]. In doing so, we first divide the sample by the median and run the model for the two subsamples. Thereafter, we run the model for countries above the first quartile and countries below the third quartile. The results are presented in the Table 1.2. The robustness check case have two objectives. First, to see whether the significance and sign of coefficients will be modified compare to the benchmark specification applied on the full sample. Second, to see whether the estimates of informality produced by different subsamples are correlated with the estimates of informality obtained with the equivalent subset of the full sample of the benchmark specification. The results in Table 1.2 show that, when we consider subsample, the observed variables keep their predictive sign from the full sample although, some of them are no longer significant. Additional robustness checks are made by changing the structure of the model. We consider as indicators some variables initially taken as causes. First, we keep the number of indicators at two and we successively replace the variable "labor force participation" by each of these variables

[4]Subsample based on our income group partition are limited in terms of size while, a quartile grouping helps to limit the instability of coefficients with respect to changes in the sample size and alternative model specifications as instability disappears asymptotically with the sample size (Dell'Anno, 2003)

(see Table 1.7). Second, we add these variables to the set of possible informal indicators (we therefore have three indicators, see Table 1.8). It emerges from our robustness checks that (i) with the exception of inflation, all the variables passing from the status of cause to the status of indicator completely lose their significance and (ii) these changes made do not modify the sign of the variables which are maintained as causes, although their significance is sometimes affected. In the next section, we compute the size of informal economy with each subsamples specification and check for the correlation coefficient with the equivalent subset of the full sample.

Table 1.2: Estimation results on subsample

	Full sample: Benchmark	Above Median	Below Median	Above first quartile	Below third quartile
Cause variables					
Constant	-1.498***	-1.382***	-1.690***	-1.456***	-1.594***
	(0.011)	(0.016)	(0.016)	(0.013)	(0.013)
Tax burden	0.073*	-0.036	0.152***	0.018	0.121***
	(0.042)	(0.072)	(0.045)	(0.055)	(0.042)
Rule of law	-0.152***	-0.334***	-0.071***	-0.172***	-0.134***
	(0.025)	(0.054)	(0.023)	(0.033)	(0.023)
Openness	-0.153***	-0.027	-0.167***	-0.174***	-0.172***
	(0.029)	(0.070)	(0.027)	(0.043)	(0.026)
Human capital	-0.028**	0.024	-0.080***	0.012	-0.054***
	(0.012)	(0.018)	(0.015)	(0.014)	(0.013)
Inflation	0.169***	0.181**	0.182*	0.146**	0.158***
	(0.063)	(0.085)	(0.095)	(0.071)	(0.059)
Poverty	0.011	0.005	0.030**	0.008	0.016*
	(0.008)	(0.010)	(0.014)	(0.009)	(0.009)
Indicator variables					
Currency outside of banks	1	1	1	1	1
Labor force participation	-0.024***	-0.070**	-0.078***	-0.068***	-0.031***
	(0.007)	(0.012)	(0.014)	(0.009)	(0.010)
Likelihood ratio test	0.000	0.000	0.000	0.000	0.000
Countries	120	60	60	90	90

Notes: Author's computation. Std errors are in parenthesis; $^{*}p < 0.10$ $^{**}p < 0.05$ $^{***}p < 0.01$.

1.3.2 The size of informality

The MIMIC procedure gives coefficients that determine relatively the estimated size of the informal economy, describing their pattern over time. To obtain the size and trend of the informal economy, one needs to convert the MIMIC index into "real word" values measured in percentage of GDP. This final step requires an additional procedure so called benchmarking or calibration procedure. For this purpose, different specifications have been developed. Here, we rely on the procedure promoted by Dell'Anno and Schneider (2006), and Dell'Anno (2008). In the first step of this procedure, using the estimates of the γ_i vector and setting the error term ξ to its mean value of zero, the predicted index values for the informal economy (η_t) can be estimated thanks to equation (1.1). In the second step, this index is converted into absolute values of the informal economy by using a base value in a particular base year. For this step of calibration, the base values used are from the year 1996 and taken from Elgin and Oztunali (2012), who present estimates of the shadow economies in 161 countries worldwide and provided model-based estimates of the informal economy. Thus, the size of the informal economy $\hat{\eta}_t$ at time t is obtained using the expression bellow:

$$\hat{\eta}_t = \frac{\tilde{\eta}_t}{\tilde{\eta}_{1996}} \eta^*_{1996} \tag{1.6}$$

where $\tilde{\eta}_t$ is the value of the MIMIC index at time t according to the structural model (1.1), $\tilde{\eta}_{1996}$ is the value of this MIMIC index in the base year 1996, and η^*_{1996} is the exogenous estimate (base value) of the informal economy in 1996 taken from Elgin and Oztunali (2012). Applying this benchmarking program, the procedure allows us to compute eight series of dataset that report the estimates of informal sector as percentage of GDP. The series concern the four specifications applied on the full sample and the benchmark specification applied on four subsamples. For each of the four series applied on full sample, the dataset has 2160 observations for 120 countries in a balanced panel framework for the year 1996 to 2013.[5] We then notice three things as reported in Table 1.3: First, our estimates of informality from the benchmark specification have a correlation of more than 98% with the three alternative specifications based on the full sample that use the GDP or the nighttime light intensity in the model. Second, our estimates have a correlation of 93% with empirical estimates from Medina and Schneider (2018). Although their procedure also relies on the GDP, their sample have more countries and a long period of time then ours. Third, at the same way, the equivalent subset of the benchmark specification is highly correlated with the estimates obtained at subsamples level (more than 95%).

Table 1.3: Correlation coefficient between estimates of different models

Correlation coefficient : Full Sample level					
Benchmark-Model	Alternative _1	Alternative _2	Alternative _3	Schneider (2017)	Medina and Schneider(2018)
	0.9946	0.9961	0.9856	0.9376	0.9320
Correlation coefficient : Subsample level					
	Above: Median	Bellow median	Above First quartile	Bellow Third quartile	
Benchmark-Model	0.9962	0.9563	0.9981	0.9990	

Notes: Author's computation

Thereby, the MIMIC model without the GDP does not really lose information on the estimate of the informal economy. The complete dataset from the benchmark specification is reported in a country-by-country and year by year basis in the appendix (see Table 1.12. Nevertheless, in this section, we report illustrative figures and various descriptive statistics. First, we estimate countries' distribution non-parametrically. For this, we rely on the income-group classification of the World-Bank (2014) and divide our sample

[5]Here we lost the year 1995 with the first difference applied on variables.

in four groups. High income countries (HIC), upper middle income countries (UMIC), lower middle income countries (LMIC) and low income countries (LIC). This is done over four period of time: P_1 (1996 - 2000), P_2 (2001 - 2005), P_3 (2006 - 2010) and P_4 (2011 - 2013). Figure 1.3 displays the results where we approximate each group of countries' informal share distribution using a non parametric kernel density function. This procedure does not impose specific functional forms on individual group distributions. One key parameter that needs to be specified is the bandwidth of the kernel. We follow the literature and use the bandwidth $h = 0.9 * sd * n^{-1/5}$ where sd is the standard deviation of log-informal share and n is the number of observations. This gives a bandwidth h of 0.129. Obviously, each income group has a different sd so, if we use this formula for h, we would assume a different h for each of the group and each period. Instead, we follow the approach of Sala-i Martin (2006) and we use the same bandwidth for all income group and periods. The author claims that, with a constant bandwidth it is very easy to visualize whether the variance of the distribution has increased or decreased over time. To get a sense of the level of informality and for each group, the figure also plots a vertical line that corresponds to the average of all periods and countries.

Figure 1.3 displays the distribution of informal share for LIC. In period one, we observe that more than a half of the distribution lies to the right part of the world average line. As time goes by, the shift from the right to the left part of the line is observed. In the fourth period; about a half of the distribution is now at the right part of the world average line. This shift of the distribution depicts a declining trend of informality as percentage of GDP over time in LIC. Figure 1.3 shows a huge shift in informality share in LMIC. This group of countries starts with more than a half of the distribution on the right of the world average part in the first period. From that period, the shift of the distribution continuous and ends up to be a third of the distribution in the last period. In this group, the informal share is steadily declining. Figure 1.3 displays the distribution of informal share for UMIC. Here, more than the third of the distribution is below the world average, and the movement of the distribution on the left of the world average line is clear. Figure 1.3 shows the distribution of the informal share in HIC. Here, it is seen that about all the size is under the world average level, and the distribution of informal share in this group seems to have moved of course, but not significantly over the studied period. Summarizing, the distribution plot highlights a declining trend of the share of informal economy over time.

Figure 1.3: Evolution of informality by income group (log share in GDP)

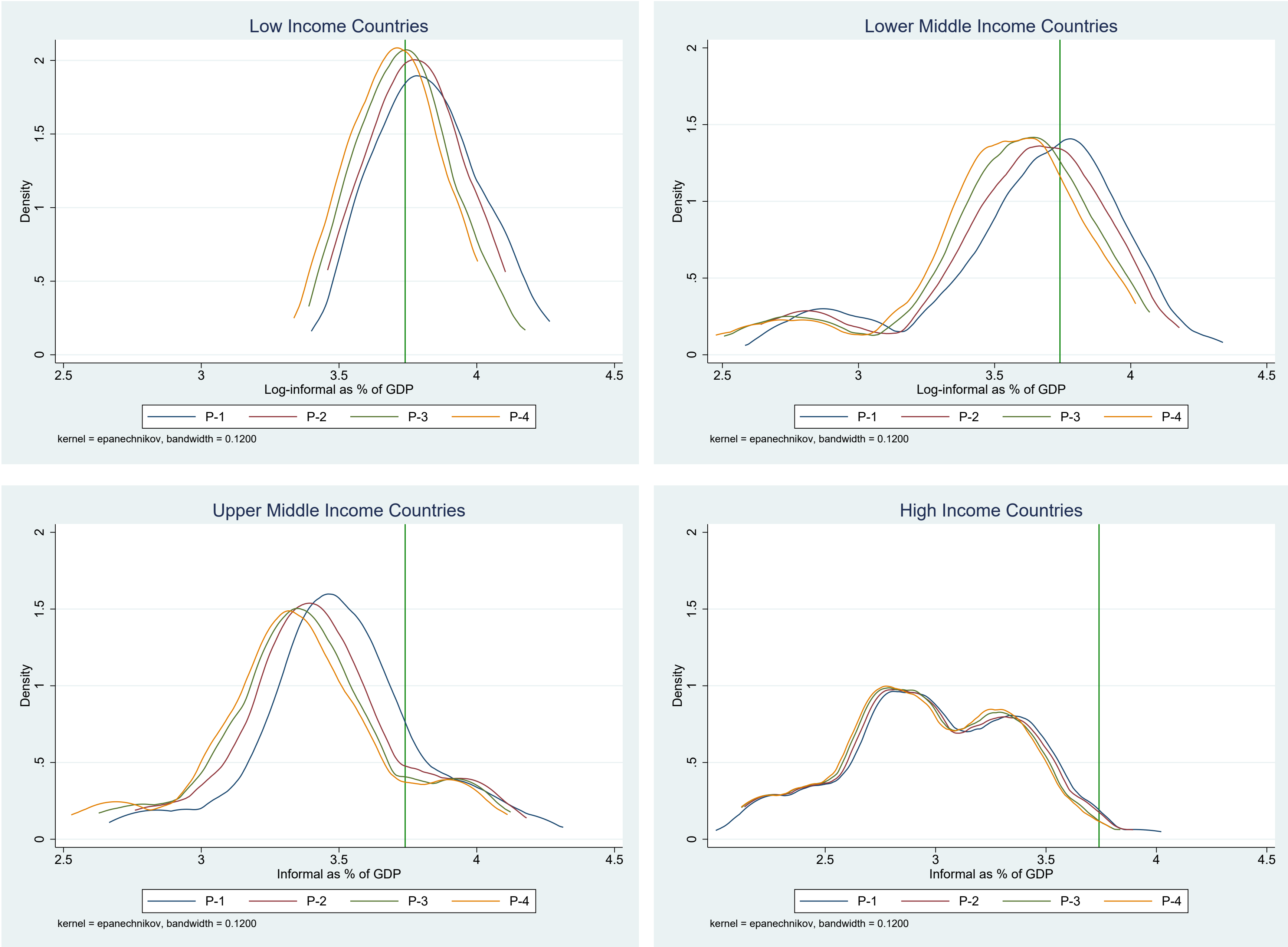
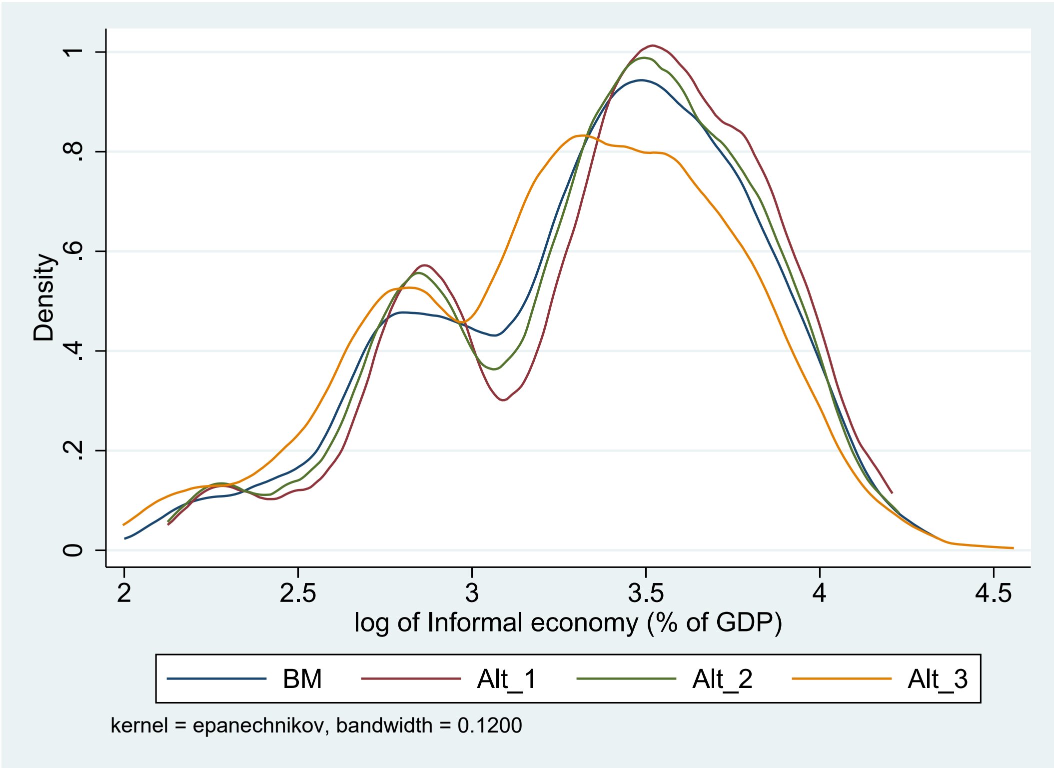


Figure 1.4: Density of estimated informal share in % of GDP under three alternative specifications



Second, we plot the density distribution of the different specifications of the model. For this, we used estimations reported in Table 1.1 and we apply the calibration procedure (using equation 1.1 and 1.6) for each of them. We then plot the distribution of the different series of the size of informal economy as percentage of GDP obtained by our benchmark model and the three alternative specification. In Figure 1.4, we observe that these distributions keep the same pattern as the benchmark specification. From this result with the correlation reported in Table 1.3, One concludes that, using the GDP or not to estimate the size of the informal economy does not affect its distribution.

Third, we use the result on the four subsamples that concern countries above the median, below the median, above the first quartile and below the fourth quartile. We compare the densities of estimates of informality shares in GDP obtained with sub-sample estimations with those of the full-sample benchmark specification. Figure 1.5 gives the result of the distribution plot. It shows that, for each quartiles, the distribution of the estimates obtained from the equivalent subset of the full sample behaves in the same manner as the one obtained at the subsamples level. This reinforces our confidence in the result obtained with our benchmark specification applied on the full sample. We went futher and use the Kolmogorov Smirnov equality-of-distributions test. In doing so, we test the equality-of-distributions between our Benchmark estimates and estimates of alternative_3 that uses GDP as cause and indicator and the on the one side and empirical estimates of Schneider (2017) that use GDP only as indicator. Result are presented in table 1.10 in appendices. Overall the result of the test shows that there are any differences in the distribution of these estimates.

Fourth, focusing on the estimates of the benchmark specification, we plot the size of informal economy as percentage of GDP vs. GDP per-capita. Here we plot average values for every country from 1996 to 2013. Figure 1.6 shows a strong negative relationship between per-capita GDP and informal output as a percentage of GDP, implying that, richer countries tend to have a smaller informal economy.

Fifth, to verify the validity of our estimates, we collected survey data from the International Labor Organization (ILO). The data from the ILO refer to the share of informal employment in total employment. The data are collected annually and cover the period 2010-2020. For our analysis, we consider 15 countries that are common with the ILO database, over the period 2011-2013 since our sample stops in 2013, and the ILO data for the year 2010 include many missing values. We then have 45 observations to compare. We compared our MIMIC estimates (expressed as % of formal GDP) and survey estimates (expressed as % of total employment) by calculating the correlation coefficient. The idea is that if the informal expressed as a percentage of employment is higher, then the informal expressed as a percentage of GDP will also be higher. We also calculated the slope of the regression of the MIMIC estimates on the survey data. Our regression equation is written as follows:[6]

$$\frac{y_I L_I}{y_F L_T} = \alpha + \beta(\frac{\hat{L}_I}{L_T}) \tag{1.7}$$

where y is the output per worker and L the employment (therefore yL is the total employment). We expect $\beta \cong \frac{y_I}{L_I} < 1$ since the informal sector is less productive than the formal sector. We represented this relationship graphically to visualize how the scatter plot behaves around the line. We found that the correlation coefficient between our estimates and the survey estimates is 0.717. The results of the regression of the MIMIC estimates on those of the surveys give a slope of 0.4 which is significant at 5% with an R^2 of 51.4%. Since the two estimates are percentage ratios, this slope suggests that the informal sector is 2.5 times less productive than the formal sector, which seems intuitive to us. As the sample of countries in the ILO database includes Lower Middle Income Countries

[6]The variables y_F and y_I are the official and informal per capita GDP respectively; L_I and L_T are informal and total labor respectively.

(LMIC) and Upper Middle Income Countries (UMIC), we also checked for the country group using of a UMIC dummy. We then get a slope of 0.33 which is still significant and the coefficient of the UMIC dummy is -0.04. This suggests that when the country is more developed, the formal sector is even more productive than the informal. This high level of correlation between these two types of estimate as well as this difference in productivity between formal and informal are such as to validate the level of our estimates expressed in % of GDP. Finally, the Figure 1.12 shows that the points on the graph reflects a strong relationship between our MIMIC estimates and the survey estimates.

Sixth and finally, with the estimates of the benchmark specification, we report descriptive statistics of informal economy as percentage of GDP by income group. To observe the difference and variation of the size of informal economy in different group of countries over time, we consider the previous repartition in four income groups. We also report descriptive statistics for the whole dataset as well. Table 1.4 shows the descriptive statistics on the estimates of informality by income group.

Figure 1.5: Evolution of informality by income group (log share in GDP)

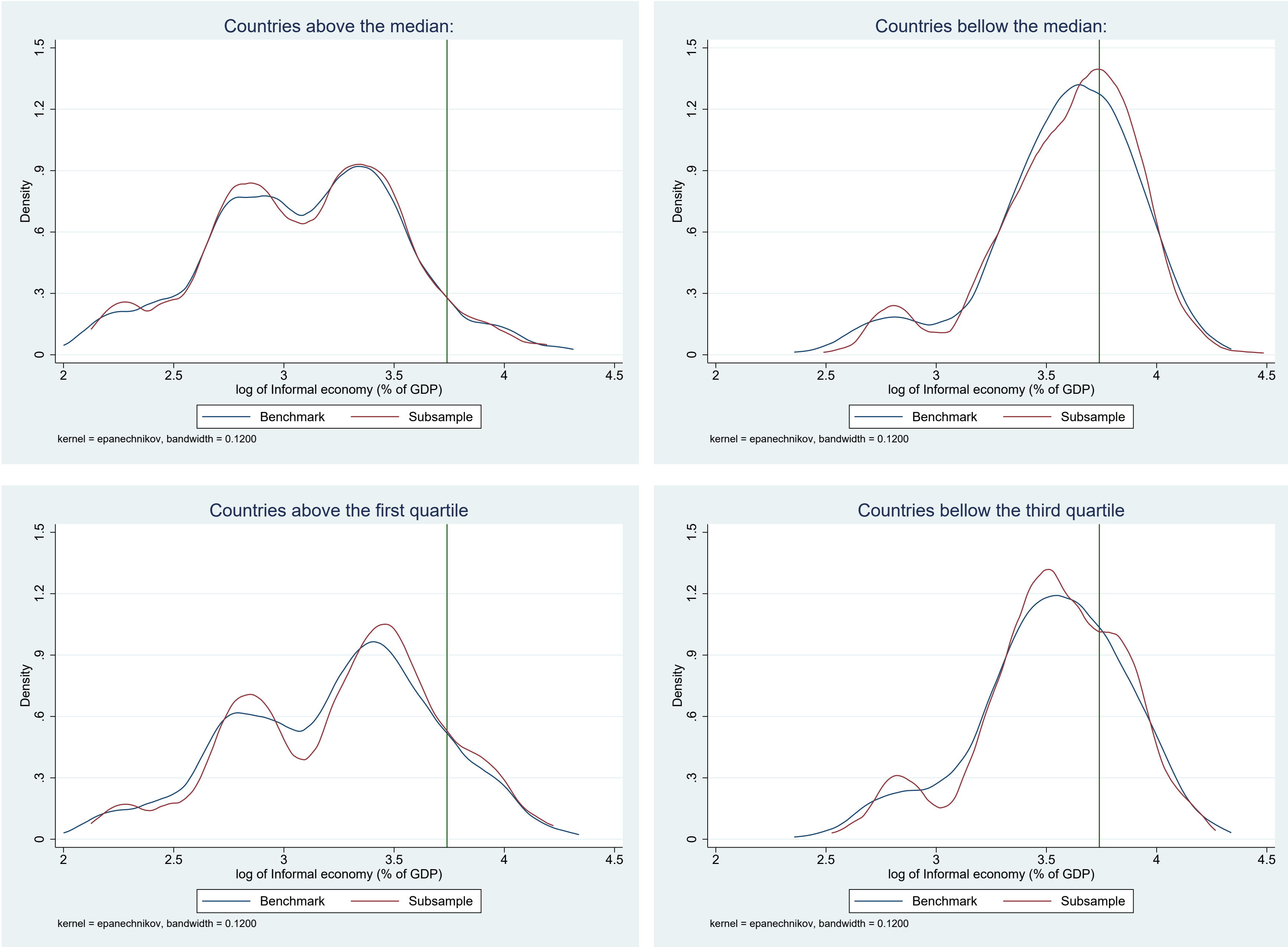


Figure 1.6: Informal economy vs GDP per-capita

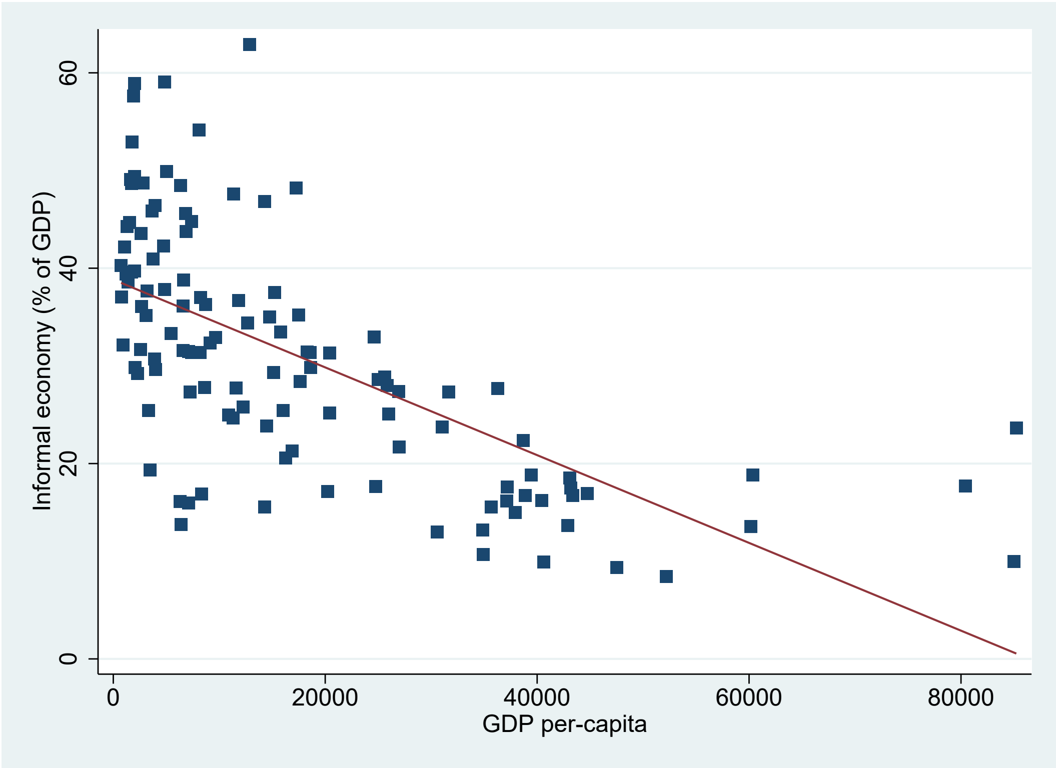


Table 1.4: Informal economy size (%of GDP)

Regions	Mean	St.dev	Min	Max
High Income countries	21.20	8.27	8.33	49.50
Upper middle income countries	32.38	10.61	12.53	66.27
Lower Middle income countries	36.95	11.21	11.89	67.94
Low income countries	43.78	7.16	28.10	65.12
Global	30.92	12.55	8.33	67.94

Notes: Author's computation

Descriptive statistics highlight a couple of essential points that confirm the facts already observed with the distribution plots. First, it is seen that, in high-income countries, informal output tends to be relatively of small size. Secondly, given the magnitude of the standard deviation, the size of the informal economy varies significantly both across and within income groups. In Table 1.5, we report the evolution of the size of informal economy in different groups over time for approximately 4-year intervals. For almost all income groups, we observe a slow but steady declining trend over time. Furthermore, the declining rate seems to depend on the initial level. The larger is the initial size of the informal economy, the higher is the decline rate. This reflects a kind of convergence in the size of informal economy across-countries. Indeed, over the eighteen years, informal activities in high-income countries have declined relatively for 5.4 percentage points. Meanwhile in the low-income countries, the decrease in the size of informal economy is for about 11.48 percentage points. This declining trend over time has been largely reported in theoretical (Elgin and Oztunali, 2012; Gutiérrez_Romero, 2010), and empirical studies (Medina and Schneider, 2018; Schneider, 2017).

Table 1.5: Evolution of informal economy size (% of GDP)

	1996-2000	2001-2005	2006-2010	2011-2013
High Income countries	21,77	21,37	20,82	20,58
Upper middle income countries	34,54	32,67	31,21	30,23
Lower Middle income countries	39,33	37,59	35,55	34,24
Low income countries	46,22	44,25	42,58	40,91
Global	32,60	31,29	29,98	29,10

Notes: Author's computation

1.3.3 Comparison between official and total outputs

As in Dell'Anno (2008) and Méon et al. (2011), we transform the estimates of informal economy obtained with our benchmark specification into unofficial (unrecorded) GDP measured on purchasing power parity (PPP). Thereafter, we compute a Total Output (TO) indicator that accounts for both official and unofficial GDP. The TO indicator is obtained by summing up the official output taken in purchasing power parity and the unofficial one via the equation 1.8. In this expression, $\hat{\eta}_t$ is the estimates of informal economy expressed in percentage of GDP, GDP(PPP) is the GDP taken in purchasing power parity and TO is the Total output indicator.

$$\mathrm{TO} = \mathrm{GDP(PPP)}\left(1 + \frac{\hat{\eta}_t}{100}\right) \tag{1.8}$$

Having obtained the TO indicator, we now compare the official output and the total output. In Figure 1.7, we graph the total output against the official output, to highlight the change that the total output causes to the official output. On the graph, a bubble whose size is proportional to the level of the population represents each country. If there had been no changes, from the official to the total output, the values would be on the 45-degree line, where the values of official are equal to the values of total output. Overlying the function y-axes equals x-axes, we see the discrepancy between the two outputs. All the values are above the 45-degree line, implying that, accounting for the informal output modifies upward the wealth patterns of countries. This rising change is also seen on Figure 1.8, which reports the distribution of the two output indicators. It is observed that, the distribution of the total output is more on the right with respect to the distribution of the official output. This implies more output, wealth, for countries when we take into account the informal output.

Figure 1.7: Recorded versus total outputs: Comparison in level

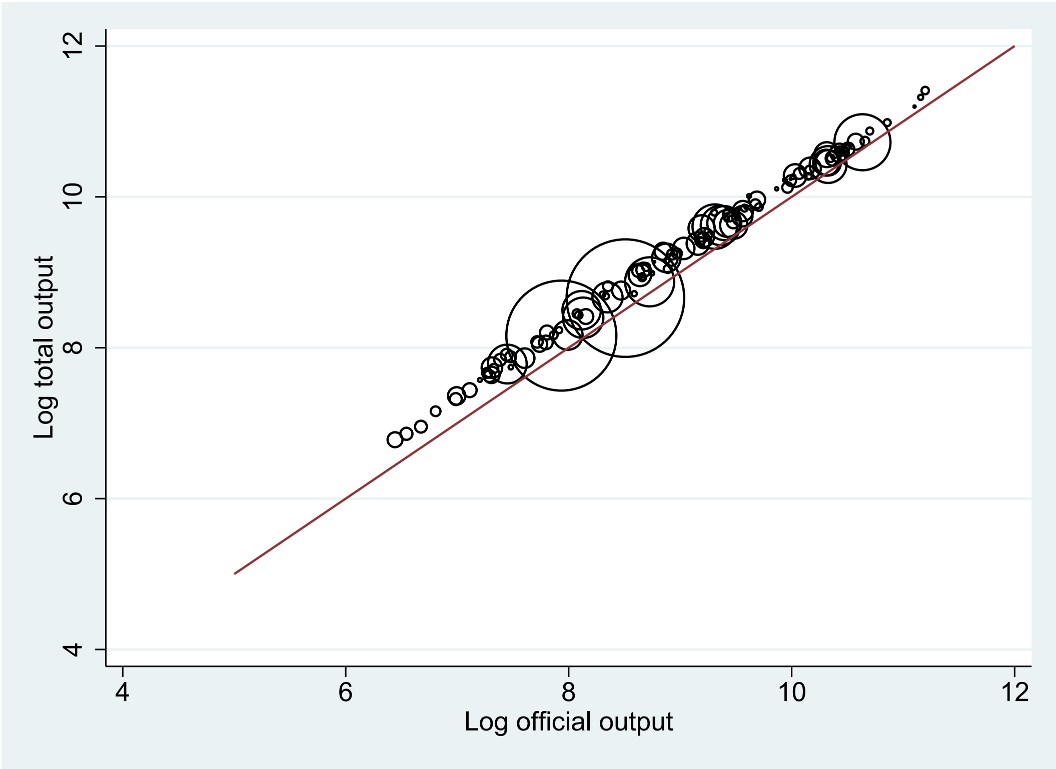
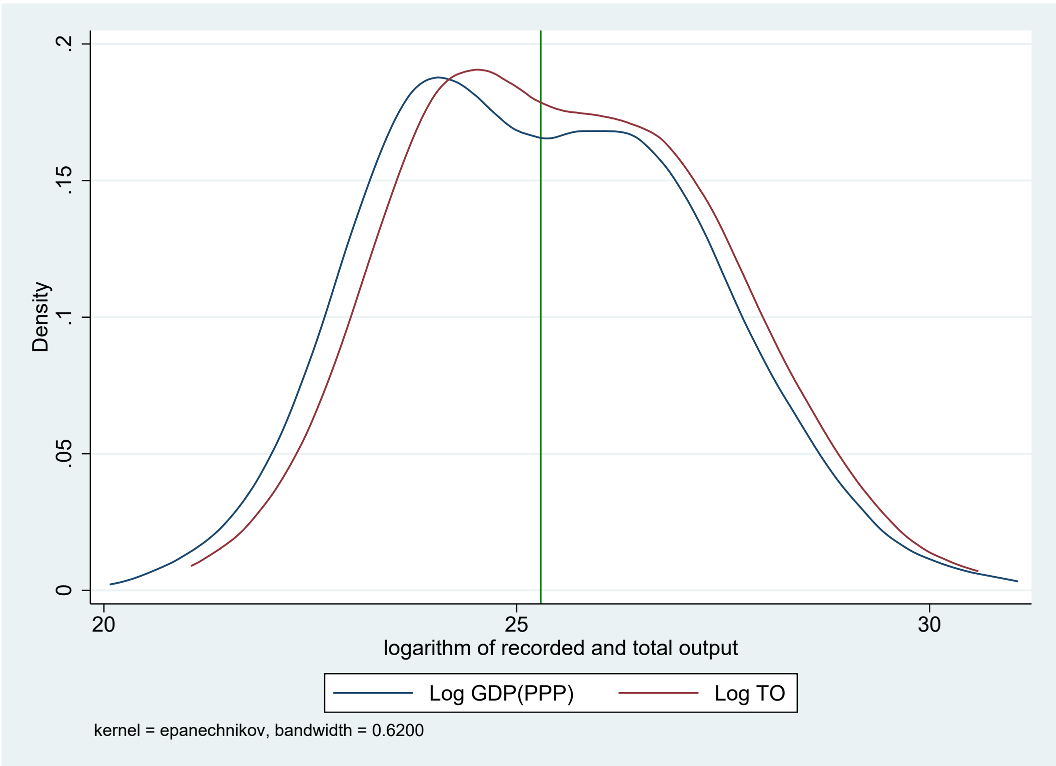


Figure 1.8: Recorded versus total outputs: Comparison of densities



1.4 Implications for convergence and inequality

In this section, we now analyze the implications of our results for the analysis of cross-country convergence and global inequality. Empirical analysis on convergence were popularized by Barro (1989); Barro and Sala-i Martin (2003); Mankiw et al. (1992); Sala-i Martin (2006). Though, they treat data from different sources and sometimes reconstructed for missing years, they all focus on official output to analyze the convergence hypothesis. For the same analysis among regions, the study by Gennaioli et al. (2014) look at satellite data from night-time lights as a proxy for per capita income. In this section, we focus to the official and total output indicators (in per-capita term) to revisit the convergence theory.

1.4.1 Sigma (σ) convergence

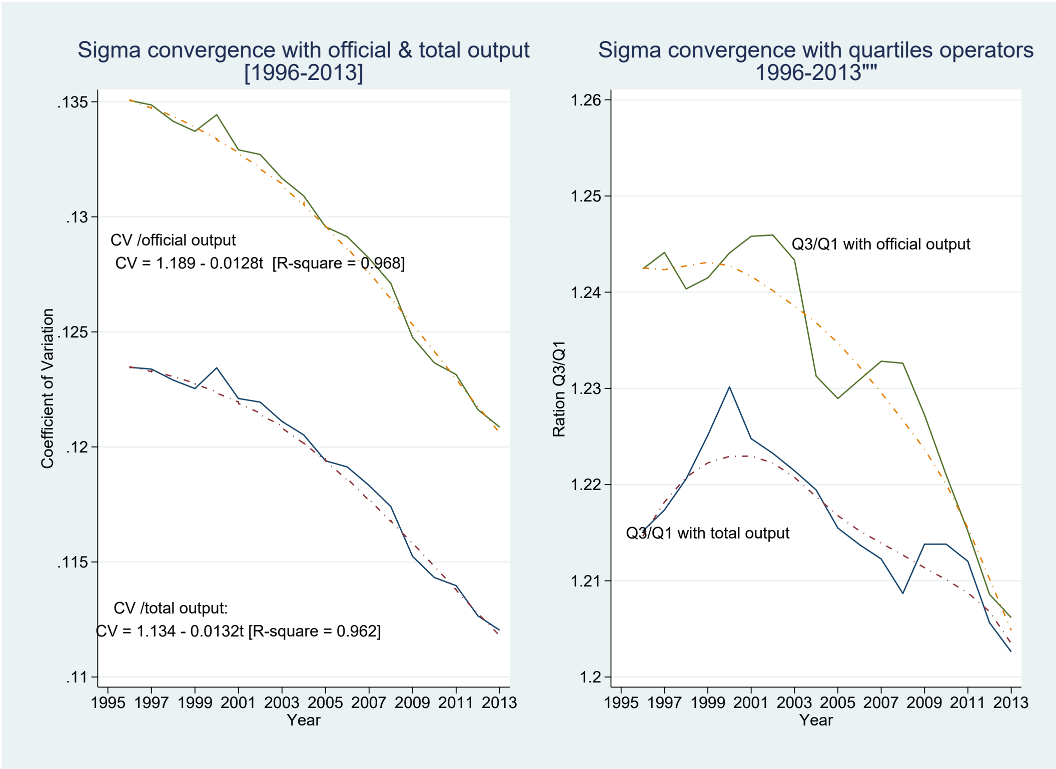
The concept of sigma convergence relates to whether the cross-country distribution of world income shrink over time or not. To check for the sigma-convergence hypothesis we first estimate the trend line of the dispersion in output levels among countries using equation 1.9.

$$CV(y_t) = \alpha_o + \alpha_1 \mathrm{t} + \varepsilon_t \tag{1.9}$$

In equation 1.9, the explained variable $CV(y_t)$ is the coefficient of variation of output (official and total levels in per-capita and logarithm terms) among economies. The explanatory variable is the time variable: $t = 1, 2, \ldots, 18$ for the period 1996-2013; ε_t is the error term. If parameter α_1 is negative, sigma-convergence does exist. Secondly, we calculate the quartiles operators $\frac{Q_3}{Q_1}$ for both the total and the official per-capita output. We then plot the CV and the ratio of quartile for the two indicators. The purpose is to check whether accounting for the informal output changes the cross-country distribution of world income as reported by the official output. The results are reported in Figure 1.9.

Considering either the coefficient of variation or the quartiles operators, the result shows that, for our two indicators of output, the sigma convergence does exist. Indeed, the figure 1.9 reports a negative and significant slope over the period. Implying that, poorer countries have not experienced income dispersion during this period. However, one notices that, with the total output indicator, the income dispersion between rich and poor countries is lower compared to that of official output. The gap between the trends of the two indicators is explained by the informal output that, when taken into account in the total output, increases countries' total wealth and reduce the income dispersion. An upmost point is the decline of the gap between the dispersion levels reported by the two indicators. The figure 1.9 shows that, the slope of the official output curve is steeper

Figure 1.9: Recorded versus total outputs: Comparison of densities



and it is catching up the curve of the total output. This is partially explained by the previously reported declining trend in the size of informal activities over the period under analysis. Also, figure 1.6 did show a negative relationship between per-capita GDP and the size of informal economy as share of GDP. This means that, as the share of informal output is declining, the share of the official output is increasing and so it is its share in the total output. This leads to a gradual decline in the gap between the trends reported by the two indicators. To find out whether the official and total output predict different trend in the σ convergence, we test the difference between the regression coefficients (α_1 for the two indicators, see in appendix; Table 1.9). The result of the test revels that there is a significant difference between the two slopes of the two regressions, implying that the total output gives different result from that of the official output. Summarizing, accounting for the informal output revises downward the income differentiation between economies over time.

1.4.2 Beta (β) - convergence

The concept of beta-convergence relates to the mobility of different economies within a given distribution of the world income. It means β convergence occurs when less developed countries grow faster than more developed countries. To check for the beta-convergence hypothesis we estimate the linear regression of the form:

$$\frac{1}{T}\log\frac{y_{i,T}}{y_{i,0}} = \beta_0 + \beta_1 \log y_{i,0} + \varepsilon_t \tag{1.10}$$

where $logy_{i,T}$ and $logy_{i,0}$ are natural logarithms of the per-capita output (official and total) in country i respectively in the last and the first year of the period under analysis; β_0 is a constant; ε_t is the error term; and T indicates the duration of the period. Convergence occurs when $\beta_1 < 0$; implying that higher initial income level negatively affects the consequent growth rate. For each of the two indicators, we conduct a linear and a quadratic regression. Table 1.6 presents the regression results.

Table 1.6: Regression result for beta convergence

	Regression with official output		Regression with total output	
	Linear	Quadratic	Linear	Quadratic
Constant	0.072***	-0.135**	0.070***	-0.154**
	(6.22)	(-1,87)	(5.53)	(-1.83)
Per_capita Income in logs	-0.0034**	0.046***	-0.0032*	0.048**
	(-2.60)	(2.69)	(-2.33)	(2.50)
Per_capita income squared in logs		-.0028***		-.0029***
		(-2,90)		(-2.68)
β	0.33%		0,31%	
Adj R-squared	0.0464	0.1028	0.0358	0.0838
Number of countries	120	120	120	120

Notes: Author's computation. Std errors are in parenthesis; $*p < 0.10$ $**p < 0.05$ $***p < 0.01$..

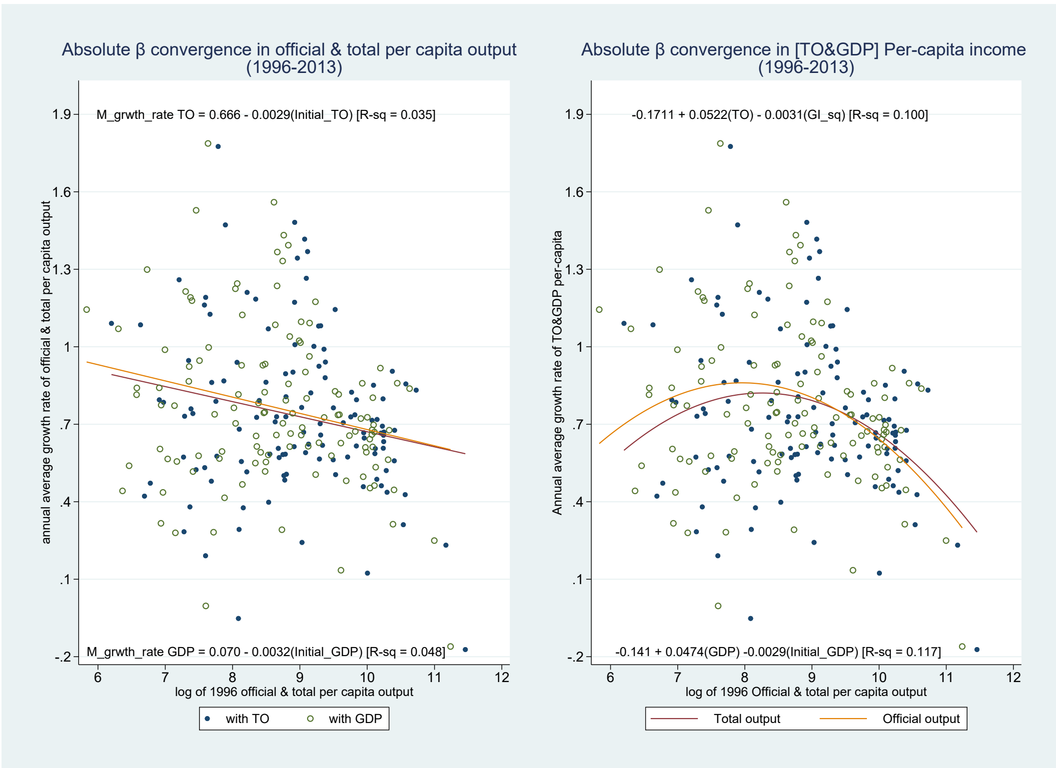
Focusing on the regression results, the linear regression reports a value of the slope, that is $\hat{\beta}_1$. The negative sign of the slope implies the existence of beta-convergence across-countries. The quadratic regression uses the equation 1.10 after adding the squared of $logy_{i,0}$ as explanatory variable[7]. The result reports a negative sign for the squared value. This implies that, bottom countries tend to diverge while top countries convergence. This result pleads for the club convergence hypothesis for which countries with similar initial conditions, here the per-capita output (official and total), converge towards the same steady state. To see whether the official and total output predict different trend in the β convergence, we test the difference between the regression coefficients for the two indicators (see result in appendix; Table 1.9)). The test reveals that there is no statistical difference between the coefficients across regressions, implying that the total output gives similar result as the official output. Summarizing, accounting for the informal economy does not significantly affect the mobility of the different economies within the given distribution of world income.

To analyze the trend in beta convergence for the two indicators, we plot the log of initial output against growth rate of the current output in Figure 1.10. The first part of the figure reports a linear and negative relationship between the initial output level

[7]Equation 9 becomes: $\frac{1}{T}\log\frac{y_{i,T}}{y_{i,0}} = \beta_0 + \beta_1\log y_{i,0} + \beta_2(\log y_{i,0})^2 + \varepsilon_t$

and its growth rate for the two indicators. That is, the beta convergence does exist for the two indicators. The second part of the figure informs that this relationship is not linear but is hump shaped, implying a divergence for bottom countries with lower level of initial income and a convergence for top countries with high initial level of income. Focusing on the total output indicator, the result shows a slower pace of divergence for bottom countries in comparison to official output. The large share of informal output in these countries could explain this fact. In general, large size of informal output tends to hide the true level of their total revenue and thus underestimating their wealth. However, when accounting for the informal output in a total output indicator, the divergence level decreases. In top countries where convergence does exist, the total output curve is above the official output, traducing a higher pace of convergence among top countries when we account for the informal output. However, the trend in official output and total output is not statistically different as previously explained. This implies that accounting for informal output does not make poor countries to grow faster than richer ones.

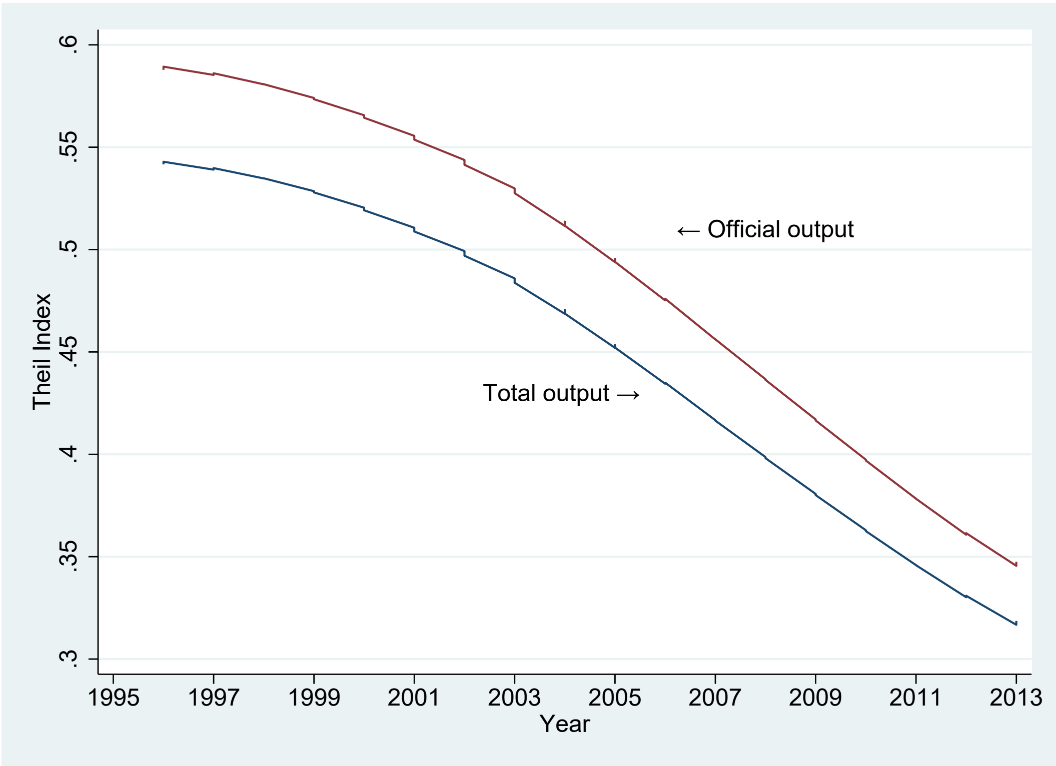
Figure 1.10: World beta-convergence of official and total outputs



1.4.3 Global inequality

The analysis of the world income inequality has called the attention of many academics (Dikhanov and Ward, 2001; Milanovic, 2002; Theil and Seale JL, 1994) and policy-makers. For example, in 2001 Human Development Report, the United Nations Development Program (UNDP), based on the Gini coefficients for a given number of countries, argues that global income inequality has been rising over time. Sala-i Martin (2006), by integrating individual income distributions for 138 countries between 1970 and 2000 and combining national accounts GDP per capita to anchor the mean with survey data to pin down the dispersion, reports a reduction in inequality. The novelty brought at this level is that in order to assess inequality across countries, we first use the official output and later we account for the informal output in a total output indicator to report the between country Theil index[8] of inequality over 1996-2013. We show in the appendixes; Table 1.11 the results of estimating the indexes for each year. Below, Figure 1.11 reports the trend in inequality for the two indicators. Over time, one observes that, although the two

Figure 1.11: Trend in the Theil index of inequality



indicators report a declining trend of inequality, the total output exhibits a lower level of inequality compared to the official output (0.455 and 0.496 on average). This result comes from the increase of countries' wealth as the informal output is accounted for in the total output. Summarizing, accounting for informal output revises downward the cross-country inequality for 4% in average over eighteen years, when looking at the Theil index. An important aspect of the yearly evolution of the Theil index for both the official and the total output is the shrink in the gap between the two inequality series reported

[8] Theil index is one of the measures widely used in the literature on inequality and it is easily decomposable (Sala-i Martin, 2006). The index has the advantage of being additively decomposable into indicators of between-and-within group inequalities. However, in order to be more explicit about the measured inequality, we computed an additional indicator of inequality, that is the GINI coefficient

by the two indicators over time. Indeed, the gap started at 4.5% in 1996 to 2.8% in 2013, implying a decrease of 53.3% in the gap of the two indicators. Still, this reduction in the gap is due to the reported decrease in the informal activities over the eighteen years around the world. This means that, as the informal output is shrinking over time, the trend of the official output is catching up the one of the total output. The same figure 1.11 also reports a drastic decrease in inequality across countries around the world. Over the studied period, the Theil index of total output decreased for about 40%. The key force behind this drastic declining trend, as reported by Milanovic (2016) might be the strong increase in GDP per capita in emerging countries, namely China and India. Since 2000, the average per capita growth rate of emerging economies has consistently been greater than average per capita growth rate of the rich world. Emerging economies had a growth rate of 4.7 percent per annum, compared with only 1% for the rich countries. However, this declining trend in inequality is not only a Chinese or Asian phenomenon. Indeed, when China and India are removed from the sample, we still observe a declining trend, though with a less steep slope. A mean-difference test on the values of the two series of the Theil index reports significant difference between them (see in appendixes; Table 1.9). This implies that overlooking the informal output exacerbates the cross-country inequality between countries.

1.5 Conclusion

Based on a sample of 120 countries for the period of 1995 to 2013, this paper we focuses on the role of informal output in the world distribution of income. To estimate the informal output, we developed different specifications of the MIMIC model where the benchmark specification disregard the GDP neither as a cause nor as an indicator in order to face the so criticized endogeneity issue. The other specifications use either the GDP or the nighttime light intensity as a proxy of global activity. The estimated informal output as a percentage of GDP reveals that the size of the informal economy varies significantly across countries where, high-income countries tend to be relatively of small size. Worldwide, the informal economy exhibits a declining trend over time and the larger is the initial size for a country, the higher is the declining rate. We then construct a Total Output indicator that accounts for both the official (formal) and the informal output. Thereafter we use it to revisit the literature on income convergence and the global inequality. Our results show that, in comparison with the official output, accounting for the informal output revises downward the level of cross-country income disparities. Focusing on the beta-convergence, we notice that, although bottom countries seem to diverge at a slower pace and top countries to converge at a raising pace, accounting for the informal output does not significantly affect the mobility of economies within the distribution of the world income. Within the framework of total output indicator, the level of global inequality across countries is lower worldwide. However, as the informal output declines over time, the gap between the trend in official and total output is also declining.

Our findings have implication from the viewpoint of both the estimates of the informal economy and the world distribution of income. From the viewpoint of the estimate of the informal economy, using the GDP or not to estimate the size of the informal economy does not change its distribution. This potentially confirms the Dell'Anno (2003)'s view on the efficacy of the MIMIC model. From the viewpoint of the world distribution of income, the total output revises downward the dispersion of poor countries and the inequality level across-countries with potential implication on the world and countries' development policies and strategies. Although the total output revises downward the level of dispersion and inequality, as the size of the informal output decreases, the results reported by the two indicators, namely the official and the total output converge. The MIMIC approach we used stills bear some weaknesses. The latter, in the opinion of most authors, result mainly from the "initial stage of research" in this field of informal economy and its measurement instead than the inappropriateness of the technique. Overcoming these caveats to gather informal output estimates will be most helpful for future researchers.

1.6 Appendix

1.6.1 Robustness on the structure of the MIMIC model

Table 1.7: Robustness with two indicators

Causes variables	Tax	*Human capital*	Inflation	Poverty
Intercept	-1.535***	-0.698***	-1.575***	-0.712***
Taxe per-woker	-	0.0566	0.0611	0.0605
Inflation	0.158***	0.189	-	0.213
Property right	-0.156***	-0.253***	-0.147***	-0.263***
trade freedom	-0.155***	-0.211***	-0.156***	-0.223***
Human capital	-0.0284*	-	-0.0268*	-0.0256
Poverty	0.0118	0.0113	0.0112	-
Indicator variables				
Currency outside of banks	1	1	1	1
Taxe per-woker	-0.00105	-	-	-
Human capital	-	-0.0220	-	-
Inflation	-	-	0.0290***	-
Poverty	-	-	-	-0.0206
Countries	120	120	120	120

Notes: Author's computation. $^{*}p < 0.10$ $^{**}p < 0.05$ $^{***}p < 0.01$.

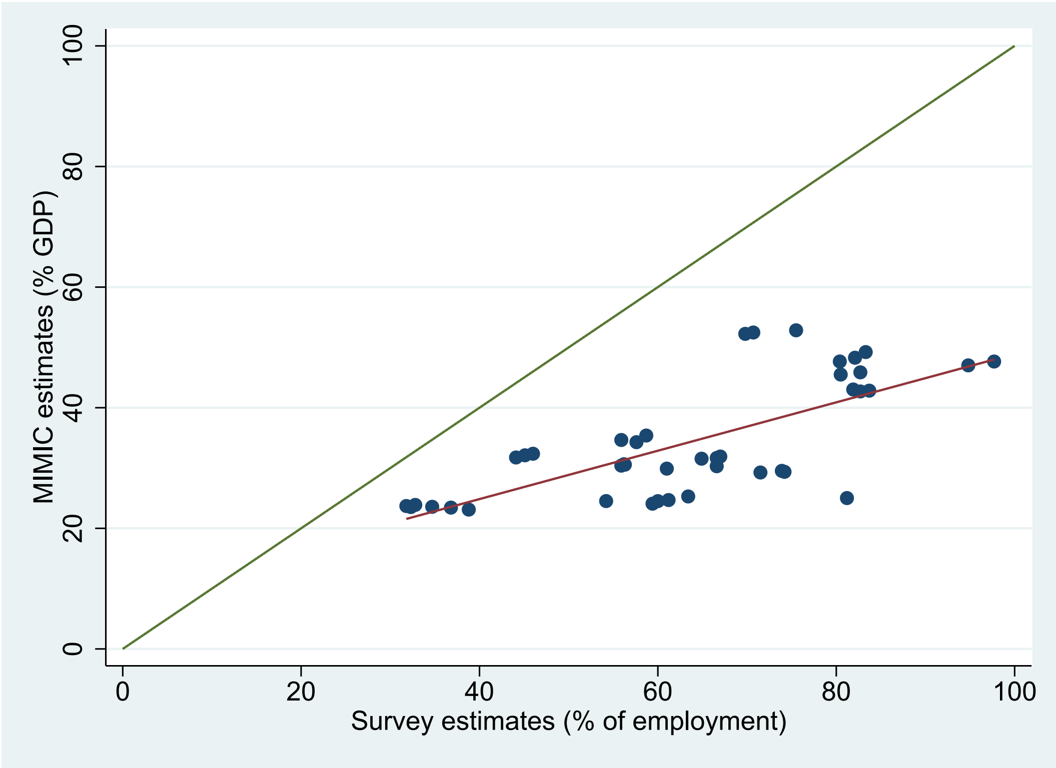
Table 1.8: Robustness with three indicators

Causes variables	Tax	*Human capital*	Inflation	Poverty
Intercept	-1.729***	-0.912***	-1.782***	-0.920***
Taxe per-woker	-	-0.0609	0.0623*	0.0645
Inflation	0.150**	0.168	-	0.192
Property right	-0.0130	-0.207***	-0.138***	-0.219***
trade freedom	0.0245	-0.184***	-0.151***	-0.196***
Human capital	-0.0287**	-	-0.0272**	-0.0268
Poverty	0.0309	0.0109	0.0112	-
Indicator variables				
Currency outside of banks	1	1	1	1
Labor force participation	-0.00103	-0.000925	-0.00104	-0.00105
Taxe per-woker	0.00994	-	-	-
Human capital	-	-0.0220	-	-
Inflation	-	-	0.0290***	-
Poverty	-	-	-	-0.0252
Countries	120	120	120	120

Notes: Author's computation. $^{*}p < 0.10$ $^{**}p < 0.05$ $^{***}p < 0.01$.

1.6.2 Correlation between MIMIC estimates and Survey estimates

Figure 1.12: Comparison between MIMIC estimates (Y axis) and survey estimates (X axis))



1.6.3 Result of the test of coefficients and mean difference

To test the significance of the difference between coefficients reported by the regressions on official and total output, we applied a seemingly unrelated estimation test that is a test for intramodel and cross-model hypothesis based on the Hausman specification (see Clogg, Petkova, and Haritou, 1995) [9] . For the Inequality, we applied a mean difference test.

[9]Clogg, C. C., E. Petkova, and A. Haritou. 1995. Statistical methods for comparing regression coefficients between models. American Journal of Sociology 100: 1261–1312. (With comments by P. D. Allison and a reply by C. C. Clogg, E. Petkova, and T. Cheng).

Table 1.9: Test of coefficients and mean difference

(Sigma Convergence: Linear regression)	
Test : α_1(Total output) $-$ α_1(Official output) = 0	
Chi2(1)	6916.04
Prob > chi2	0.0000
(Beta Convergence: Linear regression)	
Test : β_1(Total output growth) $-$ β_1(Official output growth) = 0	
Chi2(1)	1.80
Prob > chi2	0.1791
(Beta Convergence: Quadratic regression)	
Test : β_1(Total output growth) $-$ β_1(Official output growth) = 0	
Chi2(1)	0.26
Prob > chi2	0.6096
Test : β_1(Total output growth square) $-$ β_1(Official output growth square) = 0	
Chi2(1)	0.01
Prob > chi2	0.9160
(Inequality (Theil index))	
(Mean difference test : Theil (Total output) vs (Official output)	
Inequality (Mean official output)	0.4913
Inequality (Mean total output)	0.4504
Inequality (Difference)	0.0408
Prob —D> 0—	0.0000

Notes: Author's computation

1.6.4 Result of Kolmogorov Smirnov equality-of-distributions test

Table 1.10: Kolmogorov Smirnov equality-of-distributions test

Smaller group	Difference	P-value	Smaller group	Difference	P-value
Alternative_3	0.0028	0.983	Schneider(2017)	0.0313	0.123
BM result	-0.0259	0.234	BM result	-0.0212	0.381
Combined K-S:	0.0259	0.462	Combined K-S:	0.0313	0.246

Notes: Author's computation

The first line tests the hypothesis that estimates for alternative_3 contains smaller values than those for the Benchmark specification. The largest difference between the distribution functions is 0.0028. The approximate p-value for this is 0.983, which is not significant.

The second line tests the hypothesis that estimates for alternative_3 contains larger values than those for the Benchmark specification. The largest difference between the distribution functions in this direction is -0.0259. The approximate p-value for this small difference is 0.234.

Finally, the approximate p-value for the combined test is 0.462. showing that that these distributions are not statistically different

1.6.5 GINI and Theil indexes of inequality

Table 1.11: GINI and Theil indexes of inequality

	GINI INDEX			THEIL INDEX		
Year	Total Output	Official Output	Gap	Total Output	Official Output	Gap
1996	0.5509	0.5706	-0.0197	0.5402	0.5861	-0.0459
1997	0.5494	0.5693	-0.0199	0.5372	0.5830	-0.0458
1998	0.5483	0.5683	-0.0200	0.5386	0.5847	-0.0462
1999	0.5436	0.5652	-0.0216	0.5320	0.5799	-0.0479
2000	0.5436	0.5646	-0.0210	0.5318	0.5772	-0.0454
2001	0.5374	0.5587	-0.0213	0.5194	0.5642	-0.0448
2002	0.5331	0.5537	-0.0206	0.5094	0.5514	-0.0419
2003	0.5253	0.5471	-0.0217	0.4927	0.5363	-0.0435
2004	0.5174	0.5400	-0.0225	0.4767	0.5204	-0.0437
2005	0.5078	0.5320	-0.0243	0.4571	0.5032	-0.0461
2006	0.5005	0.5225	-0.0220	0.4398	0.4826	-0.0428
2007	0.4885	0.5106	-0.0221	0.4155	0.4574	-0.0419
2008	0.4818	0.5014	-0.0196	0.3998	0.4375	-0.0377
2009	0.4643	0.4831	-0.0188	0.3693	0.4044	-0.0350
2010	0.4551	0.4739	-0.0188	0.3525	0.3870	-0.0345
2011	0.4496	0.4670	-0.0175	0.3416	0.3739	-0.0323
2012	0.4436	0.4607	-0.0171	0.3319	0.3630	-0.0312
2013	0.4378	0.4537	-0.0158	0.3228	0.3514	-0.0286
Max	0.5509	0.5706	-0.0158	0.5402	0.5861	-0.0286
Min	0.4378	0.4537	-0.0243	0.3228	0.3514	-0.0479
Mean	0.5043	0.5246	-0.0202	0.4505	0.4913	-0.0408
textTotal decrease	0.2053	0.2049	0.1955	0.4026	0.4005	0.3759

Notes: Author's computation

1.6.6 Summary statistics on the estimates of informal sector

Table 1.12: Summary statistics on the size of informal sector

Country	Mean	S.D	Min	Max	Country	Mean	S.D	Min	Max
Albania	31.4	1.5	30.0	35.2	Ecuador	27.8	2.1	24.6	30.8
Algeria	27.7	1.8	25.2	30.6	Egypt	37.0	1.4	34.2	38.6
Argentina	20.6	2.6	15.8	23.6	El Salvador	45.6	1.7	42.8	48.4
Armenia	49.9	2.9	45.8	54.4	Estonia	31.3	1.1	29.8	33.2
Australia	15.0	0.3	14.6	15.6	Fiji	31.5	1.4	29.4	33.8
Austria	9.9	0.1	9.8	10.2	Finland	17.6	0.2	17.2	18.0
Bahrain	16.7	0.2	16.2	17.2	France	15.6	0.2	15.4	15.8
Bangladesh	39.7	1.9	36.2	42.0	Gabon	48.2	1.5	46.4	50.6
Barbados	35.0	0.7	34.2	36.6	Gambia	44.7	3.2	41.0	49.2
Belgium	22.4	0.4	21.8	23.0	Ghana	36.1	3.1	31.4	41.0
Belize	44.8	1.0	43.0	46.8	Greece	27.4	0.9	26.0	29.0
Benin	48.7	2.1	46.2	52.4	Guatemala	48.5	2.4	45.6	52.8
Bolivia	59.1	6.7	47.6	68.0	Haiti	49.1	4.3	43.2	56.6
Botswana	36.7	1.6	34.6	39.4	Honduras	45.9	2.5	42.8	50.2
Brazil	34.4	1.6	31.8	37.0	Hong Kong	16.2	0.2	15.8	16.4
Bulgaria	25.8	2.3	24.0	33.8	Hungary	25.2	0.8	24.0	26.8
Burkina Faso	39.4	1.3	37.4	41.6	Iceland	16.1	0.5	15.4	17.0
Cambodia	57.6	2.6	54.6	61.4	India	25.4	0.4	24.4	26.0
Cameroon	31.7	0.5	31.0	32.6	Indonesia	15.9	1.3	14.4	18.4
Canada	16.7	0.2	16.4	17.0	Iran	15.6	1.8	12.6	18.2
Chile	21.3	0.6	20.4	22.2	Ireland	17.5	0.3	17.2	18.2
China	16.1	0.7	14.8	16.8	Israel	21.7	0.3	21.2	22.4
Colombia	32.9	1.7	31.2	36.2	Italy	27.7	0.8	26.8	28.8
Congo - Brazzaville	42.2	3.8	37.8	48.6	Ivory Coast	48.7	1.2	46.0	50.0
Costa Rica	24.7	1.1	23.2	26.6	Jamaica	31.4	2.2	28.4	34.8
Croatia	31.5	0.8	30.8	34.2	Japan	10.7	0.1	10.6	10.8
Cyprus	27.3	0.5	26.4	28.2	Jordan	16.9	0.5	16.4	17.6
Czech Republic	17.7	0.3	17.4	18.4	Kazakhstan	37.5	3.2	33.0	42.2
Denmark	18.5	0.3	18.0	19.0	Kenya	29.2	1.5	26.6	31.4
Dominican Republic	32.3	1.7	30.4	35.4	Kuwait	17.7	0.9	16.6	18.8
Laos	35.1	3.6	31.6	43.6	Romania	23.9	2.5	21.6	29.6
Latvia	33.5	1.3	31.6	35.2	Russia	29.8	3.6	25.6	36.0
Lesotho	29.8	1.7	27.4	31.8	Rwanda	42.2	1.7	39.0	45.4
Lithuania	35.2	0.8	33.8	36.4	Saudi Arabia	16.9	0.7	15.8	18.0
Luxembourg	10.0	0.2	9.6	10.4	Senegal	49.4	1.1	47.6	51.2
Madagascar	38.6	1.6	36.4	41.8	Singapore	13.5	0.2	13.2	13.8
Malawi	32.2	2.5	28.2	37.0	Slovakia	17.1	0.4	16.6	18.0
Malaysia	31.4	1.0	30.0	33.0	Slovenia	28.9	0.7	28.0	30.2
Mali	39.6	1.6	37.4	41.8	South Africa	25.0	0.9	23.6	26.4
Malta	28.0	0.5	27.2	28.8	South Korea	28.6	0.9	27.6	30.2
Mexico	29.3	1.1	27.8	31.6	Spain	23.8	0.4	23.2	24.6
Moldova	37.7	3.6	32.8	43.6	Sri Lanka	43.8	2.8	39.4	47.2
Mongolia	13.8	1.4	11.8	15.8	Swaziland	36.1	2.0	32.6	39.2
Morocco	33.3	1.7	30.6	36.4	Sweden	18.8	0.1	18.6	19.0
Mozambique	40.3	2.1	37.0	44.6	Switzerland	8.4	0.1	8.4	8.6
Namibia	27.3	2.1	24.0	30.8	Tanzania	52.9	3.9	47.6	60.2
Nepal	39.6	2.5	35.6	43.6	Thailand	47.6	2.1	44.6	50.6
Netherlands	13.6	0.2	13.4	14.0	Trinidad and Tobago	33.0	1.8	30.0	35.4
New Zealand	13.0	0.2	12.6	13.2	Tunisia	36.3	0.9	34.8	37.6
Nicaragua	40.9	2.5	36.0	44.2	Turkey	25.4	3.9	21.6	34.2
Niger	37.1	1.0	35.8	39.0	Uganda	44.3	2.5	40.6	48.6
Nigeria	46.5	3.1	43.6	52.4	Ukraine	38.8	3.1	34.4	44.6
Norway	18.8	0.5	18.0	19.6	United Arab Emirates	23.6	1.1	21.8	25.0
Pakistan	30.7	1.4	28.2	33.2	United Kingdom	13.2	0.2	13.0	13.6
Panama	62.9	2.4	58.4	66.2	United States	9.3	0.2	9.0	9.6
Paraguay	31.6	1.9	29.2	35.6	Uruguay	46.9	2.0	43.4	49.6
Peru	54.1	2.0	52.0	58.2	Vietnam	19.4	0.8	18.2	20.4
Philippines	37.8	2.0	35.0	40.4	Yemen	29.6	1.8	27.0	32.2
Poland	28.4	0.8	27.6	29.6	Zambia	43.6	3.8	38.4	49.6
Portugal	25.1	0.4	24.4	25.8	Zimbabwe	58.9	4.0	51.2	65.2

Notes: Author's computation

Chapter 2

Earning Structure and Heterogeneity of the Labor Market: Evidence from the Democratic Republic of the Congo[1]

[1]We would like to thank Dejemeppe Muriel, Docquier Frédéric and two anonymous referees for their invaluable comments. We are grateful towards UCLouvain for supporting this research through the "Fond de coopération au développement"

2.1 Introduction

Due to sectoral wage determination process, labor market heterogeneity in developing economies has received a great deal of attention in the literature. This stems from the fact that, it exhibits considerable diversity that includes substantial segments (hereafter called sectors) of different characteristics (Fields, 2009) that depend on the specific environment in which workers operate (Deakin, 2013). Though in a general neo-classical framework, the level of earning and earnings of the workforce is determined by demand for and supply of labor, in developing countries, this is affected largely by strict labor market sectors and strong entry barriers across different sectors of the labor market. In self-employment activities, workers enjoy non-earning features such as greater flexibility (in terms of working hours, work relationships, responsibilities, etc.) and maximize more their utility then their earnings. In the public sector, earnings are mainly determined through a political process or service regulations (Gunderson, 1979). In the private formal sector, earnings are determined by the demand and supply conditions of the labor market and in the private informal sector, earnings are said to be low and volatiles (Günther and Launov, 2017). These differences bring a natural interest in the type and the size of the labor market's sectors and mainly, their earning outcomes. Because these outcomes are a robust indicator of the livelihood status of the workforce and hence, of the populations who derive all or the great bulk of their earnings from their work on the labor market.

Drawing on an empirical case study on DRC, this paper provides insights on the functioning of labor markets in Sub-Saharan African (SSA) countries. With a selectivity-corrected Mincerian earning equation, we investigate the role of individuals characteristics in determining their earnings. An Oaxaca-Blinder decomposition of the earning gap between sectors is carried out, allowing us to disentangle differences in characteristics (e.g. education) and difference in the return on these characteristics across sectors as drivers of the earning gap. Our initials descriptive statistics give five different sectors with two higher-paid completely formal sectors (Public and Private formal) and two lower-paid largely informal sectors (Self-employment and Private informal) that are significantly heterogeneous across them when it comes to earnings. Based on earning equations, we show that, though the effect of education on earnings provides a clear support to the human capital theory, basic education has no significant impact on earnings in higher-paid sectors. Likewise, tertiary education matters for earnings in lower-paid sectors as well. We then decompose the earning gap across sectors taken in pairs and show that workers of the lower-paid sectors earn less not only because they are less skill-endowed but also because they earn lower returns on such skills. However, when higher-paid and lower-paid sectors are concerned, the coefficient effect in the upper end of the distribution is negative, implying that the labor market provides an *informal employment earning premium* to some individuals of the lower-paid sectors whose, given their characteristics, would not do better in the higher-paid sector.

We conduct our analysis on the Democratic Republic of Congo (DRC), a typical example of an economy where labor market is badly out of balance. Galloping demography is continuously increasing the demand for jobs while, since 1990 firm's labor demand has been falling steeply as a result of looting, wars and other shocks to the economy. This state of affairs combined with high poverty, high unemployment[2] and inexistent public provisions for unemployment insurance has favored the attractiveness of public jobs and the emergence of the informal sector where self-employment dominates. As a consequence, in the formal sector, salaries are negotiated in a context of strong demand for employment, 78.4% of formal firms report to compete with informal ones (World-Bank, 2005) and the wage bill constitutes now the largest item in public sector spending as the scantiness of private jobs and the economic and political instability have fostered the rush for public jobs. Accordingly, the country is gaining insights into the factors influencing the self/paid-employment selection, the public/private sector selection, and sectoral earning determination processes for paid employees. This provides rich evidence of the labor market heterogeneity that stands out as an interesting case for analysis in the SSA context.

Due to important consequences on economic growth, poverty and inequality, the economic literature has intensively explored the heterogeneity of the labor market and earning differences between type of workers. For example, labor market and earnings display substantial heterogeneity with respect to gender (Oaxaca, 1973; Polachek and Xiang, 2009), education levels (Beaudry and David, 2003), self/paid-employment (Bernhardt, 1994) and to public/private sector (Christofides and Panos, 2020; Tansel, 2005). There is also a positive formal-informal earning differential (Falco et al., 2010) which is explained both by individuals characteristics and by the return to these characteristics (Garcıa, 2017). The limit of these studies is that, they concentrate on a dual analysis where market is a matter of two to three sectors, which is a quite narrow selection of employment opportunities in the context of SSA. Furthermore, they completely exclude from the analysis a substantial number of individuals that are outside the labor market and can be defined as unemployed.

In this paper, we provide the first study on labor market heterogeneity and earning differential in DRC. The contribution of this study is twofold: First, we begin with the structure of the labor market and then link earnings to that structure. In so doing, we let workers decide first whether to enter the labor force by accounting for the unemployment

[2]See IMF (2015)'s country report for detailed statistics on unemployment and poverty in DRC

segment. Indeed, given the high rates of unemployment in developing countries, on the one hand it points to the fact that being at the bottom end of the earnings distribution may not be the worse outcome for those in the labor market. On the other hand, many individuals either work in the household for no explicit pay, or as unpaid apprentices or are simply paid in kind. Therefore workers choose between paid and unpaid work in first stage, rather than between sectors as in the aforementioned studies. Second, we allow individuals to have a larger variety of choices and combinations on the labor market. For instance, they choose whether to work as self-employed or as wage employees and within wage employment whether in the public or the private sectors. Within the private sector, they can also choose between the formal or the informal sector. Accounting for all the possible sectors is important because high degree of heterogeneity imposes greater resource costs on the economies, especially these of developing countries, by causing the failure of the market to move the 'right' resources into 'right' sectors (Berry and Sabot, 1978). Of course, as some of these sectors choices are constrained by rationing, workers may not necessarily end up where they wish. We then examine the earning differential and its decomposition across the observed sectors taken in pairs for a detailed investigation. With such a comprehensive analysis, we aim to help developing countries to understand how do labor market operate, to identify the extent of its heterogeneity and how the latter is responsible for labor market outcomes and distortions that lead to earning differentiation across workers.

In the next sections, we outline the data used and give some descriptive statistics. In section 2.3 we explain our estimation procedure before presenting and discussing our results in section 2.4. Section 2.5 concludes.

2.2 Data and descriptive statistics.

In 2005 and 2012, the DRC's National Institute of Statistics (INS) conducted, in partnership with different actors (including Afristat and the World Bank), a broad household and expenditure survey that followed the so-called 1.2.3 methodology. Each of these numbers refers to a collection phase. Phase 1 provides detailed information on employment, unemployment, household and individual socio demographic characteristics. Phase 2 specifically gathers information on the characteristics of firms and firm owners in the informal sector. Phase 3 is a survey on household expenditures. In this paper, our sample focuses on the phase_1 of the 2012 survey limited to individuals between the age 15 and 65 years involved or not on the labor market. The sample covers about 88 600 individuals from 11 Provinces of the DRC. The survey provides data on five sectors of the labor market. 1) Individuals out of the labor market; this sector includes both the participants and non-participants. The participants are unemployed who are looking for employment and are available upon request. The non-participants are unemployed individuals unavailable for work, due to studies, disabilities, early retirement and others. 2) The self-employed[3], who at the time of the interview, work in a business that they was owned entirely or partly. 3) Public workers hired by the public administration. 4) Private formal employees and 5) Private informal employees. Such data gives an opportunity to study the heterogeneity of the labor market and the resulting earning differential. To check whether these sectors are different when it comes to earnings, we run a simple linear regression with monthly log-earning as the dependent variable. The independent variables are the different labor market sectors namely; self-employment, public workers, private formal and private informal workers entered as dummies. In this paper, we use monthly earnings instead of hourly earnings[4]. Given the often constrained working hours in the informal and self-employment sectors, monthly earnings reflect better earning opportunities in some sectors than hourly earnings.

In Table 2.1, the summary statistics provided reveal that there are considerable differences among workers employed in these sectors. It shows that self-employment and unemployment sectors are markedly large. Indeed, as previously said, the unemployment sectors includes non-participants and participants. This last category represents 8.5% of the total unemployed individuals. For the self employment sector, its magnitude comes from the fact that it includes all individuals with small business and street or market vendors who work for themselves. More than 90% of this sector is made up of informal entrepreneurs. As regard to earnings, that is the monthly earning of the main activity held in the last 30 days, the data show a remarkable difference in average earnings across the different sectors. Indeed, Table 2.1 shows that the national average monthly earning is CDF 51 946.8 . Furthermore, before the change that occurred in 2018 , the minimum wage in DRC amounted to CDF 1680 per day (i.e., a monthly earning of CDF 50 400). If we consider the minimum wage used by the "Institut National de la Statistique" (INS, 2014), after updating the minimum wage by taking into account the annual increase implemented by the law and the increase in the general price level, the minimum wage in DRC is set to 54 128 CDF. Taking the two statistics, we see that there is, on the one side two "higher-paid" sectors, namely public and private formal sectors, where the average monthly earning is above the national average and far above the updated minimum wage. Individuals in the two higher-paid sectors are 100% formal workers. On the other side, we have two "lower-paid" sectors, namely Self-employment and Private informal sectors,

[3] Includes independents with and without wage employee.

[4] Indeed, many activities in the informal as in self employment sector are part-time jobs by nature and employees in those sectors could not easily increase their earning by simply providing more labor hours per month.

where the average monthly earning is below the national average and even bellow the updated national minimum wage. Furthermore, individuals in the two lower-paid sectors are more than 95% informal workers.

This variation in earning between higher and lower-paid sectors is explained by some important differences between sectors. For example, in regard to education, we see that public and private formal workers are more endowed for secondary and tertiary education than workers in others sectors of the labor market. Indeed, on average, higher-paid workers have two times the years of schooling compared to lower-paid-sectors. Private informal workers are the less educated even compared to the unemployed. Low skilled individuals escape unemployment in informal sector, while most high skilled ones often prefer to stay unemployed until they find a formal job. This is stated by Reyes et al. (2017), skilled individuals could have the ability to sustain a longer job search to find a more suitable match. Indeed, those with higher education typically have the economic backing from their families to engage in a longer job search. Moreover, unemployed, self-employed and private informal workers are more likely to be female and less educated, while most self-employed and private informal workers live in rural areas. For other variables, the Table 2.1 shows that, individuals whose father works as public worker or as self-employed are more likely to be public servant or self employed. The descriptive statistics show that 48.63% of self-employed individuals have a father working as self-employed. We do not test the differences between these characteristics across the different sectors of the labor market. This is because if earning is different across sectors, the cross earning gap can result from how the labor market discriminates between workers of same characteristics across different sectors or how the labor market reward individuals of different characteristics across sectors.

2.3 Estimation strategy.

We want to examine the heterogeneity in the labor market and study the earning differential between workers across sectors of the labor market. To start with, we estimate the following equation in order to illustrate earnings differential across sectors i[5] of the labor market.

$$Y_i = \lambda_o + \lambda_1 SE_i + \lambda_2 PS_i + \lambda_3 PFS_i + \lambda_4 PIS_i + \epsilon_i \tag{2.1}$$

Y_i stand for the monthly earning in log form, SE; Self-Employment, PS; Public sector, PFS; Private Formal Sector, PIS; Private Informal Sector and ϵ is the error term.

Next, we examine earning differential across sectors and explore the determinants of

[5]Here, we only consider four sectors by excluding the unemployment sector as it not concerned by earnings.

monthly earnings within each sector by regressing monthly earnings on education, controlling for personal and geographical characteristics. We rely on a quantile decomposition methodology (as introduced by Koenker and Bassett (1978)), allowing us to analyze the earning gap across sector at different points of the distribution. We follow Firpo et al. (2009)'s method to compute unconditional quantile and decompose the effect of each covariate at different parts of the distribution. Unconditional Quantile Regression (UQR) uses the influence function[6] as the dependent variable at different quantiles. In the case of quantile q_τ of an outcome variable Y and explanatory variables X, the The Recentered Influence Function[7] (RIF) is

$$RIF(Y; Q_\tau) = Q_\tau + \frac{\tau - I(Y \leqslant Q_\tau)}{f_Y(Q_\tau)} \tag{2.2}$$

Where $f(.)$ is the density and 1(.) indicates that Y is at or above the quantile q_τ. It is recentered as q_τ is added and as a consequence, the expected value of the RIF is q_τ itself. Indeed, Firpo et al. (2009) show that this property extends to the conditional-on-controls RIF and the earning equation of the RIF model at quantile τ, with $\tau \in (0, 1)$, is then:

$$RIF(Y; Q_\tau) = X\beta_\tau + \nu_\tau \tag{2.3}$$

Where Y is the natural logarithm of monthly earning and X is a vector of K explanatory variables (including the constant), β_τ is the corresponding coefficient vector and ν_τ is the corresponding error term. As explanatory variables in the earning equation we include the variables usually included in a Mincerian earning regression-education (These are dummy variables for the highest completed education of the individual; basic, secondary and tertiary education) and control for gender, age (and age square), experience, and dummies of job trained and leaving in urban areas. Furthermore, dummy variables for province of residence are included to control for differentials in cost of living and labor-market opportunities.

[6]The influence function measures how robust a distributional statistic is to outliers. it is given by : $IF(Y; Q_\tau) = \tau - I(Y \leqslant Q_\tau)/f_Y(Q_\tau)$, where $I(Y \leqslant Q_\tau)$ is an indicator function taking value one if the condition in (.) is true, zero otherwise.

[7]Since the explanatory variables do not enter into the transformation of equation (2), although the X's in the model change, the interpretation of the estimated effects does not vary, and so alternative models can be compared and different sources of socioeconomic inequality incorporated. The main advantage of this method over conditional regression is that the estimated effects do not depend on the set of explanatory variables in the model. Moreover, as in the conditional regression, the estimates are robust to outliers.

2.3.1 Accounting for selection

The distribution of workers among the four different sectors is not random. In estimating the earning equations, the selection into different sectors for which we observe earnings must be taken into account. Potential biases could result from ignoring sample selection (Heckman, 1974). To take this into account, we assume that individuals face five mutually exclusive choices: Unemployed ($j = 0$), Self-employment ($j = 1$), public administration employee ($j = 2$), Private formal employee ($j = 3$) and Private informal employee ($j = 4$). Worker's tastes and preferences as well as human capital and other characteristics will determine the sectoral choice. We assume a conditional multinomial logit model for the probability that the individual chooses alternative j relative to that of being in an arbitrarily chosen reference sector (which is non-employment) as follows:

$$P_j = exp(Z\alpha_j)/\left[1 + \sum_{j=1}^{4} exp(Z\alpha_j)\right] \tag{2.4}$$

Where Z is is a vector of explanatory variables affecting sectoral choice and α_j is a vector of unknown parameters of the alternative j. The selection equation is in reduced form, in the sense that the earning rate is not included as an explanatory variable and we only consider someone's actual state. Information on preferred labor market state or job search is not taken into account. We adopt the two-step estimation method developed by Lee (1983) and Trost and Lee (1984). In the first stage, we estimate the sectoral choice probabilities by maximum likelihood logit method and construct the selection term for the alternative j as follows:

$$\lambda_j = \phi(H_j)/\Phi(H_j) \tag{2.5}$$

Where $H_j = \Phi^{-1}(P_j)$, ϕ is the standard normal density function, and Φ is the standard normal distribution function. In the second stage, the estimated λ_j is included among the explanatory variables of the earning equations. The implied earning equations are then estimated using a RIF regression, providing consistent estimates of the parameters. The term λ_j plays the same role as Mill's ratio in the usual Heckman (1979) procedure but it is here quantile-specific Töpfer (2017). A statistically significant positive value of the self-selection term indicates that persons who – after controlling for observable characteristics – are more likely to work in sector j also have, ceteris paribus, higher expected wages in this sector (see Dimova et al. (2011)). A negative estimate, by contrast, would indicate that persons who are more likely to work in sector j have lower expected wages in this sector.

In order to achieve identification of the parameters in the multinomial logit, we introduce in its equation variables that influence labor-force participation and sector choice but may be excluded from the earning equations. We include; unearned income (whether a family member owns shares, securities or financial investments), risk aversion (proxied by whether the individual has a health insurance or not), whether or not a close member of family lost his job and dummies of father working in the public sector and father working as self-employed. These variables provide labor-market information and should be associated with labor market participation but not earnings. For instance, as suggested by Schultz (1990), unearned income is expected to reduce the probability of participation by raising the shadow value of a person's time in non-market activities and in self-employment. A risk averse individual will prefer being employed to working as self-employed or being unemployed (Christofides and Panos, 2020). Having a father working in the public sector may increase the probability of employment of an individual in the public administration sector, and similarly for the other variables and sectors. For these variables, the relevance of the exclusion restrictions in terms of their predictive power of sector choice can be directly tested from the model estimates. However, no formal over-identification test has been developed for this specific framework. We are aware of the fact that, as usual, the validity of our exclusion restrictions is debatable, because it can be argued that the selected variables might be related to unobserved determinants of earnings. This would be especially true in the case that the list of control variables in the earning equation(s) does not include all the relevant features of the current sector. Nevertheless, in Table 2.6 in the appendixes we provide an informal means of testing both for the relevance of the exclusion restrictions and for the excludability of the aforementioned variables from the outcome equations.

2.3.2 Decomposition strategy

The earning gap between workers of different sectors can exist not only because of disadvantages in terms of remunerated characteristics, but also because the returns to these characteristics can be different for workers of different sectors. To assess the earning gap between sectors, the Oaxaca (1973) and Blinder (1973)'s decomposition methodology is implemented. Based on separate estimation of earning equations, the Oaxaca and Blinder's decomposition generates the counterfactual on the basis of which the difference in average earnings between sectors is broken into two additive components: one attributable to differences in average characteristics of the individuals (the characteristics effect), and the other to the differences in the rewards associated with these characteristics (the coefficient effect). The method by Firpo et al. (2009) allows to conduct a detailed Oaxaca and Blinder's type decompositions using unconditional quantile regression that accounts for selection (Töpfer, 2017). Indeed, the earning gap is, as in the standard two-fold Oaxaca-Blinder decomposition, decomposed into characteristics (explained) and coefficients (unexplained) components. The decomposition for the τth quantile $\hat{\triangle}_\tau = \overline{RIF}_j$ - $\overline{RIF}_i$ takes the form hereunder:

$$\underbrace{\hat{\triangle}_\tau}_{Earning\ gap} = \underbrace{\bar{X}_j - (\bar{X}_i)\hat{\beta}_j}_{Characteristics\ effect} + \underbrace{\bar{X}_i(\hat{\beta}_j - \hat{\beta}_i)}_{Coefficient\ effect} \tag{2.6}$$

Where $\overline{RIF}_j$ and $\overline{RIF}_i$ are the estimated mean $\widehat{RIF}$ of sectors j and i respectively. Taking one sector (the sector with high mean income) as the reference category, the method has the advantage that it computes detailed decomposition and allows for the unconditional mean interpretation of the coefficient estimates. When the earning gap is mainly attributable to the coefficient effect, earning differences are attributable to differences in returns on skills between sectors. When, the earning gap is primarily explained by the characteristics effect, earning differences between sectors are due to differences in workers' endowments. In our estimation, we use a bootstrap procedure with fifty replications to estimate standard errors for the estimated coefficients, for earning gap as well as for the characteristics and coefficient effects.

2.4 Estimation Results.

In this section, we present the results obtained from the Multinomial logit to examine the probability of being in a given sector, the regression result of earning's determinant in each sector of the labor market on the $10th$, $50th$ and $90th$ percentiles and the results of the earning decomposition across the different sectors of the labor market. But first, the results of equation (1) where y stand for the monthly earning in log form, and each sector is entered as dummy variable are presented in Table 2.2. The last four columns of Table 2.2 represent different reference groups for the four sectors of the labor market where workers get monthly earnings. Given the significance of the coefficients, the regression results show that the four sectors are statically different from each other, when monthly earning is considered. A separate earning function is then needed for each of the sectors to capture the sector-specific earning determinants.

2.4.1 Multinomial Logit Estimates

Multinomial logit estimates of sector choice are shown in Table 2.3. The table gives the marginal effects of each variable on the probability of joining a particular sector calculated at the mean values of the variables, with the unemployment sector as the reference group. The results indicate that, males are less likely to work in private informal sector but more likely to be in the three others sectors, particularly in the self-employment. Basic education does not matter for working in the formal private sector meanwhile it decreases the probability of being employed in the private informal and the public sectors compared to unemployed individuals. However, individuals with basic education are more likely to work as self-employed. On the one hand, secondary and tertiary education significantly increase the probability of working in the private formal and in the public sector. Furthermore, the higher the education level the higher its contribution to the participation in the two aforesaid sectors. On the other side, these individuals with secondary and tertiary education are less likely to work informally or being self-employed and the higher the education level the higher its probability of no participation in the private informal or self-employed sector. As the two sectors, namely self-employment and informal employment are the lower-paid sector of the labor market, educated individuals tend to avoid them while queuing for public and private formal jobs. Except for being employed in the private formal sector, experience and marital status significantly increase the probability of employment in all the others three sectors and mainly the self-employment sector. Probably that, with more experience, private formal workers who have accumulated wealth prefer to shift and work as self-employed.

The job training significantly increases the probability of employment in all of the four sectors, with more weight for the self-employment sector.Urban individuals are more likely to be private formal workers than unemployed. This can be explained by the fact that formal firms are more located in urban areas. Income effects on participation are measured by the unearned income of the individuals. The result indicate that individuals with unearned income are more likely to be private formal or public workers than unemployed. As hypothesized previously, risk aversion individuals are less likely to be self-employed. This said, the risk aversion significantly decreases by 8.7 percentage point the probability of working as self-employed but, it increases significantly the probability to participate on the labor market as wage employee. If a family member lost his job, individuals are more likely to work informal. This is more likely if the lost family member was providing income to the household, the remaining unemployed members of the family should find faster how to compensate for the revenue lost. The easiest way is then to go informal as the entry barriers are weak in this sector. As expected, having a father working as public servant significantly increases the probability of participation as employee in the public administration or working as self-employed. This effect is however, negative for participation in the private informal and nil in the private formal sector. Similarly, individuals with parents working as self-employed are more likely to work as independent, but less likely to work in the private sector. In the second step, we used the estimates from the multinomial logit regression as a selection equation to address the selection bias into the wage equations.

2.4.2 The Wage Equations

Selectivity-corrected estimates of the sectoral wage equations for employed of different sectors are given in Tables 2.4 and 2.5, respectively. All of the wage equations are statistically significant overall. From these tables, the results for the unconditional quantile regressions at the sector level show that mostly the selection terms "δ" are statistically significant in all sectors. These results indicate a presence of sample selection bias for individuals across the earning distribution in the sectors of the labor market. Tables 2.4 and 2.5 summarize the results for quantile regressions at the 10th, 50th and 90th percentiles for each sector. Focusing on the gender variable, the results highlight a significant earning gap between men and women across all four sectors of the labor market that have earnings and the gap appears to be larger in lower-paid sectors. For example, in self-employment sector, man's expected earnings at the bottom of the earning distribution is about 25.2% higher than a woman's expected earnings. This difference goes to 43.9% at the top of the earning distribution. In the private informal sector, in the middle quantiles of the distribution, a woman's expected earnings at the 50th percentile are approximately 58.8% lower than a man's expected earnings and 48.6% lower at the 90th percentile.

Sidelined by the scarcity of positions in higher-paid sectors given their characteristics, many males have joined the lower-paid sectors and have become more competitive in sectors often reserved to their female counterparts. Of course, similar discrimination is found for higher-paid sectors but with lower intensity. In general, whatever the sector, the results show more pronounced barriers for women competing for jobs. Linear and quadratic terms in age have the expected positive and negative signs respectively, in all sectors. This implies that age dividend does exist in all the four sectors with more weight in the middle of the distribution in the private informal sector. In the latter tier of the distribution, one additional year increases for 10,2% the monthly earnings of individuals. In regard to education, basic and secondary education do matter for earnings only in lower-paid sectors. Indeed, as previously said, the self-employment sector is made up of individuals with small business and street vendors where more than 90% of them are informal self-employed individuals. The two sectors, self-employment and private informal sectors are thus of lower productivity where workers are said to face low and volatile earnings in small firms with labor-intensive activities and without job security. These facts make the two sectors weakly attractive to tertiary educated individuals but more attractive to low educated individuals who barely have room in higher-paid sectors. However, the tertiary education significantly increases earnings in all the sectors. In higher-paid sectors, tertiary education increases by 68.1 percentage point the monthly earning in the public sector while it almost double (97.6 percentage point) the monthly earning of individuals in the middle tier of the private formal sector. Overall, tertiary educated people have advantage over the rest whatever the sector of employment. The higher earnings associated with age, education provides clear support to the human capital theory in the public and private sector (Becker, 1964; Mincer, 1974). However, the relative low/insignificant returns to basic education in higher-paid sectors that are known to be legal as they operate formally erodes the individual's real incentive to obtain low levels of education and questions the quality of schooling in DRC. Furthermore, higher levels of education tend to yield significant positive returns in lower-paid sectors (largely informal and less productive) as well. For a country with a relatively high level of informality, this is area calls for further attention and investigation. It is known that informality decreases with education. But, on the one side, if there is no earning premium to basic education in higher-paid (that are formal sectors) labor sectors in a country where it is so costly to study, people may choose working underground instead. On the other side, if education, both basic and tertiary levels are rewarded in the lower-paid (mostly informal sectors) as well the latter, as more flexible, with less entry and exit barriers and without or less tax burden, will be attracting more and more educated people eroding the country's productivity and wealth in the long run. The ending point could be a vicious circle of poverty where the country and its population are caught in an informality trap.

Surprisingly job training effect, thought positive, is insignificant in the private formal sector. But this can probably be explained by the presence of highly educated individuals in this sector. Table 2.1 showed that 38.53% of workers of that sector have tertiary education. This level of education might act as job training and as seen previously, higher education influence strongly the workers' earnings. The significant returns to job training in the public sector can be liked with benefit in the form of incremental salary, additional allowance, or promotion in accordance with the government policy. Concerning individual experience, its positive effect is significant only in higher-paid sectors. Finally, living in urban areas often procures higher returns to workers regardless of the sector of employment on the labor market.

2.4.3 Decomposition Results

For a better-detailed analysis and a clear and simplified view, this section presents the decomposition results in the form of graphs[8]. Figure 2.1 plots the estimated wage gap, correcting for selection for six cases representing the interaction of four sectors of the labor market taken in pairs. Taking the sector with high mean income as the reference sector in each pair, the six pairs of sectors are as follow; a) Earnings in public sector vs in self-employment, b) Earnings in private formal sector vs in self-employment, c) Earnings

[8]Estimates can be made available upon request.

in self-employment vs in private informal sector, d) Earnings in private formal sector vs in public sector, e) Earnings in public sector vs in private informal sector and f) Earnings in private informal sector vs in private informal sector. Equal weights are assigned to the two sectors for each pair. Figure 2.1 contains six subfigures and reports the earning gap of the six pairs.

Considering the subfigures A, B, E and F where a higher-paid sectors is the reference sector over a lower-paid sector, we observe that the estimated earning gap is positive throughout the earning distribution. This confirms that workers in higher-paid sectors are more remunerated than their counterpart in the lower-paid sectors. However, the earning gap is not homogeneous throughout the earning distribution. In the bottom part of the distribution it is large (this is more pronounced when the public sector is the reference sector). Around the fifth percentile, it goes decreasing to be lower in the second half of the distribution. This heterogeneity in the distribution of the earning gap indicates that not all individuals in the lower-paid sectors are the same compare to individuals in the higher-paid sectors when it comes to earnings. Focusing on each set of factors, the coefficient effect dominates the characteristic effect in the bottom of the distribution. Subsequently, much of the earning gap at the bottom of the distribution result from workers in the lower-paid sectors being remunerated less than workers of the higher-paid sectors for the remunerated characteristics. The coefficient effects decrease over the distribution, whereas the characteristic effects increase in the second half of the distribution, particularly toward the upper end of the distribution where characteristics effects dominate the coefficient effects. This fact indicates that on the one side, workers of the lower-paid sectors earn less not only because they are less skill endowed but also because they earn lower returns on such skills. On the other side, individuals at the top end of the distribution in lower-paid sectors earn less compare to their counterpart of the higher-paid sectors because the latter have superior skills.

This position of workers in the lower-paid sectors vis-a-vis of workers in the higher-paid sectors shows two different types of lower-paid workers in the DRC's labor market. First, those at the bottom of the distribution, who, despite *identical* (or almost) characteristics to their counterparts' workers at the bottom of the distribution in the higher-paid sectors, earn less on such characteristics. They represent the disadvantaged group, working in the "easy-entry" segment of the lower-paid sectors. Second, workers of lower-paid sectors at the upper tiers of the distribution, where the characteristics effect explains much of the earning gap. In this upper tier, we observe a relative lower earning gap between workers of the higher-paid and lower-paid sectors. Both types of workers could have no significant differences in the rates of return on characteristics but the difference in earnings is due to differences in skills. This group of workers may be associated with the advantaged segment of lower-paid sectors and might be linked with entrepreneurship. Largely informal, these workers endowed with some particular skills prefer the combination of monetary rewards and greater flexibility (in terms of working hours, work relationships, responsibilities, etc.) in the informal sector (Fields, 1990). In this advantaged tier of the distribution, lower-paid individuals might be enjoying non-earning features and maximize more their utility rather than their earnings. More interestingly, in the top end of the distribution, the coefficient effect is negative in the four subfigures. This informs that, in this tier of the distribution, moving individuals of the lower-paid sector to higher-paid sectors and rewarding them according to their characteristics could even lower their actual earnings. We then have a "*lower-paid sector employment earning premium*" on the labor market. Due to the prevalence of informal individuals in the lower paid sector as previously showed, the lower-paid sector employment earning premium is indeed an "*informal employment earning premium*" for individuals who, given their characteristics, wouldn't do better in the higher-paid (formal) sector. In this context, the significant rationing of higher-paid jobs and relative abundance of informal workers, particularly those with very low qualification levels, undermine the benefits of higher-paid sectors for low-skilled individuals.

In the Subfigure D of Figure 2.1 , the pair of sector concerns the two higher-paid sectors of the labor market where private formal is the reference sector. The figure shows a positive earning throughout the earning distribution implying that Private formal workers earn more than Public workers. Both for this pair and the previous ones, the earning gap is not homogeneous along the distribution. However, differently from the previous case, the earning gap is lower at the bottom of the distribution and higher in the large part of the second half of the distribution. As public jobs are financed by means other than those operating in the private sector and that, earnings are mainly determined through a political process or service regulations (Gunderson, 1979), earning disparities between lower and higher earners within the public sector are less important. However, in the private formal sector, productivity matters for earnings. This fact generally leads to a large earning disparity between lower and higher productive workers (earners) within the private formal sector. This difference in the determinants of earnings between the two higher-paid sectors leads to an increasing earning gap between them throughout the distribution. As regard to each set of factors, the coefficient effect exceeds the characteristics effect for 80% of the earning distribution. From this, we learn that the earning difference between the two higher-paid sectors in DRC is mainly due to the difference in skill remuneration rather than skill composition with a higher return in the Private formal sector. Finally, the pair of lower-paid segment is given in the Subfigure D of Figure 2.1. Due to the relative lower earning gap, we conducted, in addition to the result of the linear regression in Table 2.2

that showed a significant difference between the two lower-paid sectors, a mean test difference of the earnings between them and a likelihood-ratio test for combining alternatives sectors in Table 2.7. Overall, the result showed that their mean earnings are significantly different and that the two lower-paid sectors are significantly distinguishable with respect to the variables in the model. Thus, no categories should be combined. Given that, we did separate the two sectors in this decomposition section. Between the two lower-paid sectors, a relative earning gap, mainly explained by the skill remuneration, exists in the first half of the distribution but disappears in the second half. This positive gap can be due to the presence of formal entrepreneurs in the Self-employment sectors whose skill remuneration in the formal sector makes the difference. However, as one moves up along the earning distribution, the earnings in Self-employment is caught by that of Private informal sector. The vanishing earning gap can be partially explained by the fact that formal entrepreneurs who are the higher earners of the self-employment sector are paying tax on their higher earnings meanwhile higher earners of the Private informal sector are not. In any cases, the vanishing gap is consistent with greater freedom of choice between Self-employment and Informal job as individuals move up along the distribution. With a marginal earning gap in this tier of the distribution, workers in the Private informal sector may to some extent be willing to accept lower earnings to avoid the administrative cost of social security in regulated sectors, when it is perceived as costly and ineffective Garcıa (2017).

2.5 Conclusion

In this paper, we take advantage of a large-scale survey in DRC to examine the heterogeneity in the labor market and examine the earning gap and its decomposition across different sectors throughout the earning distribution. Using a Mincerian selectivity-corrected sectoral wage equations estimated for each sector, the role of workers' personal characteristics is explored in the first step. In the second step, an Oaxaca-Blinder decompositions of the wage differentials between sectors are carried out. Initial data description reports five different sectors in the DRC labor market. A basic earning regression with four sectors entered as dummies shows clearly that they are statistically different when earnings are concerned. The analysis of sector earnings combined with the stylized facts of DRC's labor market allows to group the four sectors into two mains group. On one side two higher-paid sectors, fully made up of formal individuals, where the average monthly earning is above the national average and far above the updated national minimum wage. On the other side two lower-paid sectors, mainly made up of informal individuals, where the average monthly earning is below the national average and far below the updated national minimum wage.

The earning functions show individuals self-select in different sectors of the labor market. The higher earnings associated with age, education and others provides a clear support to the human capital theory. The result shows that basic education is not significantly rewarded in higher-paid sectors, meanwhile tertiary education matter for earnings in lower-paid sectors as well. The decomposition results report that, when higher-paid sectors are taken as reference sectors, the earning gap is positive but not homogeneous throughout the earning distribution. Furthermore, this positive earning gap is due both to skill remuneration in the bottom part of the distribution and skill composition in the upper tier of the distribution implying that, workers of the lower-paid sectors earn less not only because they are less skill endowed but also because they earn lower returns on such skills. More interestingly, when the pair of sector concerns higher-paid vs lower-paid sectors, the coefficient effect is negative in the upper tier of the distribution, highlighting an informal employment earning premium on the labor market as lower-paid sectors are mainly informal. When the two higher-paid sectors are concerned, private formal workers earn more than public workers. However, differently from the previous case, the earning gap, mainly attributable to skill remuneration, is lower at the bottom of the distribution, but higher in the large part of the second half of the distribution. When the two lower-paid are concerned, the earning gap appears only in the first half of the distribution and vanishes in the second half, undermining the benefits of Self-employment.

Our results have policy implications for development strategies, in DRC and the developing world that aim, whether to combat earning discrimination on the labor market for well-being purpose, whether to safeguard diversity of sectors in the labor market with specific regulations, for entrepreneurial spirit and household subsistence strategies purposes. The coexistence of diverse sectors on the labor market with differences in earning outcomes calls for various strategies to face the large share of lower-paid sector and to limit informality that, in some circumstances, provide a premium to workers but increases discrimination on the labor market. Specific policies have to be constructed for each particular group or sector on the labor market. This is important because, on the one side we show that different mechanisms may be working at the group of sector level or even at each sector level and on the other side. because most people, especially the poor, derive all or the great bulk of their income from the work they do.

2.6 Appendix

2.6.1 Descriptive statistics

Table 2.1: Descriptive statistics

Variables	Total	Unemployed	Self-employed	Public workers	Private Formal workers	Private Informal workers
Size	55.652	21043	22331	3289	1069	7920
Monthly earning (CDF)	51946.8	-	42954.26	94901.43	178185.8	41451.6
Male (Proportion)	48.40	44.94	51.96	78.47	78.92	32.50
Age(year)	32.82	25.89	37.70	41.95	38.25	32.58
Education(year)	6.76	7.68	5.62	11.87	12.41	4.92
Education (Proportion)						
-Illiterate	21.66	14.05	27.94	5.02	2.89	34.63
-Primary school	24.61	21.31	30.06	4.72	4.01	29.96
-Secondary school	47.59	57.47	40.22	63.52	54.57	33.03
-Tertiary school	6.14	7.17	1.78	26.73	38.53	2.38
Job training (Proportion)	0.05	0.00	3.17	51.26	41.14	5.67
Experience (year)	8.18	1.28	13.35	11.83	7.99	10.43
Living in urban (Proportion)	50.54	66.89	35.97	66.52	93.17	35.80
Marital (Proportion)	58.98	29.57	78.59	82.82	72.22	70.11
Head of hh (Proportion)	34.85	9.82	55.21	77.96	72.40	20.97
Others(proportion)						
Unearned income (Proportion)	1.45	1.90	0.79	3.05	6.06	0.93
Father in the public sector(Proportion)	18.84	21.93	14.61	35.94	36.95	13.02
Father is Self-employed (Proportion)	40.42	33.28	48.63	26.79	16.09	45.16
Risk averse (Proportion)	4.79	6.90	1.91	11.21	18.38	3.13
Family member lost a job (Proportion)	2.55	3.18	1.93	2.09	3.54	2.73

Notes: Authors computation using 1.2.3 data

2.6.2 Linear regression

Table 2.2: Linear regression

Variables	(1)	(2)	(3)	(4)
Constant	10.06***	10.97***	11.63***	9.97***
	(0.007)	(0.018)	(0.032)	(0.012)
Self employment	-	-.909***	-1.57***	.088***
	-	(0.019)	(0.033)	(0.014)
Public sector	.909***	-	-.667***	.997***
	(0.019)	-	(0.037)	(0.022)
Private-Formal sector	1.57***	.667***	-	1.665***
	(0.033)	(0.037)	-	(0.034)
Private-Informal sector	-.088***	-.997***	-1.665***	-
	(0.014)	(0.022)	(0.034)	-
Observation	33241			
$F(3, 33237)$	1459.92			
Prob $\succ$ F	0.0000			
Adj R-squared	0.1164			

Notes: Authors computation. Notes: Log monthly earnings is the dependent variable. ***,**,* Denote significance at 1%, 5% and 10% levels, respectively. Standard errors are in parenthesis. In regression (1), (2), (3) and (4), Self-employment, Public, Private Formal, and Private Informal sectors are respectively taken as the base sector.

2.6.3 Determinants of sectors' choice

Table 2.3: Determinants of sectors' choice

	Self-employed		Public workers		Private formal workers		Private informal workers	
	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.
Male	0.0753***	(0.0036)	0.0010***	(0.0017)	0.0081***	(0.0013)	-0.0825***	(0.0030)
Age	0.0232***	(0.0009)	0.0010***	(0.0004)	0.0012***	(0.0003)	-0.0054***	(0.0007)
Age_sq	-0.0003***	(0.0000)	0.0000***	(0.0000)	0.0000***	(0.0000)	0.0000***	(0.0000)
Education								
Primary	0.0224***	(0.0051)	-0.0100***	(0.0027)	0.0014	(0.0016)	-0.0184***	(0.0043)
Secondary	-0.0157***	(0.0049)	0.0286***	(0.0024)	0.0130***	(0.0014)	-0.0557***	(0.0042)
Tertiary	-0.1324***	(0.0098)	0.0653***	(0.0034)	0.0287***	(0.0024)	-0.0656***	(0.0087)
Experience	0.0135***	(0.0002)	0.0004***	(0.0001)	-0.0003***	(0.0001)	0.0058***	(0.0002)
Job training	0.3137***	(0.0645)	0.0343***	(0.0014)	0.0140***	(0.0015)	0.2114***	(0.0277)
Living in urban	-0.0258***	(0.0039)	-0.0281***	(0.0019)	0.0212***	(0.0021)	-0.0563***	(0.0034)
Marital	0.0741***	(0.0040)	0.0082***	(0.0019)	-0.0006	(0.0013)	0.0150***	(0.0034)
Unearned income	0.0187	(0.0177)	0.0035***	(0.0047)	0.0054**	(0.0026)	-0.0244	(0.0160)
Risk aversion	-0.0874***	(0.0104)	0.0102***	(0.0026)	0.0079***	(0.0015)	0.0177**	(0.0084)
Member of family lost his job	-0.0466***	(0.0111)	-0.0140***	(0.0045)	-0.0025	(0.0029)	0.0330***	(0.0087)
Father is a public servant	0.0227***	(0.0051)	0.0099***	(0.0017)	-0.0008	(0.0012)	-0.0258***	(0.0045)
Father is an Self-employed	0.0534***	(0.0038)	0.0004***	(0.0019)	-0.0052***	(0.0015)	-0.0073**	(0.0031)
Province fix effect	Yes		Yes		Yes		Yes	
Number of observation	57 204							
Prob ≻ F	62692.57***							

Notes: Notes: Author's calculations based on 1.2.3 data. The reference sector is the unemployment sector. ***,**,* Denote significance at 1%, 5% and 10% levels, respectively. Standard errors are in parenthesis. Illiterate is the excluded categories in education. All models contain Province dummies.

2.6.4 Quantile regressions

Table 2.4: Quantile regressions : Lower-paid sectors

	Self-employment			Private Informal workers		
	10	50	90	10	50	90
Male	0.252***	0.410***	0.439***	0.227***	0.588***	0.486***
	(0.038)	(0.023)	(0.027)	(0.047)	(0.053)	(0.056)
Age	0.071***	0.065***	0.054***	0.073***	0.102***	0.067***
	(0.010)	(0.006)	(0.007)	(0.010)	(0.010)	(0.009)
Age_sq	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Education						
Primary	0.102**	0.091***	-0.002	0.075	0.214***	0.118***
	(0.046)	(0.027)	(0.026)	(0.052)	(0.054)	(0.039)
Secondary	0.288***	0.303***	0.217***	0.083	0.461***	0.470***
	(0.045)	(0.027)	(0.029)	(0.057)	(0.059)	(0.053)
Tertiary	0.547***	0.587***	1.210***	0.194**	0.976***	1.423***
	(0.092)	(0.071)	(0.144)	(0.076)	(0.109)	(0.224)
Experience	-0.001	0.001	0.001	-0.005*	-0.012***	-0.004
	(0.002)	(0.001)	(0.001)	(0.003)	(0.003)	(0.003)
Job training	0.226***	0.238***	0.354***	-0.064	0.274***	0.566***
	(0.075)	(0.055)	(0.089)	(0.062)	(0.074)	(0.126)
Formal	-0.222***	-0.348***	-1.186***	-	-	-
	(0.055)	(0.126)	(0.333)	-	-	-
Without Employee	0.003	-0.075***	-0.104***	-	-	-
	(0.047)	(0.028)	(0.034)	-	-	-
Urban	0.188***	0.270***	0.425***	0.185***	0.458***	0.485***
	(0.037)	(0.024)	(0.029)	(0.047)	(0.054)	(0.052)
δ	0.052***	0.015**	-0.008	0.021***	0.059***	0.048***
	(0.011)	(0.006)	(0.009)	(0.008)	(0.010)	(0.012)
Constant	8.317***	9.421***	12.713***	7.589***	8.196***	10.607***
	(0.268)	(0.291)	(0.697)	(0.188)	(0.190)	(0.206)
Province fe	Yes	Yes	Yes	Yes	Yes	Yes
Prob $\succ$ F	33.51***	91.65***	64.61***	19.16***	115.04***	41.03***
Number of obs		21732			7216	

Note: ***,**,* Denote significance at 1%, 5% and 10% levels, respectively. Standard errors are in parenthesis

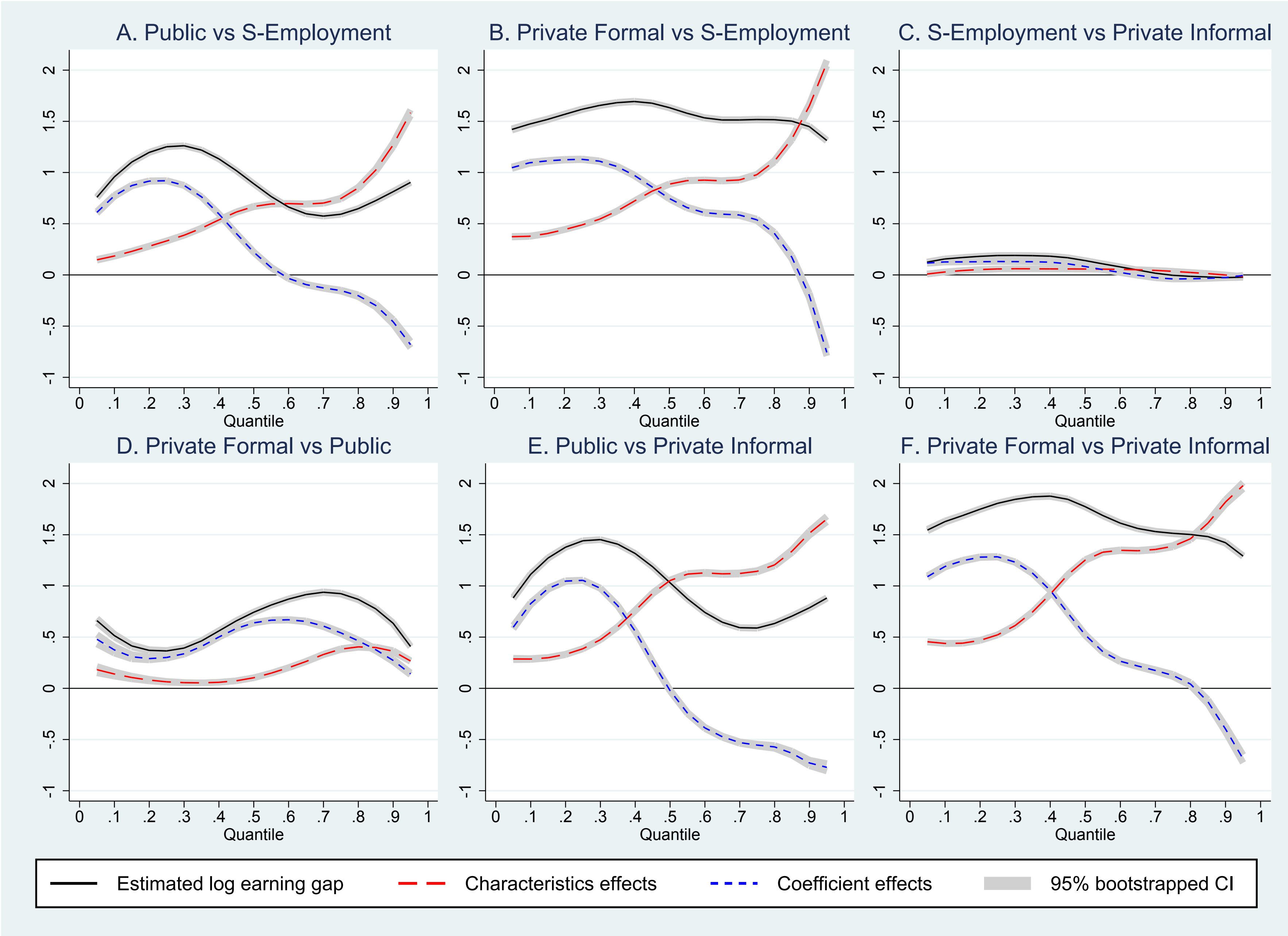
Table 2.5: Quantile regressions : Higher-paid sectors

	Public workers			Private Formal workers		
	10	50	90	10	50	90
Male	0.032	0.032***	0.218***	0.077	0.163**	0.135*
	(0.073)	(0.013)	(0.063)	(0.122)	(0.082)	(0.074)
Age	0.021	0.014***	0.040***	0.064*	0.059***	0.024
	(0.022)	(0.003)	(0.015)	(0.037)	(0.020)	(0.019)
Age_sq	0.000	0.000***	0.000**	-0.001	-0.001**	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Education						
Primary	-0.483**	-0.156***	-0.144	0.827	-0.252	-0.050
	(0.240)	(0.035)	(0.127)	(0.520)	(0.232)	(0.131)
Secondary	-0.126	-0.061**	0.040	1.001**	-0.328*	0.006
	(0.163)	(0.026)	(0.106)	(0.441)	(0.189)	(0.098)
Tertiary	-0.069	0.057**	0.681***	0.997**	0.103	0.422***
	(0.177)	(0.029)	(0.135)	(0.451)	(0.202)	(0.120)
Experience	0.022***	0.003***	0.005	-0.010	0.006*	0.007**
	(0.004)	(0.001)	(0.004)	(0.008)	(0.004)	(0.004)
Job training	-0.154**	0.036***	0.179***	0.012	-0.009	0.024
	(0.078)	(0.012)	(0.062)	(0.103)	(0.071)	(0.085)
Urban	0.976***	0.098***	0.423***	1.410***	0.476***	-0.059
	(0.087)	(0.013)	(0.045)	(0.352)	(0.113)	(0.077
δ	0.050**	-0.001***	0.015**	0.014	0.028***	0.028***
	(0.020)	(0.002)	(0.007)	(0.014)	(0.010)	(0.007)
Constant	9.359***	10.648***	10.963***	7.064***	10.281***	12.030***
	(0.569)	(0.083)	(0.348)	(0.958)	(0.526)	(0.436)
Province fe	Yes	Yes	Yes	Yes	Yes	Yes
Prob $\succ$ F	14.82***	61.03***	20.56***	6.08***	36.34***	6.81***
Number of obs		3248			1043	

Note: ***,**,* Denote significance at 1%, 5% and 10% levels, respectively. Standard errors are in parenthesis

2.6.5 Earning decomposition

Figure 2.1: Earning differentials between formal and informal sector over different quantiles of the earning distribution by province



2.6.6 Testing the validity of the exclusion restrictions

Table 2.6: Testing the validity of the exclusion restrictions

Sectors	Relevance	Excludability		
		10	50	90
	Wald Test ($P-value$)	Wald Test ($P-value$)	Wald Test ($P-value$)	Wald Test ($P-value$)
Unemployed	364.15 (0.0000)	– –	– –	– –
Self-Employed	364.15 (0.0000)	2.64 (0.0216)	0.90 (0.4812)	1.00 (0.4180)
Public	73.89 (0.0000)	1.22 (0.2979)	1.60 (0.1584)	0.20 (0.9632)
Private_Formal	23.23 (0.0003)	0.63 (0.6420)	0.79 (0.5595)	0.40 (0.8046)
Private_Informal	121.66 (0.0000)	0.98 (0.4320)	2.07 (0.0674)	0.27 (0.9283)

Note: Relevance: *statistical significance of the exclusion restrictions in the five sector choice equations.*
Excludability: *statistical insignificance of the exclusion restrictions in each earning equation.*

Here, we seek to provide evidence of the validity of the elicited exclusion restrictions of variables incorporated in the multinomial model for identification. For them to be valid, the variables should be relevant determinants of sector choices but not directly related to earnings, once we have conditioned for the employment sector and other attributes. The validity of the "relevance" condition can be directly tested from the estimates of the multinomial selection equation. Table 2.6 above contains several Wald tests for the joint statistical significance of the exclusion restrictions for each estimated model. As can be seen, the relevance of the exclusion restrictions for the whole multinomial model is clearly not rejected by the data. Taking each equation separately, the variables included are good predictors of the differences in the likelihood of each sector (in relation to be unemployed), As for the "excludability" condition, no formal overidentification test has yet been developed in this framework. Therefore, this condition has to be informally checked by examining the joint statistical significance of the exclusion restrictions in the outcome equation(s) of sectors, conditional on other determinants of earnings. The results of these Wald tests — which are reported in the four last column of Table 2.6 - suggest that the exclusion restrictions are not jointly significant in the outcome equations at any conventional significance level, with the exception of Self-employed sector on the tenth percentile and private informal sector on the fifteenth percentile, in which the null hypothesis that the exclusion restriction's coefficients are jointly equal to zero is not rejected when considering a significance level of 5%. Overall, the evidence obtained when adopting this informal approach to demonstrating the validity of the exclusion restrictions suggests that the model is well identified.

2.6.7 Test of significance

Table 2.7: Test of significance

Mean difference test			
Groups	Observation	Mean	P-value
Self-Employment	22331	10.061	
Private informal	7920	9.97	0.0000
Difference		0.088	

Based on the above, There is a significant mean difference of log earnings between the two sectors

Likelihood-Ratio test for combining alternatives

	chi2	df	P≻ chi2
Unemployed & Self-Employment	36224.949	26	0.000
Unemployed & Public	27430.368	26	0.000
Unemployed & Private Formal	9773.842	26	0.000
Unemployed & Private Informal	22323.603	26	0.000
Self-Employment & Public	13218.431	26	0.000
Self-Employment & Private Formal	4590.221	26	0.000
Self-Employment & Private Informal	3720.946	26	0.000
Public & Private Formal	1323.106	26	0.000
Public & Private Informal	10228.398	26	0.000
Private Formal & Private Informal	3675.629	26	0.000

Based on the above, No categories should be combined

Chapter 3

Spatial Inequality, Labor Market Frictions and Informality in the Democratic Republic of the Congo[1]

[1]This chapter is coauthored with Zainab Iftikhar and Frédéric Docquier. The authors thank Célestin Bucekuderhwa Bashige, Muriel Dejemeppe, Georg Duernecker, Leo Kaas, William Parienté, Michel Tenikue, and Bruno Van der Linden for their helpful comments. I am grateful to UCLouvain for financial support through the "Fond de coopération au dévelopment". Zainab Iftikhar thanks the Joachim Herz Foundation for financial support.

3.1 Introduction

Despite some timid progress toward the Sustainable Development Goals (SDGs), Africa continues to lag behind when it comes to per capita income growth and economic convergence. More than half of the global poor (i.e., people earning less than USD 1.90 PPP per day) live in Africa, and the region of Central Africa is struggling to improve its SDG indicators. African nations have experimented with diverse approaches to economic development, prioritizing infrastructure and capital accumulation, fight against corruption and poor governance, trade and economic integration, human development, etc. Most of these actions have failed, possibly because achieving sustainable development is a multifaceted challenge requiring a combination of favorable conditions that are difficult to reconcile. In particular, abundance of mineral wealth has not generated export-led growth, and has even become a curse rather than a blessing as appropriate conditions for generating value added are not met. The o-ring model of development is likely to apply to many African countries. Although the original theory focuses more on micro aspects (Kremer, 1993),[2] a macroeconomic generalization of the o-ring theory implies that policies targeting a specific part of the economy have low value if other complementary components of the development machinery are not working properly. It is also in line with theories showing that the extreme poor need a "big multidimensional push" to escape from the poverty trap (Banerjee et al., 2015). Our findings are in line with this view, and illustrate the difficulty of reducing income inequalities in the Democratic Republic of Congo (DRC, henceforth) using a two-sector model with labor market frictions.

Our focus on DRC is justified by two facts. First, DRC is among the world's poorest countries, and suffers from weak levels of public infrastructure and human capital, poor governance, and an overwhelmingly large informal sector. Despite greater monthly wage levels in the formal sector, one third of the skilled and ninety percent of the unskilled are employed informally. These numbers potentially reflect a labor market with large frictions. Second, microdata are available to characterize both levels of spatial and within-region inequalities, and the source of income earned by each individual. In particular, the *1.2.3* household survey is quite unique in the sense that it documents the level of income of Congolese working age people by education level, by sector of activity (formal vs. informal), by type of activity (entrepreneurship vs. employment) and by province. The database provides a unique opportunity to analyze interactions between sectors and the functioning of the labor market. It reveals that DRC provinces are strongly heterogeneous in labor market characteristics and productivity levels. Kinshasa is the richest province with a level of income per worker that is 2.2 times greater than the country-wide average level, and 3.4 times greater than the level observed in one of the poorest provinces, namely

[2]Say, the importance of firm-level interactions between tasks

Kasai Occidental. Spatial inequalities are important, and account for approximately fifty percent of the Theil index between broad groups of income earners. Within-province inequality are large as well. In all provinces, the informal sector is not only huge but drains a large share of both skilled and unskilled workers. On average, workers in informality have low levels of earnings compared to similar workers in the formal sector.

We build a two-sector model that exactly matches the distribution of income in DRC, and use it to quantify the effect of different types of "policies". In each province, a single good is produced by formal and informal firms, using different technologies. The formal sector is characterized by labor market frictions. The informal sector is characterized by small businesses run by educated entrepreneurs who hire informal workers on a competitive labor market. In line with the reality, we ignore trade and migration flows; they are almost inexistent between provinces.[3] The economy of each province is characterized by five sets of parameters, namely (i) socio-demographic indicators (say, the share of secondary-educated workers), (ii) the level of public infrastructure, (iii) technological parameters in the formal sector, (iv) technological parameters in the informal sector, and (v) labor market frictions in the formal economy (partly reflecting poor labor market institutions and arbitrary regulations).

We parameterize the model to exactly match income disparities between groups, and conduct a set of numerical experiments. The calibration reveals large differences across provinces, both in observed characteristics (i.e., the education structure of the population and the level of public infrastructure per worker) and in identified parameters (i.e., total factor productivity scale factors in both sectors, labor market frictions). These differences explain the average income gap with the richest province of Kinshasa as well as within-province inequality. We use counterfactual simulations to analyze the role of these province-specific characteristics in explaining the income gap with the richest province of Kinshasa as well as within-province inequality. Somewhat unsurprisingly, we show that income disparities are mostly determined by technological characteristics, reflecting both endowment in mineral resources, geographic position, topography, institutional quality, etc. Most DRC provinces are characterized by chronic political instability and corruption, and provinces with rich subsoil have experienced violent conflicts for decades. More interestingly, we find that stimulating TFP in the formal sector increases the province-wide average level of income, but has smaller effect on low-skilled workers' income. Stimulating TFP in informality induces smaller aggregate gains, but greater benefits for the unskilled.

[3]Internal exchanges are hampered by insufficient and poor conditions of the road network. Chronic economic mismanagement and internal conflicts have led to serious under-investment in infrastructure over many years. To put it in perspective, only 5.4% of the road network over the country is asphalted. This also significantly limits the mobility of labor. Indeed, Using the 1.2.3 database, the proportion of migrants accounts for 4.2% of the working aged population only. These includes both intra- and inter-province labor mobility. The inter-province mobility itself is likely to be significantly lower.

Acting on both sectors is desirable to increase the average income of both types of workers.

A development policy that disregards the situation of the informal sector has low or even detrimental effects on inequality and extreme poverty. In particular, policies targeting education and public infrastructure have smaller effects as they mostly impact productivity in the formal sector, and reduces the skill ratio and productivity in informality, where many unskilled workers are trapped. These policies taken in isolation induce potential undesirable effects on the distribution of income, inequality and extreme poverty. More generally, the effectiveness of each policy taken in isolation is limited, due to complementarities between them and to the low mobility of unskilled workers across sectors. We quantify the high level of complementarity between policies and highlight strong o-ring patterns of spatial inequality. With regard to the low mobility across sectors, it is due to large labor market frictions. Reducing frictions at the levels observed in Kinshasa alone or combined with a policy targeting productivity in the formal sector can be worse than targeting the technology only as it mainly benefits skilled workers. A much more dramatic reduction of labor market frictions would be needed to benefit unskilled workers.

Our paper relates to the literature on the effectiveness of development policies. In the discussion of William Estearly's "Elusive Quest for Growth," Wacziarg (2002) wrote: "Over the last decades, the list of proposed panaceas for growth in per-capita income included high-rate of physical capital investments, rapid human capital accumulation, low-income inequality, low fertility, being located far from the equator, a low incidence of tropical diseases, access to the sea, favorable weather patterns, hands-off government, trade-policy openness, capital-markets development, political freedom, economic freedom, ethnic homogeneity, British colonial origins, a common-law legal system, the protection of property rights and the rule of law, good governance, political stability, infrastructure, market-determined prices (including exchange rates), foreign direct investment, and suitably conditioned foreign aid." Most of these miracle growth policies have proven disastrous or ineffective, which might be due to the complex system nature of development policies (Mueller, 2020). Growth miracles require a combination of favorable and mutually reinforcing factors. In particular, creating incentives that are conducive to growth-enhancing behaviors (such as private investments in physical and human capital) requires a supportive environment with political stability and sound governance, a good level of public infrastructure, and functioning markets.

Our approach formalizes the interactions between these ingredients. First, it accounts for large frictions observed in the labor market. In many sub-Saharan African countries, a small number of firms operate in the (productive) formal sector and offer relatively high wages. However, the disfunctioning labor market implies the existence of overwhelming informal sector characterized by lower earnings and lower productivity (Dickens, 1985;

Jütting and De Laiglesia, 2009; Maloney, 1999, 2004). The informal sector is defined as the part of an economy that is not taxed or monitored by any form of government. It drains a large proportion of both skilled and unskilled workers in developing countries (Docquier and Iftikhar, 2019; Jütting and De Laiglesia, 2009; Schneider, 2012). In DRC, most people who have created informal enterprises have done so because they lack alternative employment opportunities such that informality is now the norm for many workers. The odds of finding productive employment on the labor market are disappointingly low for the workforce. Reyes et al. (2017) show that in the country, workers find it increasingly difficult to participate in the formal labor market, and education does not shield workers from informality. Job creation was not fast enough to meet the demand from a growing working-age population between 2005 and 2012.

Second, we account for the role of infrastructure.[4] There is evidence that public infrastructure – transport infrastructure, electricity, sanitation, water and sewer lines, communication systems, etc. – is a key determinant of productivity, growth and income (Calderon and Serven, 2010, 2014; Dufflo and Pande, 2007; Irmen and Kuehnel, 2009; La Porta and Shleifer, 2014b; Wang and Wu, 2015). Yet, a deficit in public infrastructure is observed in many developing countries (Bhattacharya and Kharas, 2011; Estache, 2010; Fay et al., 2011; Ingram and Kessides, 1994). This is particularly the case in DRC, where equipment are outdated, maintenance levels are insufficient, and new investments are low.[5]

Third, human capital is usually seen as a key determinant of development potential through its effects on health, knowledge, skills, resilience of people (Bloom et al., 2004; Hanuschek, 2013). On this basis, the World Bank has launched a *Human Capital Plan* for Africa in 2019 with precise targets for 2023. Again, the lack of schooling infrastructure, reflected in the low number of schools and teachers per capita, is a major obstacle. In DRC, the education system is primarily financed by parents, the school enrolment rate is low, and illiteracy is high among the population (Gyimah-Brempong, 2011). Enrolments have decreased due to isolation of some areas, parents' increasing inability to pay school fees, failure to maintain infrastructure, etc. The national program calls for universal primary education by 2018, but this objective seems unattainable. Furthermore, public spending in education is unevenly redirected towards less developed regions (?).[6]

[4]Not that informality can generate vicious circles as the difficulty to raise fiscal revenues is one of the main reasons why investments in infrastructure are small in developing countries. Besley and Persson (2014) report that low-income countries have small tax rates (varying between 10 and 20% of GDP), while in developed countries the average level of taxation is close to the 40% of GDP; they argue that this discrepancy is due by the larger size of the informal sector which makes the sensitivity of taxable income to the tax rate much greater than in developing countries.

[5]As shown by Herderschee et al. (2012), only four out of 10 province capitals are linked by road to the national capital (Kinshasa), and shipping costs by rail transport are substantially greater than in other countries. Access to electricity and telecommunication are low.

[6]In the DRC, the poorest provinces are not the prime recipient of public resources. For instance, in

Finally, and perhaps above all, it has long been recognized that good institutions and stability are needed to enable growth miracles. Good institutions provide secure property rights and relatively equal access to resources (Acemoglu et al., 2005). They emerge when there are accountability constraints on power-holders, and few rents to be captured by the elite. In Africa, expected revenues from mining extraction has led to conflict, corruption and autocracy. Natural resources abundance is usually a political and social curse (Bohn and Deacon, 2000; Tsui, 2014). DRC is rich in mineral resources. Rather than bringing economic prosperity and political stability, natural endowments have led to illegal exploitation, both internal and external, conflicts and corruption (Asiedu and Lien, 2011; Bhattacharyya and Hodler, 2010; Olson and Congdon Fors, 2004). This translates into lower levels of total factor productivity and incentives to invest for formal and informal firms.

The rest of this paper is organized as follows. Section 3.2 describes the *1.2.3* and highlights ten stylized facts characterizing the economy of DRC, and helping us motivate the specification and parameterization of our model. In Section 3.3, we develop a labor market model which is consistent with the stylized facts and anecdotal evidence. Results of our numerical experiments are presented in Section 3.4. Finally, Section 3.5 concludes.

3.2 Stylized Facts

Our study focuses on the Democratic Republic of Congo (DRC), which is the most populous francophone nation of the world (about 100 million inhabitants) and among the world's poorest countries. In the next section, we model the labor market of each Province of DRC. Our model endogenizes the size and structure of the formal and informal segments of the labor market as well as their implications for the distribution of income. We parameterize our model to match ten important stylized facts that are illustrated in the three tables and in Figure 3.1 below. The microdata used to characterize the economy of each Province are mostly extracted from the *1.2.3* database, a broad household and expenditure survey conducted between 2005 and 2012 by the DRC's National Institute of Statistics (hereafter INS) in partnership with different actors (including Afristat and the World Bank).

The *1.2.3* data were collected in three phases. Each of these numbers refers to a collection phase. Phase 1 provides detailed information on employment, unemployment, household and individual socio-demographic characteristics. Phase 2 focuses on the informal sector and gathers information on the characteristics of firms and firm owners in informality. Phase 3 is a survey on household expenditures. We exploit Phases 1 and

2013, the government spent \$4 per capita in education in Kasai Oriental, the poorest province in the DRC, against \$57 for Kinshasa, the province with the lowest number of poor.

2 of the 2012 survey, which covers about 88,600 individuals. Given its large scale and recognized quality, this database provides a unique opportunity to study the relationships between informality, productivity and income distribution.

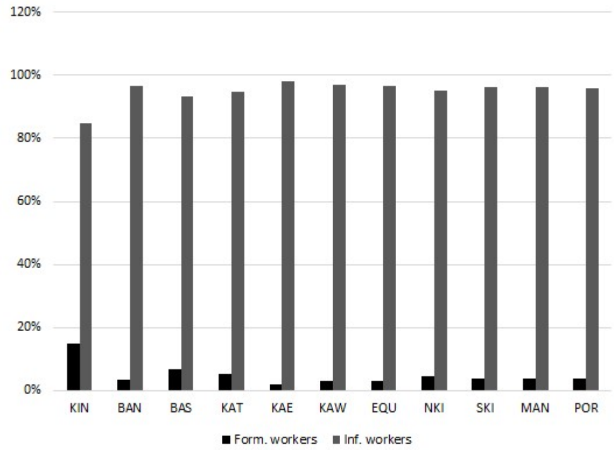
The ten stylized facts (labeled as SF1 to SF10) are the following:

SF1. DRC is one of the least developed countries of the world. – Despite an abundance of natural resources, DRC has one of the world's lowest GDP per capita and is characterized by low institutional quality, as illustrated in Table 3.1. For many decades, the majority of the population has lived in extreme poverty. In 2019 (Human Development Report of the United Nations), DRC showed a HDI index of 0.459 and ranked 179 out of 189 countries included in the database. Almost 64% of the Congolese population live below the national poverty line, and 76% live with less than USD 1.90 per day (in PPP value). Poverty rates range from 36.8 percent of the population in the richest province of Kinshasa to more than 70 percent in the least developed provinces. Inequality is among the highest in sub-Saharan Africa with Gini index of 0.45 in 2012 and a per capita GDP level of USD 767.4 (in PPP value), i.e. 13.5% of the sub-Saharan African mean, 9.4% of the Central African mean... and 3.6% of the worldwide average level.

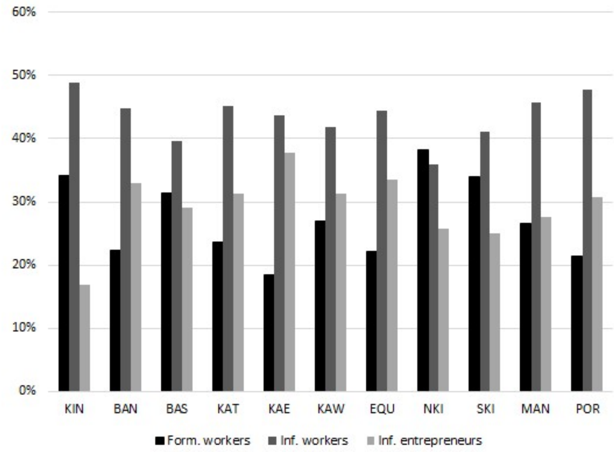
The Corruption Perception Index (CPI), which ranges from 0 to 100, scores and ranks countries based on how corrupt is the public sector. In 2019, the CPI score of DRC was equal to 18, one-half the average level of sub-Saharan African countries and 41.8% of the world average. This ranks DRC in the 168th position out of 180 countries included in the database. It is worth emphasizing that DRC has lowered its CPI rank for the recent years, which suggests that corruption is continuously increasing. Although the DRC constitution protects the ownership of private property, enforcement is virtually nonexistent. As a consequence, the Heritage Foundation assigns a property right score of 30.1 (on a scale from zero to 100) which is way below the mean level observed in Central Africa and sub-Saharan Africa. In addition, the Heritage Foundation assigns a government integrity score of 13 (on a scale from zero to 100), which is the lowest score in central Africa and less than one-half the sub-Saharan African average level. Although the tax burden is high in the country, public revenues are mostly fed by taxes on mineral extraction industries. Many of these payments fail to reach the government budget for a variety of reasons. Finally, the inflation rate is four times as large as the sub-Saharan African average (and 6 times as large as the world average). This reflects a high level of political and economic turmoil.

Hence, DRC suffers large social and economic disparities which are usually perceived as resulting from a strongly embedded culture of corruption and the lack of appropriate economic reform. Between 2007 and 2012, the country experienced an increasing economic

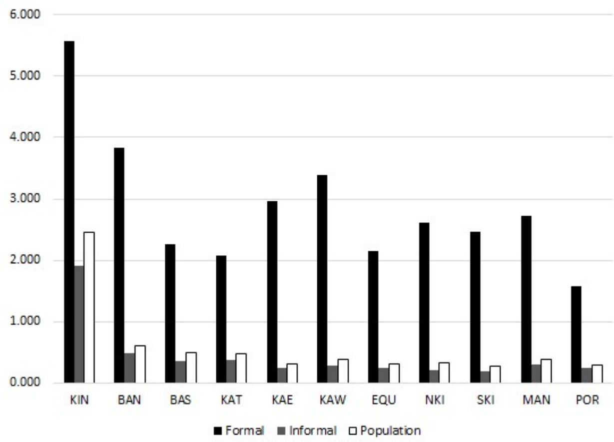
Figure 3.1: Labor market characteristics by sector and by skill group



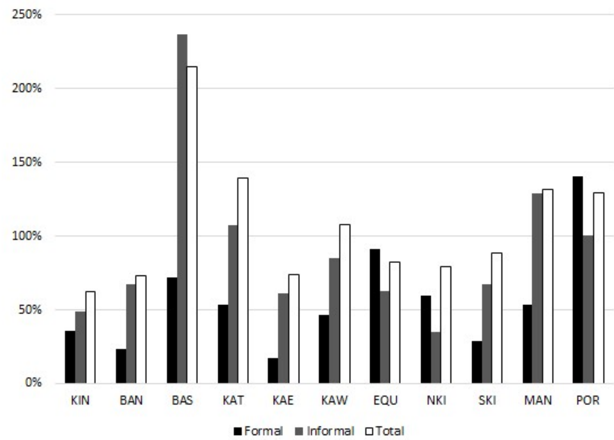
(a) Labor allocation of the unskilled



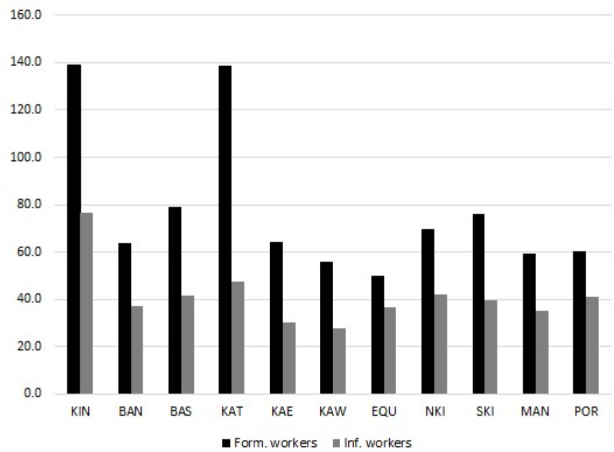
(b) Labor allocation of the skilled



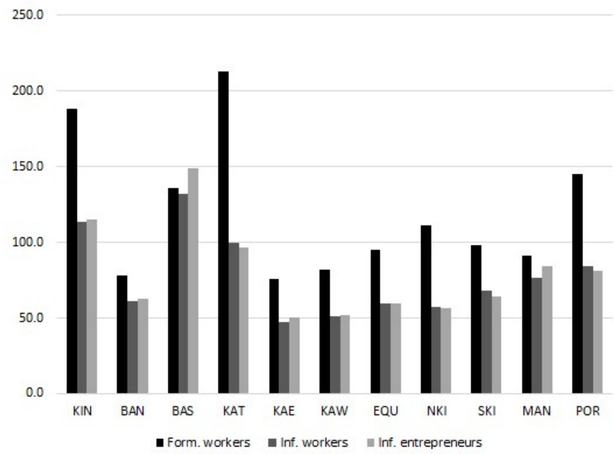
(c) Skill ratio by sector



(d) Skill premium by sector



(e) Monthly earnings of the unskilled



(f) Monthly earnings of the skilled

growth with a cumulative nominal growth of 33% (5.8% per year). However, the impact of economic growth on living standards has been very limited.

SF2. DRC provinces exhibit heterogeneous labor market characteristics. –

Table 3.1: DRC comparison with selected countries

Countries	HDI[a]	CPI[b]	GDP p.c. PPP value[c]	Property rights[d]	Gov. integrity[e]	Tax burden[f]	Inflation rate[g]
DR Congo	0.459	18	767.4	30.1	13.1	74.4	29.3
Angola	0.574	26	6814.3	36.9	15.1	87.3	19.6
Cameroun	0.563	25	3828.2	45.3	20.8	74.8	2.4
Gabon	0.702	31	18495.9	36.9	36.7	74.3	4.8
Equatorial Guinea	0.588	16	22709.7	38.1	15.1	75.1	1.3
Congo, Rep.	0.608	19	6798.9	40.7	23.1	63.3	1.1
Sao-Tomé & Principe	0.609	46	3324.0	41	37.4	88.3	7.9
Chad	0.401	20	2415.3	32.4	15.1	45.8	2.5
Central Africa	0.563	25.1	8144.2	37.7	22.0	72.9	8.6
Sub-Saharan Africa	0.541	32	5661.9	44.0	28.9	76.0	7
World	0.731	43	21385.6	56.6	43.8	77.3	4.8

Notes: Author's computation based on: [a] UNDP's 2019 Human Development Report; [b] Transparency International; [c,g] World bank indicator and IMF as summarized by Heritage Foundation; [d,e,f] Heritage Foundation. Data refer to the year 2019.

The country is currently divided into the city-province of Kinshasa and 25 other provinces. However, before 2015, the country had 11 provinces. As we use data for 2012, our model focuses on these 11 provinces and Table 3.2 relies on the *1.2.3* database to shed light on their labor market characteristics. A random stratification technique has been implemented to guarantee that each of the eleven provinces of DRC has at least one thousand household-level observations.

DRC provinces are strongly heterogeneous in size and in labor market characteristics. Differences in resources endowments are huge, and translate into different business opportunities, productive capacity and labor market composition. Looking at mean monthly wages, larger provinces exhibit larger average levels of earning (the correlation between population size and monthly earning equals 0.58). In particular, Kinshasa (10.5 million inhabitant) is 3.4 times richer than the poorest province of Kasai Oriental (7.1 million inhabitant), 3.1 times richer than the smallest province (Maniema with 2.2 million inhabitant), and 2.1 times richer than the country's average. Yet, geographic mobility is low in DRC. Unreported results from the *1.2.3* database show that the fraction of working aged individuals who have lived administratively in their province of residence for less than 10 years equals 13.4%. And subtracting those who moved for non-economic reasons (study, family reunification, war displaced, etc.), this fraction is reduced to 4.2% of the population. This group of movers include both within-province and between-province movers. The latter group itself is likely to be significantly lower. Our model will disregard the geographic mobility across provinces.

The size of the informal sector is above 80% in all provinces except Kinshasa; it is even greater than 90% in three provinces (Kasai Oriental, Equateur and Province Orientale). The correlation between average monthly wages and the size of the informal sector equals

-0.88. Education is low as around one quarter of the population has completed secondary education. The share of secondary-educated workers varies between 15% in Province Orientale and 59.9% in Kinshasa. The correlation between average monthly wages and the share of secondary educated equals 0.86. Due to these disparities and their potential impact on the creation of wealth, the constitution of February 18, 2006 in its article 181, instituted a common equalization fund so that the richest provinces participate in the development of the poorest provinces through the payment of 10% of their national revenue.

Table 3.2: Heterogeneity in labor market characteristics across DRC Provinces

Province	Population (x 1,000)	Monthly wage	Informal job (as %)	Secondary+ (as %)
Kinshasa	10,558	127,432	62.5	59.9
Bandundu	8,954	46,078	87.6	28.9
Bas-Congo	5,215	72,407	82.9	29.4
Katanga	12,240	93,735	87.1	26.3
Kasaï Oriental	7,190	37,147	93.5	16.8
Kasaï Occidental	5,757	37,151	89.3	21.2
Equateur	8,121	43,572	91.5	16.7
Nord-Kivu	6,240	54,681	85.8	23.1
Sud-Kivu	5,411	56,732	88.9	16.5
Maniema	2,187	40,672	88.9	19.5
Province Orientale	8,589	45,137	91.3	15.0
Unweighted mean	7,315	59,522	86.3	24.8
Coef. of variation	0.363	0.452	0.093	0.487

Notes: Author's computation. Population data are from INS country's statistical report (2015). Wage, Share of informal job and share of Secondary+ are computed from 1.2.3 database.

SF3. Cross-province disparities in public infrastructure is even larger. – The level of public infrastructure is extremely low in DRC. Our data by province are taken from INS country's statistical report (2015). From 2010 to 2015, the public capital spending amounted to 4.0% of total revenues, on average. As shown in Table 3.3, access to drinkable water and electricity is limited (17.5% and 13.4% of the population). Of a hydropower potential of 106,000 MW, only 2,417 MW of hydropower capacity is installed and only 42.1% of it is available for use. Ground transport in DRC has always been difficult. Only 5.4% of the road network is paved. The railways are poorly maintained, crowded and dangerous, and less than one third of the provinces have an international airport. Furthermore, until the end of 2015, there was no airline company, neither public nor private, serving the whole country. The low number of airports and their deplorable state limit exchanges and transactions between provinces and countries. Combined with the insufficiency and poor condition of the road network, this hampers internal exchanges and the development of local potential in a country where only 20% of the national territory is covered by the telephone network and about 7% of the population has access to internet (INS, 2015).

Remarkably, looking at the coefficients of variation related to each infrastructure proxy, regional disparities in public infrastructure are larger (sometimes three times larger) than disparities in labor market characteristics (reported in Table 3.2). In Equateur, only 7% of the population has access to electricity and 2.3% has access to drinkable water. Eight provinces have no international airport. The percentage of paved roads is below 1% in Equateur and Kasai Occidental, and smaller than 10% in eight provinces. In general Kinshasa has much better infrastructure than the rest of the country.

Table 3.3: Heterogeneity in public infrastructure across provinces

Province	Capital spending per worker	Acc. Electricity as % of HH	Acc. water as % of HH	Paved as % of road	Intern. airport	Pub. cap. index
Kinshasa	958.8	74.0	89.0	90.1	Yes	84.4
Bandundu	217.2	2.2	5.6	5.2	No	4.3
Bas-Congo	1243.0	16.1	20.9	20.1	No	19.0
Katanga	1669.1	13.0	20.6	5.5	Yes	13.0
Kasaï Oriental	240.3	0.5	8.8	3.1	No	4.1
Kasaï Occidental	160.6	0.4	3.1	0.8	No	1.4
Equateur	413.0	7.0	2.3	0.6	No	3.3
Nord-Kivu	655.0	5.2	8.3	20.5	Yes	11.3
Sud-Kivu	703.2	10.8	19.8	7.9	No	12.8
Maniema	1182.2	8.8	3.1	6.3	No	6.1
Province Orientale	379.3	9.0	11.5	2.3	Yes	7.6
Unweighted mean	711.1	13.4	17.5	14.8	0.272	15.2
Coef. of variation	0.664	1.479	1.345	1.674	1.633	1.474

Notes: Author's computation based on INS statistical report (2015). The allocation of capital expenditure across Provinces is provided by the capital expenditure plan of the Ministry of Budget. The public capital index in the last column is the unweighted mean of Cols. (2), (3) and (4).

SF4. An overwhelming majority of unskilled workers are employed informally. – Turning out attention to the labor market, the informal sector is overwhelmingly large in DRC. As a consequence of favoritism, looting, wars and other shocks to the economy, the country's labor market is badly out of balance. Since 1990, labor supply has been falling steeply while galloping demography is continuously increasing the demand for jobs. Associated with high levels of poverty and inequality, absence of public provisions for unemployment insurance has favored the emergence of the informal sector. According to IMF (2015), more than 80% of the active population operates outside the labor market regulations and the World-Bank (2005) reports that 78.4% of formal firms are competing with informal firms.

The *1.2.3* survey report of 2012 shows that the informal sector drains 88.6% of assets nationwide, and is likely to constitute an obstacle to faster development. It is seen as a key factor reducing the potential tax base, thereby minimizing the infrastructure spending that the country needs. The informal sector is particularly attractive for the low skilled. In this study, we define skilled individuals as workers with at least a “state diploma degree" (12+ years of education), while the unskilled are those who have not completed secondary education. Figure 3.2a shows that more than 95% of them are employed in the informal sector in virtually all provinces. Two exceptions are Kinshasa (85%) and Bas-Congo (93%).

SF5. A majority of well-educated workers is in the informal sector. – Although the phenomenon is less pronounced for holders of a state diploma, Figure 3.2b show that informality also drains a large fraction of skilled workers. The share of skilled workers employed in the formal sector is close to one third in Kinshasa, Bas-Congo and sud Kivu. The highest share (38%) is obtained in Nord-Kivu. In the other provinces, it varies between 20% and 25%. These results are in line with other studies revealing that the informal economy is recognized for being low-skilled intensive, even if a non negligible fraction of educated workers has informal jobs (Docquier and Iftikhar, 2019; Verick, 2008).

SF6. The formal sector is skill-intensive while the informal sector is not. – As a corollary of the two previous stylized facts, the skill ratio, defined as the ratio of secondary-educated to less educated workers, varies drastically across sectors. The black bar in Figure 3.2c show that the skill ratio is rather large in the formal sector. It varies between 5.6 in Kinshasa and 1.6 in Province Orientale. On average, there are three times as many skilled workers as low-skilled workers in the formal sector. This is way above the skill ratio obtained in the national population, which is around 0.5. Hence, the formal sector in DRC is clearly not representative of the national economy. By contrast, the skill ratio is very low in the informal sector, with the exception of Kinshasa. The country-wide average skill ratio in informality is around 0.4, but it falls to 0.3 when excluding Kinshasa. This is slightly smaller than what we observed in the national population (0.5).

SF7. For both skill groups, formal jobs are better remunerated. – Looking at Figures 3.2e and 3.2f, the levels of monthly earnings are greater in the formal sector for both groups. On average, unskilled workers earn 1.9 times more in the formal sector than informal employees, while secondary-educated workers earn 1.6 times more. These ratios are rather stable across provinces and are also representative of the situation of Kinshasa. There are a few exceptions such as Equateur and Province Orientale where unskilled formal employees earn 40% more than informal employees only, or Bas-Congo where skilled workers exhibit the same average levels of income in both sectors.

Some workers choose to voluntarily work in the informal economy for a host of reasons (e.g., women who are in their prime fertility age or older workers who would like to take advantage of more flexible working hours in the informal sector). In addition, several mechanisms can be used to generate a wage differential between sectors in the competitive labor market setting (such as unobserved heterogeneity in workers' abilities, or a differential in risk, in exposure to rent-seeking, etc.). The facts in Figures 3.2e and 3.2f suggest, however, that informality at large is not a choice despite the fact that some skilled entrepreneurs may find it optimal to operate in informality (see discussion under SF10). Some models assume competitive labor markets and perfect mobility of workers between sectors. Such models fail to explain why a large income differential exists between sectors and why informality attracts workers from all skill groups. Search-and-matching models a la Pissarides (2000) are more compatible with SF7 and better account for endogenous job creation by formal firms (Docquier and Iftikhar, 2019).

SF8. The skill premium varies across sectors and is sometimes larger in informality. – A related issue concerns the levels of the skill premium in both sectors. Figure 3.2d show that skill premia are positive and rather large in both sectors. They vary, however, across provinces. On average, the largest skill premia are observed in provinces with low levels of human capital such as Bas-Congo, Maniema or Province Orientale, which is in line with a standard neo-classical production function with decreasing marginal productivity of skilled labor. The lowest levels are obtained in Kinshasa, Bandundu and the two provinces of Kivu. Skill premia are also correlated across sectors (0.3). What is remarkable is that in the skill premium is on average larger in the informal sector (95%) than in the formal sector (59%). This is the case in eight provinces. Exceptions are Province Orientale and the two provinces of Kivu. The gap between secondary-educated and unskilled workers is larger in informality. Hence, assuming that the informal sector offers a subsistence level to all workers is wrong. This is defendable for unskilled workers, but not for the skilled.

SF9. The informal sector is governed by an entrepreneurial structure embedding land/capital owners and workers of both skill types. – The Congolese informal sector consists of small scale businesses that are easy to launch, require relatively few specific skills, and are characterized by precarious conditions. The *1.2.3* database shows that these activities are concentrated in the agricultural sector, mining industry and small retail business. With an average size of establishments of 1.3 people, it is an atomized sector which mainly consists of micro-units. More than 50% of the informal production units in Congolese agglomerations operate without specific professional premises, and 31.2% carry out their activity at home (Makabu et al., 2006). The sector employs both skilled (15%) and unskilled (85%) individuals, who declare themselves as self-employed, entrepreneurs, unemployed, or searching for a job. Importantly, the sector follows an implicit (and sometimes explicit) entrepreneurial structure, which is mostly governed by the high heterogeneity in access to credit (Mushagalusa-Mudinga et al., 2014; Sara Geenen and Iragi-Mukotanyi, 2013).[7]

[7]The *1.2I3* database reveals that the lack of access to credit is the first obstacle encountered by informal workers, as 98.5% of informal workers report that they do not have access to credit. See also

In agriculture and mining, land ownership is concentrated in the hands of relatively wealthier people, local politicians and churches. Land owners are not in a position to exploit their land directly (Mushagalusa-Mudinga, 2014). Land owners rent out their farmland to peasants, in return for payments which are very rarely made in cash, but much more frequently made in kind or in labor hours. We find it reasonable to assume that land owners have secondary education, and act as entrepreneurs. This is in line with Reyes et al. (2017), who shows that skilled individuals have a low probability of operating alone in the informal sector, with the *1.2.3* database, which reveals that a large fraction of skilled people in informality declare themselves as entrepreneurs, and with Adoho and Doumbia (2018), who documents that top-performer entrepreneurs in the informal sector are well educated. Most of them own large plots of land and plantations, and operate in agriculture using the abundant and available peasant workforce. Other workers (mostly unskilled) produce, and return part of their sale revenues to land owners. The situation is very similar in the retail sector. Due to lack of capital, small retailers fit into structures where they work for wholesalers. They obtain merchandise and small equipment from wholesalers, sell the merchandise, and pay back the agreed amount to wholesalers.

We thus represent the informal sector as a set of skilled entrepreneurs providing capital to workers (mostly unskilled but include some skilled). The latter produce and/or sell the merchandise, return part of their sales revenue to the owners (assimilated to profit in our model) and live with the rest (assimilated to wages in our model). As unemployment benefits do not exist and wages are way greater in the formal sector, we assume that entrepreneurs and workers in informality are queuing for a better job in the formal sector.

As in many low-income countries, informality also includes other odd jobs carried out at the corner of the street and outside any entrepreneurial structure. One often sees them in the streets of big cities (shining of shoes, repairing shoes, selling candy, etc.). However, most of these odd jobs are done by children. In 2000, the ILO estimated that nearly 2 million children aged 10 to 14 were economically active in the DRC, with almost equal numbers of girls and boys.[8] The model we developed does not capture child labor as our data covers the adult population only.

SF10. In informality, skilled workers and entrepreneurs earn similar levels of income. – On average, as apparent from Figure 3.2f, skilled workers and entrepreneurs in informality have similar levels of monthly earnings. This is the case in all provinces. Our model thus assumes that, in equilibrium, skilled people in informality are indifferent between acting as entrepreneur or wage earner. This being said, Mohammad (2014a)

https://www.farmlandgrab.org/post/view/26683.

[8]See the ILO Committee of Experts on the Application of Conventions and Recommendations (CEACR), Reports, Individual Observations, general Observations and Direct Requests (2008-2010), published in 2000

has shown that, although informal firms are smaller than formal firms in DRC, some micro firms are highly productive. The author identifies a upper-tier segment of the informal economy where firms are dynamic and efficient, and where entrepreneurs are well remunerated. Some of them are likely to prefer running their business in informality to searching for alternative employment in the formal sector.[9] The same heterogeneity can be observed in the formal sector where some jobs can be very well remunerated. On average, the level of monthly earnings of informal entrepreneurs is below the formal wage rate, which implies that informality at large is not a choice.

3.3 Model

For all provinces of DRC ($p = 1, ..., P$), we develop a labor market model which is consistent with the stylized facts described above. Workers are infinitely lived and risk neutral. They discount the future at the exogenous market rate r. There are two skill groups, the skilled, in number H_p, corresponding to individuals with at least a "state diploma degree" (12+ years of education), and the low skilled, in number L_p, corresponding to individuals who have not completed secondary education. The total population in province p is given by $N_p \equiv L_p + H_p$ and the skill ratio in the working-age population is defined as $Z_p \equiv \frac{H_p}{L_p}$. At each moment in time, a single homogeneous final good is produced in two different sectors, the formal and informal sectors (labeled F and I). The final good is the *numéraire* and its price is normalized to unity. Formal firms employ skilled and unskilled workers whereas in the informal sector, skilled entrepreneurs employ both skilled and unskilled workers, in line with SF9.

The informal labor market is competitive, whereas the formal labor market is characterized by search frictions, wage bargaining and involuntary informal employment. In each province p, the workers from each skill group ($S = H, L$) are found in one of the two sectors. A fraction e_p^S of the type-S labor force is employed in the formal sector at a wage rate w_p^S, and produces intermediate goods for the final sector. Those who do not find a formal job are absorbed by the informal sector and keep searching for a job in the formal sector. Hence, a fraction $i_p^L = 1 - e_p^L$ of the unskilled labor force is employed in the informal sector; these informal employees earn a competitive wage ω_p^L which is smaller than w_p^L, in line with SF7. Similarly, a share $i_p^H = 1 - e_p^H$ of the skilled labor force is employed in informality. In line with SF9, we distinguish two types of occupation for

[9] This can be reinforced by the fact that the low level of regulations, the weak enforcement of labor laws (Reyes et al., 2017) and the higher tolerance for informal activities observed in DRC can make the informal sector more attractive to some entrepreneurs (Garcıa, 2017). Despite having job opportunities in the formal sector because of such skills, some individuals might prefer the combination of monetary rewards and greater flexibility (in terms of working hours, work relationships, responsibilities, etc.) in the informal sector (Mohammad, 2014b).

them. A fraction $b_p^H i_p^H$ of the high-skilled labor force acts as informal entrepreneurs and makes a business profit equal to π_p^H, whereas the others (i.e., a fraction $i_p^H(1-b_p^H)$ of the labor force) are employed as workers and earn a competitive wage ω_p^H. Skilled individuals in informality are perfectly mobile between the two occupations, which ensures that in equilibrium the earnings of workers and entrepreneurs are equalized ($\omega_p^H = \pi_p^H$). This is in line with SF10 and implies that there is no incentive for high-skilled workers to move from one occupation to the other. Furthermore, SF7 suggests that $\omega_p^H < w_p^H$.

3.3.1 Technology

The **formal sector** F in province p produces a quantity Y_p of final good using a CES combination of intermediate inputs, Y_p^L and Y_p^H, given by:

$$Y_p = A_p \left[\alpha_p {Y_p^L}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_p) {Y_p^H}^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}, \tag{3.1}$$

where A_p denotes the level of total factor productivity (TFP) in the formal sector, α_p is a province-specific parameter governing the preference for the two inputs and their income shares (reflecting the sectoral composition of the economy), σ is the common elasticity of substitution between the two intermediate goods.

Intermediate inputs are produced by unskilled and skilled workers using a linear technology:

$$Y_p^H = e_p^H H_p \tag{3.2}$$

$$Y_p^L = e_p^L L_p, \tag{3.3}$$

where e_p^L and e_p^H are the employment rates of unskilled and skilled workers in the formal labor market of province p.

The level of TFP in the formal sector is increasing in both infrastructure per worker g_p and the skill ratio of intermediate inputs $z_p \equiv \frac{Y_p^H}{Y_p^L}$, which differs from Z_p, the skill ratio in the working-age population. We have:

$$A_p = \overline{A}_p z_p^\eta g_p^\varphi, \tag{3.4}$$

where $\overline{A}_p$ is an exogenous scale factor, η and φ are the elasticity of TFP w.r.t the skill ratio and the amount of infrastructure per capita, respectively. We assume $\eta + \varphi < 1$.[10]

The final good sector is perfectly competitive, implying that the price of each inter-

[10] Bom and Ligthart (2014) find an average elasticity of output to core infrastructure of 0.17. See also As IMF (2014). The upper bound of the range reported in Calderon and Serven (2014) equals 0.1.

mediate input equals its marginal product, which is given by:

$$y_p^L = \overline{A}_p z_p^\eta g_p^\varphi \alpha_p \left[\alpha_p + (1-\alpha_p) z_p^{\frac{\sigma-1}{\sigma}}\right]^{\frac{1}{\sigma-1}}, \tag{3.5}$$

$$y_p^H = \overline{A}_p z_p^\eta g_p^\varphi (1-\alpha_p) z_p^{\frac{-1}{\sigma}} \left[\alpha_p + (1-\alpha_p) z_p^{\frac{\sigma-1}{\sigma}}\right]^{\frac{1}{\sigma-1}}. \tag{3.6}$$

The modeling of the **informal sector** heavily relies on SF9 and SF10 described in Section 3.2. The final good is produced by skilled entrepreneurs with the help of unskilled and skilled employees. Production is governed by a Cobb-Douglas production function. All skilled entrepreneurs are homogeneous in terms of their productivity. The production function of each entrepreneur is given by:

$$\hat{y}_p = B_p h_p^{\psi_p} \ell_p^{\chi_p} \tag{3.7}$$

where B_p is the TFP level, h_p and ℓ_p represent the number of skilled and unskilled workers employed by each entrepreneur, respectively. If follows that total output in informality is given by $\hat{Y}_p = H_p i_p^H b_p^H \hat{y}_p$. Parameters ψ_p and χ_p represent the province-specific elasticities of output per entrepreneur with respect to unskilled and skilled labor, respectively. When calibrating the model without constraining parameter levels, it comes out that $\psi_p + \chi_p$ is almost constant across provinces, and well below one. This decreasing returns to scale property ensures that profits are positive. However, the ratio $\frac{\psi_p}{\chi_p}$ varies with the share of educated workers in the economy. This reflects the heterogeneity of the informal sector. It comprises small-scale businesses run by entrepreneurs and employing a mix of unskilled and skilled workers in most advanced provinces; it is made of family-based, low productivity businesses with a very low demand for skilled labor in the least advanced provinces. We need $\psi_p + \chi_p = \varrho_p < 1$ in all provinces to ensure that profits are positive, and $\frac{\psi_p}{\chi_p} = \kappa_p Z_p^\mu$, so that:

$$\psi_p = \frac{\varrho_p \kappa_p Z_p^\mu}{1 + \kappa_p Z_p^\mu} \tag{3.8}$$

$$\chi_p = \frac{\varrho_p}{1 + \kappa_p Z_p^\mu}, \tag{3.9}$$

where ϱ_p is the share of labor income in total revenues of the informal sector of province p, and μ is the elasticity of entrepreneurs' relative preference for skilled workers with respect to the province-wide supply of skilled labor.

The level of TFP in the informal sector is increasing in the level of infrastructure per worker g_p:

$$B_p = \rho_p \overline{A}_p g_p^\phi, \tag{3.10}$$

where $\rho_p \overline{A}_p$ is the exogenous scale factor (expressed as a fraction ρ_p of the TFP scale factor in the formal sector, without loss of generality), and ϕ is the elasticity of TFP with respect to the amount of infrastructure per capita such that $\phi \leq \varphi$. The latter condition implies that a rise in the level of public infrastructure increases the productivity gap between the formal and informal sectors.

The wage paid to unskilled workers in informal sector is denoted by ω_p^L and that paid to skilled workers is denoted by ω_p^H. At each moment in time, the entrepreneur maximizes profit π_p^H and this determines the demand for unskilled labor, ℓ_p and skilled labor h_p per entrepreneur. The instantaneous profit is given as follows:

$$\pi_p^H = B_p h_p^{\psi_p} \ell_p^{\chi_p} - \omega_p^L \ell_p - \omega_p^H h_p. \tag{3.11}$$

The profit maximization conditions ($\frac{\partial \pi_p^H}{\partial \ell_p} = 0$ and $\frac{\partial \pi_p^H}{\partial h_p}$) implies that the demand for low- and high-skilled labor and the level of profit per entrepreneur are governed by:

$$\begin{aligned} \omega_p^H &= B_p \psi_p h_p^{\psi_p - 1} \ell_p^{\chi_p}, & (3.12) \\ \omega_p^L &= B_p \chi_p h_p^{\psi_p} \ell_p^{\chi_p - 1}, & (3.13) \\ \pi_p^H &= B_p (1 - \psi_p - \chi_p) h_p^{\psi_p} \ell_p^{\chi_p}. & (3.14) \end{aligned}$$

In line with SF10, the skilled workers are perfectly mobile between the two occupational states, informal employment and entrepreneurship. In equilibrium $\omega_p^H = \pi_p^H$, there is no incentive for skilled agents to move from one state to the other. This implies:

$$h_p^* = \frac{\psi_p}{1 - \psi_p - \chi_p}, \tag{3.15}$$

which implies that the number of skilled employees in informality ($i_p^H H_p$) divides into a fraction

$$b_p^H = \frac{1}{1 + h_p^*} \tag{3.16}$$

of entrepreneurs, and a fraction $(1 - b_p^H) = h_p^* (1 + h_p^*)^{-1}$ of workers. As h_p^* is a province-specific constant from Eq. (3.15), the level of earnings of skilled individuals depend on the equilibrium level of ℓ_p, which will be determined below.

3.3.2 Labor Market

Utility is an increasing function of net income. As earnings are greater in the formal sector (in line with SF7), individuals employed in informality (both workers and entrepreneurs) search for formal jobs. At each moment in time, some are matched with open vacancies through a matching function $F(i_p^S, v_p^S)$, where v_p^S is the total number of vacancies available per type-S worker in the formal sector. The matching function exhibits constant returns to scale and is increasing in both arguments. The job finding rate for agents in each period is given by $\lambda(\theta_p^S) = \frac{F(i_p^S, v_p^S)}{i_p^S}$ where $\theta_p^S \equiv \frac{v_p^S}{i_p^S}$ is defined as the market tightness. The hiring rate is defined as $q(\theta_p^S) \equiv \frac{F(i_p^S, v_p^S)}{v_p^S}$ such that $\frac{\partial q(\theta_p^S)}{\partial \theta_p^S} < 0$ and $\frac{\partial \lambda(\theta_p^L)}{\theta_p^S} > 0$. Workers in the formal sector can lose their job and return to the informal sector at an exogenous (job destruction) rate δ_p^S; this is the rate at which workers lose their jobs and go back into the pool of informally employed workers or entrepreneurs.

We first characterize the **Asset Value Functions**. In the formal labor market, firms post vacancies V_p^S at each moment in time. We assume each firm can post one vacancy and pays a cost c_p^S per moment in time for maintaining a vacancy. The costs are skill-specific and such that $c_p^H > c_p^L$. The cost includes, among others, advertising costs, interview arrangements, initial training, time and resources invested by the firms to find a worker as well as the forgone output while maintaining the vacancy. For simplicity, we assume job search costs for workers and costs for creating a vacancy to be nil. Remember that we denote by w_p^S the wage rate paid by formal firms. From the firm's perspective, the lifetime value of a vacancy (V_p^S) and of a filled job (J_p^S) that require a skill type S are given as follows:

$$rV_p^S = -c_p^S + q(\theta_p^S)(J_p^S - V_p^S), \tag{3.17}$$
$$rJ_p^S = y_p^S - w_p^S - \delta_p^S(J_p^S - V_p^S). \tag{3.18}$$

In the steady state equilibrium, the free entry condition implies $V_p^S = 0$. This implies:

$$J_p^S - V_p^S = \frac{y_p^S - w_p^S}{r + \delta_p^S}, \tag{3.19}$$
$$J_p^S - V_p^S = \frac{c_p^S}{q(\theta_p^S)}. \tag{3.20}$$

Hence, the job creation conditions are given by:

$$\frac{y_p^L - w_p^L}{r + \delta_p^L} = \frac{c_p^L}{q(\theta_p^L)}, \tag{3.21}$$
$$\frac{y_p^H - w_p^H}{r + \delta_p^H} = \frac{c_p^H}{q(\theta_p^H)}. \tag{3.22}$$

The job creation equation says that in equilibrium the marginal cost of opening a vacancy equals the marginal profit from a filled job. As for workers, let W_p^S and U_p^S represent the lifetime value of formal employment and informal employment for type-S individuals, respectively. The wage in the formal sector is taxed at the national tax rate τ, and government revenues are used to finance public infrastructure, education, and other types of public expenditures. We ignore the government budget constraint here as (i) it is not province-specific, (ii) the model is static and we disregard the dynamic implications of running a public deficit, and (iii) part of the revenue can be redirected to corruption.

Keeping in mind that $\omega_p^H = \pi_p^H$, the lifetime value of employment and informal employment for type-S workers are given as follows:

$$rW_p^S = w_p^S(1-\tau) - \delta_p^S(W_p^S - U_p^S), \tag{3.23}$$

$$rU_p^S = \omega_p^S + \lambda(\theta_p^S)(W_p^S - U_p^S). \tag{3.24}$$

This yields:

$$W_p^S - U_p^S = \frac{(w_p^S(1-\tau) - \omega_p^S)}{r + \delta_p^S + \lambda(\theta_p^S)}, \forall S = (L, H). \tag{3.25}$$

We can now characterize the conditions governing the **labor allocation and wage rates**. In steady state, the flows into and out of informal employment balances out each other, we have

$$\frac{\dot{I}_p^L}{I_p^L} = \delta_p^L(\frac{L_p}{I_p^L} - 1) - \lambda(\theta_p^L) = 0,$$

$$\frac{\dot{I}_p^H}{I_p^H} = \delta_p^H(\frac{H_p}{I_p^H} - 1) - \lambda(\theta_p^H) = 0.$$

Denoting $\frac{I_p^S}{S_p} = i_p^S$, we have:

$$i_p^L = \frac{\delta_p^L}{\delta_p^L + \lambda(\theta_p^L)}, \tag{3.26}$$

$$i_p^H = \frac{\delta_p^H}{\delta_p^H + \lambda(\theta_p^H)}, \tag{3.27}$$

which implies that $e_p^S = \frac{\lambda(\theta_p^S)}{\delta_p^S + \lambda(\theta_p^S)}, \forall S = (L, H)$ determines the share of type-S individuals employed in the formal sector. The allocation of workers determines the skill ratio in the formal sector:

$$z_p \equiv \frac{e_p^H H_p}{e_p^L L_p} = \frac{\frac{\lambda(\theta_p^H)}{\delta_p^H + \lambda(\theta_p^H)}}{\frac{\lambda(\theta_p^L)}{\delta_p^L + \lambda(\theta_p^L)}} Z_p. \tag{3.28}$$

In the informal sector, the labor market is competitive and the wage rate of unskilled workers is such that the demand for labor equals the supply:

$$\frac{H_p i_p^H \ell_p^*}{1 + h_p^*} = L_p i_p^L, \tag{3.29}$$

which, together with Eq. (3.13), determines ω_p^L and ℓ_p^*. As explained above, once ℓ_p^* is determined, $\omega_p^H = \pi_h^H$ can be computed from Eq. (3.12), or alternatively from Eq. (3.14). In other words, there is no need to equalize the supply and demand of skilled labor to determine the equilibrium level of earnings in informality as the latter can be obtained by plugging the perfect mobility conditions Eq. (3.15) and the equilibrium value ℓ_p^* into Eq. (3.12) or (3.14).

In the formal sector, once the workers are matched with firms they bargain over wage and the wage rate is determined through Nash bargaining as follows:

$$(1-\beta^L)(W_p^L - U_p^L) = \beta^L(J_p^L - V_p^L)(1-\tau), \tag{3.30}$$

$$(1-\beta^H)(W_p^H - U_p^H) = \beta^H(J_p^H - V_p^H)(1-\tau), \tag{3.31}$$

where β^S is the bargaining power of the worker of type-S workers in the formal sector. Keeping in mind that $\omega_p^H = \pi_p^H$, the wages rates are then given as follows:

$$w_p^L = \frac{y_p^L \beta^L (r + \delta_p^L + \lambda(\theta_p^L)) + (1-\beta^L)\frac{\omega_p^L(r+\delta_p^L)}{(1-\tau)}}{r + \delta_p^L + \beta^L \lambda(\theta_p^L)}, \tag{3.32}$$

$$w_p^H = \frac{y_p^H \beta^H (r + \delta_p^H + \lambda(\theta_p^H)) + (1-\beta^H)\frac{\omega_p^H(r+\delta_p^H)}{(1-\tau)}}{r + \delta_p^H + \beta^H \lambda(\theta_p^H)}. \tag{3.33}$$

3.3.3 General Equilibrium

We consider public interventions (τ), the skill composition of the labor force (Z_p), and the level of infrastructure per inhabitant (g_p) as exogenous. We also assume a standard Cobb-Douglas matching function with symmetric elasticities in the formal labor market,

$$F(i_p^S, v_p^S) \equiv \epsilon_p^S \sqrt{I_p^S V_p^S}, \tag{3.34}$$

where $V_p^S = v_p^S L_p$ is the total number of vacancies available for type-S workers.

We can now define the steady state, general equilibrium of our model as following:

Definition 1. For a set of common parameters $\{\sigma, \eta, \varphi, \phi, \tau, r, \delta^S, \beta^S \kappa\}$ and a set of province-specific parameters $X_p \equiv \{L_p, H_p, g_p, \overline{A}_p, \alpha_p, \rho_p, \varrho_p, \kappa_p, \epsilon_p^L, \epsilon_p^H, c_p^L, c_p^H\}$, the general equilibrium is a set $\Gamma_p \equiv \{A_p, B_p, y_p^L, y_p^H, \theta_p^L, \theta_p^H, i_p^L, i_p^H, b_p^H, z_p, w_p^L, w_p^H, \omega_p^L, \omega_p^H, \pi_p^H, \ell_p^*, h_p^*, \psi_p, \chi_p\}$ of endogenous variables satisfying the following 19 conditions, $\Gamma_p = f(X_p)$: (i) definition of technological externalities (3.4), (3.10), (3.8) and (3.9); (ii) profit-maximization conditions (3.5) and (3.6) in the formal sector; (iii) job creation conditions (3.21) and (3.22) in the formal labor market; (iv) labor market equilibrium conditions (3.12), (3.13), (3.14) and the equilibrium condition (3.29) in the informal sector; (v) equilibrium informal employment shares (3.26) and (3.27) for both skill groups; (vi) optimal allocation of high-skilled workers between informal employment and entrepreneurship (3.15) and (3.16); (vii) equilibrium skill ratio (3.28) in the formal sector; and (viii) wage formation conditions (3.32) and (3.33) in the formal labor market.

The other endogenous variables (e.g., asset values, N_p, z_p, e_p^S or π_p^H) can be computed as a transformation of the parameters or as by-products of the endogenous variables. For the purpose of our numerical experiments, we divide the set of province-specific parameters into five categories:

Definition 2. The set of province-specific parameters $X_p \equiv \{X_p^Z, X_p^G, X_p^F, X_p^I, X_p^L\}$ consists of five subsets of parameters, namely the human capital structure, $X_p^Z = \{Z_p\}$, the level of public infrastructure, $X_p^G = \{g_p\}$, technological characteristics of the formal sector, $X_p^F = \{\overline{A}_p, \alpha_p\}$, the technological characteristic of the informal sector, $X_p^I = \{\rho_p, \varrho_p, \kappa_p\}$, labor market frictions, $X_p^L = \{\epsilon_p^L, \epsilon_p^H, c_p^L, c_p^H\}$.

We now explain how we calibrate common and province-specific parameters to exactly match the data and stylized facts described in Section 3.2.

3.3.4 Parameterization

The parameter values are summarized in Table 3.4. Alternative parameters values are considered in the robustness analysis.

In line with the definition of the general equilibrium, we consider 8 parameters as common to all provinces, and we assign them a consensus value from the existing empirical literature. As far as the elasticity of substitution between intermediate goods is concerned, we use the elasticity of substitution between high-skilled and low-skilled workers estimated by Ottaviano and Peri (2012) for a one-sector model. This gives $\sigma = 2$. For the interest rate, we follow Satchi and Temple (2009) setting real interest rate at 4% per year and compute the monthly rate; this gives $r = 0.003$. In line with Gong and van Soest (2012) and Satchi and Temple (2009) on Mexico, the job destruction rates are set to $\delta = 0.060$ for both skill groups. As for the bargaining power of workers, most of the literature uses a value around 0.500. We use $\beta^S = 0.500$ in all provinces.[11]

For η, the existing empirical literature suggests that quantitatively large aggregate schooling externalities are unlikely to exist in developing countries (Acemoglu and Angrist, 2000; Moretti, 2004). Caselli and Ciccone (2013) argue that for a typical poor country, increasing college attainment to the level of the US in 1990 would add less than 4 years to average years of schooling, inducing a 30% increase in TFP. Transposing this to the share of secondary educated means that when the share of skilled workers in developing

[11] Satchi and Temple (2009) recommends a value of 0.700 when informal sector represents more than 30% of the workforce.

countries increases to the US level, the TFP increases by 30%. To calculate the lower bound on η, we take the average share of high school graduates of DRC (30%). Increasing this share to the US level (90%) in the year 2010 involves an increase by 300%. Following Caselli and Ciccone (2013), we assume that this shock induces a 30% rise in TFP, which implies that $\eta = 0.100$.

Finally, hundreds of papers have estimated the elasticity of aggregate output (usually proxied by the GDP per capita or by measures of private output) to public infrastructure using cross-country regressions (Bom and Ligthart, 2014; Dufflo and Pande, 2007; Irmen and Kuehnel, 2009; Wang and Wu, 2015). Using a synthetic index of infrastructure to assess its impact on GDP, Calderon and Serven (2014) obtain a long-run elasticity varying between 0.05 and 0.10. Calderon and Serven (2010) find, however, that the largest contributions of infrastructure development to growth were attained in South-Asia. In sub-Saharan countries, the contribution of infrastructure is smaller. We assume a lower-bound elasticity $\varphi = 0.050$ in the formal sector, and we halve this value in the informal sector ($\phi = 0.025$).

As far as the 12 province-specific parameters are concerned, three of them, $\{L_p, H_p, g_p\}$, are directly obtained from the data described in Section 3.2. In particular, Table 3.3 shows that the amount of capital investment per capita in Col. (1) exhibits low variability across provinces, while the other columns show that the actual level of infrastructure is way greater in Kinshasa and varies a lot across province. We proxy g_p with the public capital index reported in Col. (6) of Table 3.3. This index ranges from 1.4 in Kasai Oriental to 84.4 in Kinshasa. The others parameters in X_p are calibrated to match the same number of moments, namely the informal employment shares by skill groups (i_p^L, i_p^H), the structure of earnings by skill group and by sector $(w_p^L, w_p^H, \omega_p^L, \omega_p^H, \pi_p^H)$, the number/share of skilled entrepreneurs in informality (b_p^H) and the number of unskilled employees per entrepreneur (ℓ_p^*). We assume that $c_p^S = 0.4 w_p^L$ in all provinces, in line with Docquier and Iftikhar (2019).

As we used all the degrees of freedom of the data to identify the needed coefficients, our model is exactly identified and cannot produce a test of its assumptions. In order to establish the relevance of our parameterization method, we examine whether our province-specific parameters exhibit realistic correlations with traditional explanatory variables used in the econometric literature.

The coefficients of correlation between province-specific parameters and potential correlates are provided in Table 3.5. The TFP scale factor in the formal sector $(\overline{A}_p)$ is significantly correlated with the shares of the manufacturing industry in formal output and employment. Importantly, the correlation between $\overline{A}_p$ and population density is insignificant, which suggests that cross-province differences in productivity are not cor-

Table 3.4: Parameters – Summary

Prm.	Definition	Source	Mean	CV
Common to all provinces				
σ	Elast. of subst. btw intermediates	Ottaviano and Peri (2012)	2.000	-
η	Elast. of TFP to human capital in F	Caselli and Ciccone (2013)	0.1	-
φ	Elast. of TFP to infrastructure in F	Calderon and Serven (2014)	0.050	-
ϕ	Elast. of TFP to infrastructure in I	Calderon and Serven (2014)	0.025	-
μ	Elast. of $\frac{\psi_p}{\chi_p}$ to z_p	Calibration outcome	2.000	-
τ	Income tax rate in F	Direction Générale des Impôts (RDC)	0.132	-
r	Monthly interest rate	Satchi and Temple (2009)	0.003	-
δ^S	Monthly job destruction rate	Satchi and Temple (2009)	0.060	-
β^S	Bargaining power	Petrongolo and Pissarides (2001)	0.500	-
Province-specific				
$\overline{A}_p$	TFP scale factor in F	Calibration outcome	193.2	0.380
α_p	Income share parameter in F	Calibration outcome	0.313	0.141
ρ_p	Relative TFP scale factor in I	Calibration outcome	0.589	0.292
ϱ_p	Sum of ψ_p and χ_p	Calibration outcome	0.852	0.242
κ_p	Scale factor in $\frac{\psi_p}{\chi_p}$ function	Calibration outcome	0.886	0.194
ψ_p	Elast. of output to HS labor in I	Calibration outcome	0.235	0.468
χ_p	Elast. of output to LS labor in I	Calibration outcome	0.616	0.223
ϵ_p^L	Scale factor in LS matching fct.	Calibration outcome	0.014	0.342
ϵ_p^H	Scale factor in HS matching fct.	Calibration outcome	0.054	0.369
c_p^L	Cost of posting a LS vacancy	Equal to 0.4 times w_p^L	35.2	0.383
c_p^H	Cost of posting a HS vacancy	Equal to 0.4 times w_p^H	54.1	0.368

Notes: CV = coefficient of variation of province-specific parameters, defined as the ratio of standard deviation to the mean value.

related with urbanization and market size once TFP is deflated by infrastructure and human capital. Remember $\overline{A}_p$ is a residual scale factor; the absence of correlation with Z_p and g_p suggests that the size of technological externalities (φ and η) makes sense. In addition, $\overline{A}_p$ is not correlated with the number of people displaced due to conflicts, which is a proxy for political stability and governance quality in the province. As conflicts mostly arise in resource abundant provinces, this also suggests that benefits of natural wealth and resulting instability costs may cancel out on the aggregate.

The relative productivity in informality (ρ_p) is negatively correlated with internal displacements. It is uncorrelated with the other indicators, which comforts our assumptions that TFP in informality is less affected by the level of infrastructure per capita and uncorrelated with human capital. The elasticity of informal output to high-skilled employment (ψ_p) is positively correlated with most indicators, in contrast with the elasticity to low-skilled employment (χ_p). In particular, the greater the skill ratio in population (Z_p), the greater is the elasticity of informal output to high-skilled labor (ψ_p). This suggests that a skilled-biased externality is more likely to operate in the informal sector than in the formal sector (as α_p is uncorrelated with the skill ratio). The sum $\varrho_p = \chi_p + \psi_p$ is uncorrelated with our indicators. In other words, the informal sector is heterogeneous across provinces; its technology is more skill-intensive in provinces where human capital is less scarce. As far as labor market frictions are concerned, the efficiency of the matching function for low-skilled workers is positively correlated with most indicators, whereas the scale parameter of the matching function for skilled workers is not.

Table 3.5: Parameters – Validation

Correlate	$\overline{A}_p$	α_p	ρ_p	ψ_p	χ_p	ϱ_p	ϵ_p^L	ϵ_p^H
Population density	0.343	-0.400	0.471	0.882*	-0.791*	-0.331	0.890*	0.063
Value added in Manufacturing	0.630*	-0.450	0.284	0.859*	-0.781*	-0.362	0.846*	-0.143
Workers in Manufacturing	0.943*	-0.258	-0.043	0.658*	-0.630*	-0.398	0.571	-0.278
Good roads (as %)	0.458	-0.292	0.168	0.680*	-0.557	-0.052	0.788*	0.018
Nb. business projects	0.594	-0.154	0.317	0.828*	-0.754*	-0.351	0.827*	0.189
Nb. vacancies to be filled	0.520	-0.193	0.209	0.687*	-0.613*	-0.245	0.789*	0.166
Urban population share	0.490	-0.276	0.267	0.832*	-0.768*	-0.395	0.790*	-0.156
People displaced	0.345	0.038	-0.623*	-0.382	0.462	-0.257	0.157	-0.273
Infrastructure per capita	0.447	-0.343	0.457	0.900*	-0.819*	-0.382	0.911*	0.140
Skill ratio in population	0.367	-0.371	0.575	0.947*	-0.886*	-0.492	0.874*	0.069

Notes: Data are obtained from the INS country's statistical report (INS, 2017). Population density is the average number of inhabitants in a given area per square kilometer in the year 2013. Share of the manufacturing sector in formal output and formal employment in the year 2013, respectively. Good roads (%) represents the share of paved road (2016). Nb. business projects and Nb. vacancies to be filled are an annual mean value from 2012-2015 and represent the number of business projects and vacancies to be filled. Urban population share represents the percentage of population living in urban areas. People displaced represents the number of internal displacements due to conflicts and instability (2014). * means significant at the 5% level.

We identify four groups of provinces, as illustrated in Figure 3.2.[12] We use a radar-plot representation to highlight disparities along the five dimensions of Definition 2. The values of the parameters in each province are shown as a fraction of the value of parameter in Kinshasa. The first group includes Bas-Congo (BAS) and Province Orientale (POR), which exhibit relatively high levels of productivity in the formal sector. Bas-Congo takes advantage of its geographic position as the only gate through which the country has access to the ocean. This allows BAS to trade more with other countries, and to benefit from greater tax revenues from trade. A large portion of the products imported and exported by the DRC transits through this province. Province Orientale benefits from its subsoil assets, which are abundant in gold mineral. POR has benefited from the settlement of multinational mining companies (e.g., Kibali Mining) that contributed to shifting the exploitation from *artisanal* to industrial, and to improving institutions. The province also benefits from revenues generated by dynamic cross-border trade with the *Eastern African Community* countries.

The second group includes three provinces, North and South Kivu (NKI, SKI) and Equateur (EQU), which exhibit medium levels of productivity in both sectors. In particular, the provinces of Kivu are well endowed in mineral resources and represent the world reservoir of *columbo-tantalite* as well as of other minerals. They have an advantage of sharing borders with countries experiencing increasing economic growth rates (i.e., Rwanda and Tanzania) (Brenton et al., 2011). However, they have been experiencing

[12] We exclude Katanga, which is almost as rich as Kinshasa due to its abundant mineral resources.

violent conflicts for decades, fueled by both national and regional tensions. This permanent instability has generated huge costs in terms of human lives, social and economic development.

The third groups includes two provinces, Bandundu (BAN) and Maniema (MAN), with low productivity in the formal sector, but with a relatively successful informal sector. Maniema is handicapped by its *landlockedness* as it does not share borders with any of the 9 countries surrounding DRC. Bandundu has limited mineral resources, and suffered from the disorganized exploitation of rubber. Lowes et al. (2017) show that the greater exposure to extraction-oriented institutions has significantly affected BAN in terms of education, wealth and health outcomes. Historically, the severe rationing of formal jobs in these two provinces has contributed to the development of a relatively dynamic informal sector.

Finally, the two provinces of Kasai (KAE and KAW) combine many shortages (infrastructure, health, education, sanitation, etc.) and high labor market frictions. They are the most rural provinces in DRC, with low productivity levels in both sectors. Although they have significant mineral resources such as diamonds, their economies are dominated by small-scale and sparse *artisanal* mining and farms. The IMF (2015) report ranks these two provinces as the poorest of DRC, with an average poverty rate of 76.5%.

3.4 Quantitative Experiments

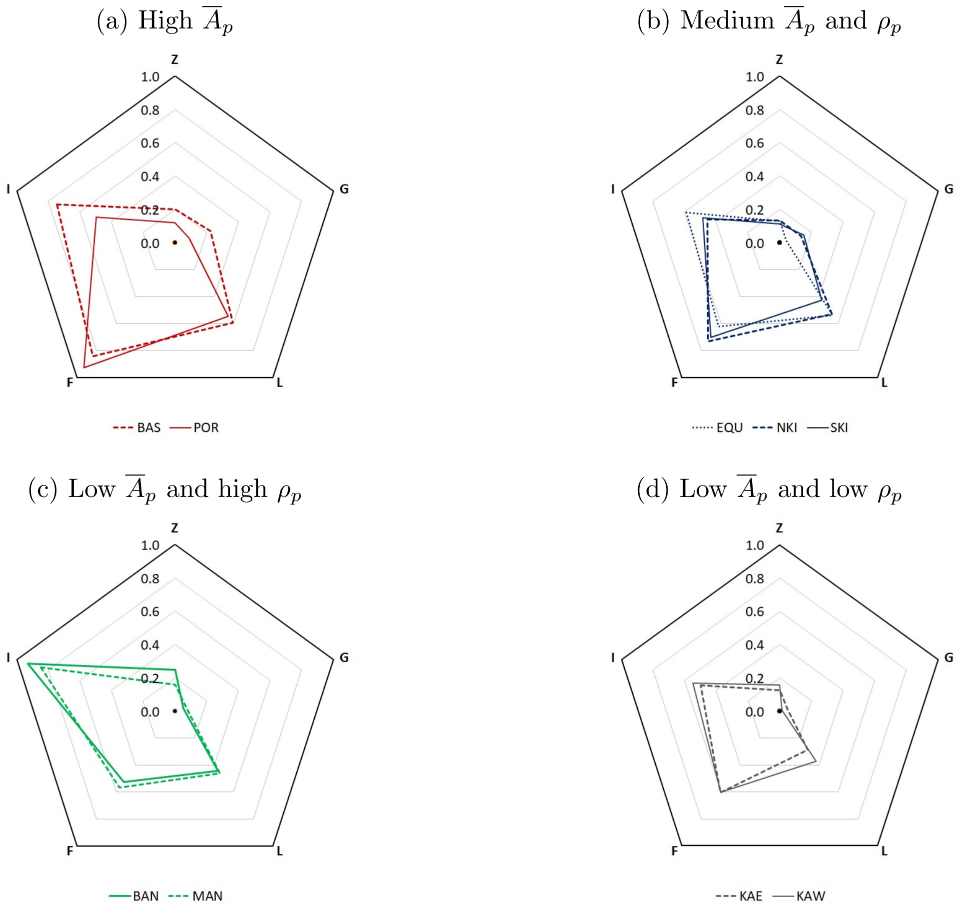
In our numerical experiments, we consider the richest province of Kinshasa (indexed by KIN) as a benchmark, simulate the counterfactual general equilibrium $\overline{\Gamma}_p = f(X_{KIN})$ when province-specific parameters are in totality or partially equalized with those of Kinshasa, and compare it with the observed equilibrium, $\Gamma_p = f(X_p)$. We care about the structure of income, as described by sector- and skill specific levels of income before taxes $(w_p^H, w_p^L, \omega_p^H, \omega_p^L)$, the average income level of unskilled workers ($\overline{w}_p^L \equiv (1 - i_p^L)w_p^L(1 - \tau) + i_p^L\omega_p^L$), and the average level of income per capita in the province ($\overline{w}_p \equiv (H_p\overline{w}_p^H + L_p\overline{w}_p^L)/(H_p + L_p)$.

3.4.1 One-At-A-Time Policy Changes

Let us first consider policy reforms targeting one specific part of the economy at a time. Building on the subsets of parameters in Definition 2, we simulate five counterfactuals:

- Education policies (Z): they lead to a counterfactual equilibrium obtained after replacing X_p^Z by the level observed in Kinshasa, X_{KIN}^Z. This gives $\overline{\Gamma}_p^Z = f(X_{KIN}^Z, X_p^G, X_p^F, X_p^I, X_p^L)$;

Figure 3.2: Broad classification of provinces



Notes: Broad classification based on calibrated parameters from Table 3.4, expressed as a fraction of the value obtained in Kinshasa.

- Infrastructure policies (G): they lead to a counterfactual equilibrium obtained after replacing X_p^G by X_{KIN}^G. This gives $\overline{\Gamma}_p^G = f(X_p^Z, X_{KIN}^G, X_p^F, X_p^I, X_p^L)$;

- Policies influencing the technology of the formal sector (F): they lead to a counterfactual equilibrium defined as $\overline{\Gamma}_p^F = f(X_p^Z, X_p^G, X_{KIN}^F, X_p^I, X_p^L)$;

- Policies influencing the technology of the informal sector (I): they lead to a counterfactual equilibrium defined as $\overline{\Gamma}_p^I = f(X_p^Z, X_p^G, X_p^F, X_{KIN}^I, X_p^L)$;

- Policies influencing labor market frictions (L): they lead to a counterfactual equilibrium defined as $\overline{\Gamma}_p^L = f(X_p^Z, X_p^G, X_p^F, X_p^I, X_{KIN}^L)$.

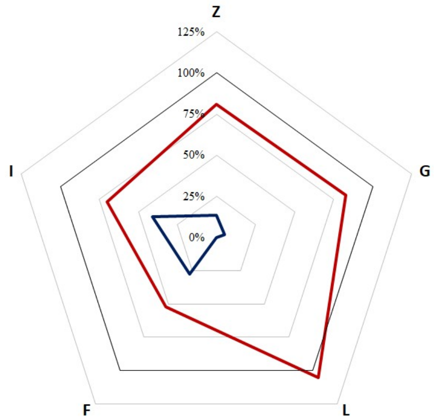
First, we use the same radar-plot representation as above to describe the effect of the each policy on the region-wide average level of income per capita. Results are depicted by the blue curves in Figure 3.3, which represent the effect of each of these policy reforms expressed as percentage of the income gap with Kinshasa, measured on the vertical axis. The closer from the 100% reference level, the higher the explained share of the income gap with Kinshasa. Red curves will be discussed in the next section. The bottom panel shows the effect obtained in nine provinces of DRC (Katanga is excluded, as it is richer than Kinshasa). The top panel shows the unweighted average of the province-specific effects.

The first key message of our analysis is that the technological characteristics in both sectors (X_p^F and X_p^I) are the key determinants of spatial inequalities, somewhat unsurprisingly, followed by human capital and infrastructure. Labor market frictions taken in isolation play a negligible role. The results by province are in line with the broad classification presented in Figure 3.2. In particular, improving the technology of the formal sector (X_p^F) – which mostly consists of increasing the TFP level ($\overline{A}_p$) – is the most effective policy by far in BAN and MAN, but also in the poorest provinces (KAE and KAW) and in EQU. Improving the technology of the informal sector (X_p^I) is the most effective policy in the most productive provinces (BAS and POR). In the two provinces of Kivu, acting on both sectors is desirable.

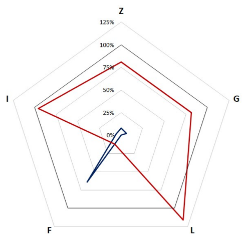
The role of the technology is mostly governed by the role of the exogenous TFP scale factor in both sectors. Remember that Table 3.5 shows that TFP in formality is correlated with the share of the manufacturing in output and employment. TFP partly relates to the quality of local institutions and mineral resource endowment. Good and stable institutions are instrumental to creating an enabling environment for socioeconomic growth. The institutional context in Kinshasa relies on a system called *Branchement* which directly connects economic actors to higher-level authorities and bypasses the role of the provincial government (Nkuku and Titeca, 2018). These alliances between economic actors and high-level political actors can be fragile in turmoil periods; they are, however, less unstable than in less developed provinces. Although the provincial governor of Kinshasa remained in post from 2007 to 2017, other provinces are characterized by chronic instability, which is reflected by a high turnover of provincial leaders. From 2007 to 2017, four governors succeeded each other in South Kivu, three in Equateur, Kasai Occidental and Bandundu. A high number of reshuffles in provincial executive governments were observed in all provinces in general, even in the relatively highly productive Province Orientale (Gérard, 2014). This instability prevents political leaders to carry out ambitious reconstruction or development programs.

Figure 3.3: Effect of one-at-a-time and quadruple policy changes on income-per-worker levels ($\overline{w}_p$)

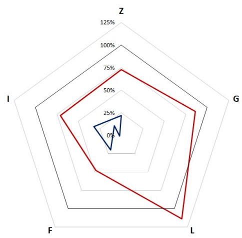
(a) Unweighted mean



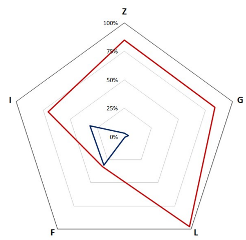
(b) Bandundu



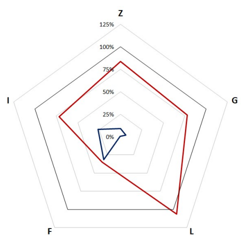
(c) Bas-Congo



(d) Kasaï Oriental

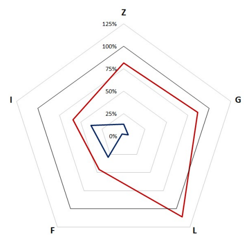


(e) Kasaï Occidential

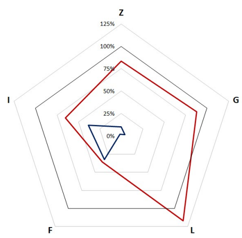


(f) Equateur

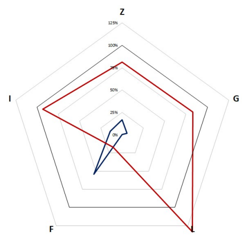
(g) Nord-Kivu



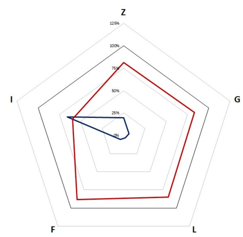
(h) Sud-Kivu



(i) Maniema



(j) Province Orientale



Notes: The blue curve is the fraction of the income gap between each province and Kinshasa that is filled when a single policy $P = (Z, G, F, I, L)$ is implemented. The red curve is the fraction of the gap that is filled when all policies but $P = (Z, G, F, I, L)$ are implemented.

Another key aspect is corruption. Political leaders have strong incentives to manage mineral wealth in their own interest. Hence, natural resource endowments have their pros and cons. On the one hand, they serve as a catalyst for the provision of public services. It is not unusual that mining companies finance projects in the electricity and transport sectors, which benefit local communities in some provinces.[13] Sometimes, they also contribute to the provision of local public services such as healthcare, agriculture extension services, water supply, and education. On the other hand, well endowed provinces are prone to corruption, fraud, interventionism, and rely on armed groups for the control of mining sites (Global Witness, 2006). This contributes to perpetuating extreme poverty and political instability.

The second major finding is that the effectiveness of each policy taken in isolation is relatively small compared to the average income gap of 60 percent. This can be better understood if one looks at their effect on the distribution of income. Figure 3.4 shows the effects of the shocks on the income structure. The background gray area highlights the relative change in the key parameter affected by each policy (as measured on the right scale), while the bars show the relative change in income for the four types of workers, skilled vs unskilled in the formal vs. informal sector (as measured on the left scale).

In Figure 3.4a, we set the technological parameters of the formal sector to the level of Kinshasa (X_p^F). Note that this implies that productivity increases in both sectors, as the productivity in informality is proportional to the TFP of the formal sector as implied by Eq. (3.10). Under a competitive labor market, a similar increase in the TFP of the two sectors would not change the allocation of the labor force. In case of a labor market with frictions, however, the change in TFP induces job creation in the formal sector which, in turn, attracts both types of workers. Given larger frictions for the unskilled, more educated workers move to formality and the skill ratio increases in this sector. On the contrary, the skill ratio decreases in the informal sector, which attenuates the gains for unskilled workers. Hence, although the role of market frictions taken in isolation is negligible, the functioning of the labor market governs the size and distribution of the gains from improving the quality of the technology. Improving the technology of the formal sector benefits all types of workers, but has greater effects on the skilled.

Figure 3.4b shows that improving the technology of the informal sector (X_p^I) has less inegalitarian effects. It induces similar benefits on skilled and unskilled workers in the informal sector. By attracting entrepreneurs and unskilled workers in informality, this policy has negative effects on the skill ratio in the formal sector. However, despite the

[13] Examples of such projects include: (i) the development of power station by Randgold in Kibali (Province oriental), (ii) the construction of roads and bridges by Banro in Twangiza (South Kivu), (iii) the construction of four hydroelectric plants, and (iv) a transmission line by Tenke Fungurume Mining in the Katanga province.

decline in skill ratio, low-skilled workers in the formal sector observe an increase in wage in six provinces. This is due to the fact that wages are determined via bargaining, and an increase in the informal TFP improves outside options of all workers, which results in higher wage for workers in the formal sector. This effect dominates the negative effect of reduced skill ratio on the wage of low-skilled workers in the formal sector. Both types of workers exhibit large income gains in the informal sector. In 5 provinces (BAN, BAS, KAW, MAN and POR), increasing relative productivity in informality is more beneficial to unskilled workers than to entrepreneurs. In 6 provinces (BAS, KAE, KAW, NKI, SKI, POR), this is the most effective policy to combat extreme poverty.

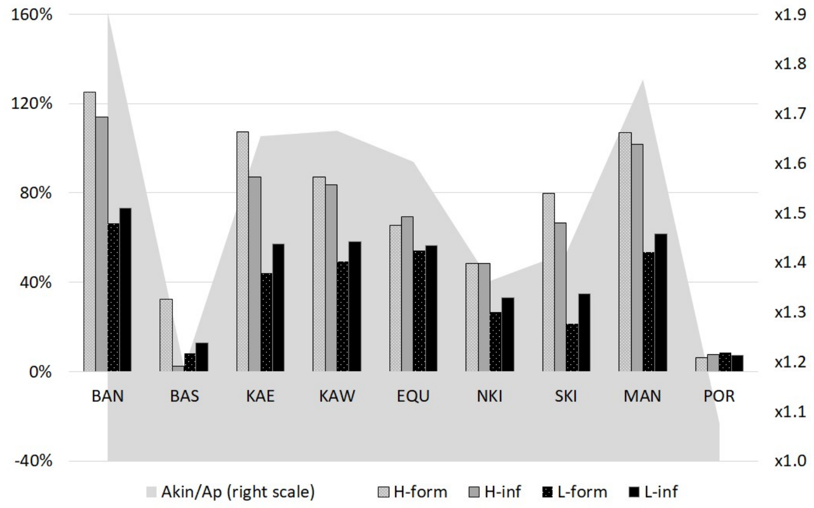
Figure 3.4c shows the effects of setting the proportion of skilled workers to the level observed in Kinshasa (X_p^Z). The shocks are large: the province-wide skill ratio (Z_p) increases by a factor that ranges from 4 in Bas-Congo to 8 in Maniema, as measured on the right scale. A rise in the skill ratio would stimulate the TFP gap between the formal and informal sectors due to technological externalities, but increases the competition between educated workers on both markets. As far as skilled workers are concerned, the competition effect dominates in all provinces. In all provinces, profits from informal businesses decrease. With regard to unskilled workers, their wage rate in the formal sector increases drastically in all provinces. Due to strong labor market frictions, more than ninety percent of unskilled workers remain in informality where their wage rate decreases. This is because entrepreneurship decreases. The majority of unskilled workers is adversely affected by this policy.

In Figure 3.4d, we depicts the effect of decreasing labor market frictions (X_p^L). This leads to small effects in most provinces. Reducing frictions reduce skill ratio in both sectors, and reduces productivity gap between the two sectors due to human capital externalities. This effect dominates the positive effect of a decline in skill ratio on the productivity of the high skilled workers, and attenuates the positive effects of job creation arising from smaller frictions. This policy taken in isolation has very limited effects on the income level of the unskilled workers in the informal sector, who marginally gain due to a decline in the skill ratio in the informal sector.

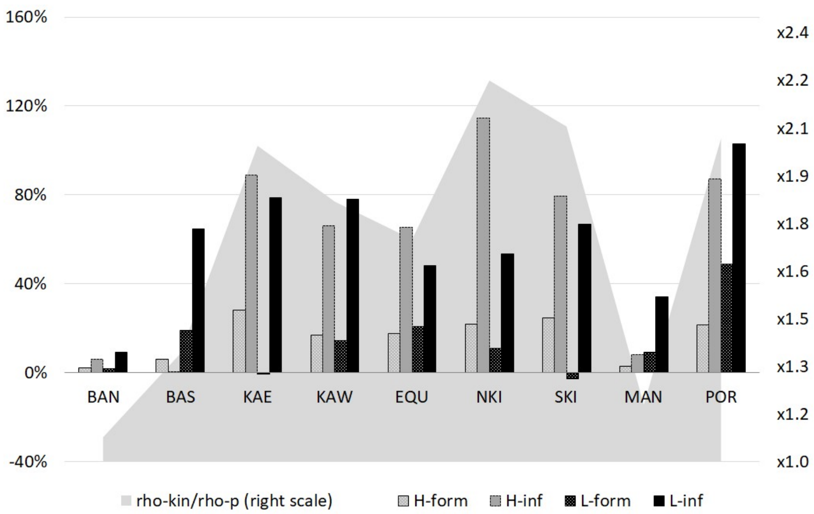
Finally, Figure 3.4e shows the effects of setting the level of public infrastructure to the level observed in Kinshasa (X_p^G). The shocks are large: the province-wide level of infrastructure per capita (g_p) increases by a factor that ranges from 6 in Bas-Congo to 57 in Kasai Occidental, as measured on the right scale. Given the low elasticity of TFP to infrastructure found in the literature, and the fact that infrastructure matters even less in the informal economy, we find small effects on income. A rise in infrastructure increases TFP in both sectors but also increases the productivity gap between the two sectors. The rise in productivity gap worsen the outside options of workers this attenuates the positive effect of infrastructure on wages. Overall, the effect is more beneficial for the skilled, who are more mobile across sectors, and for the unskilled who are employed in the formal sector. The effect on the majority of unskilled workers employed in informality is small.

Figure 3.4: Income responses $(\frac{dw_p^S}{w_p^S}, \frac{d\omega_p^S}{\omega_p^S})$ to policy reforms

(a) Technology in the formal sector (X_p^F)



(b) Technology in the informal sector (X_p^I)



(c) Skill ratio in the population (X_p^Z)

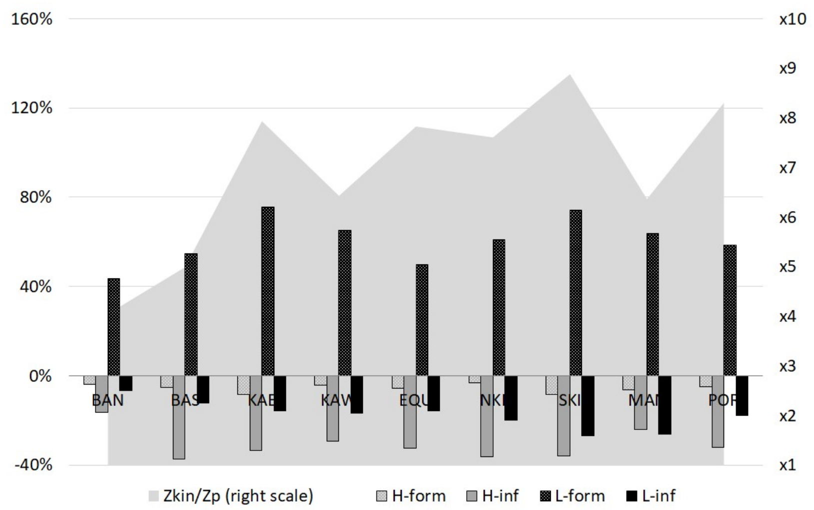
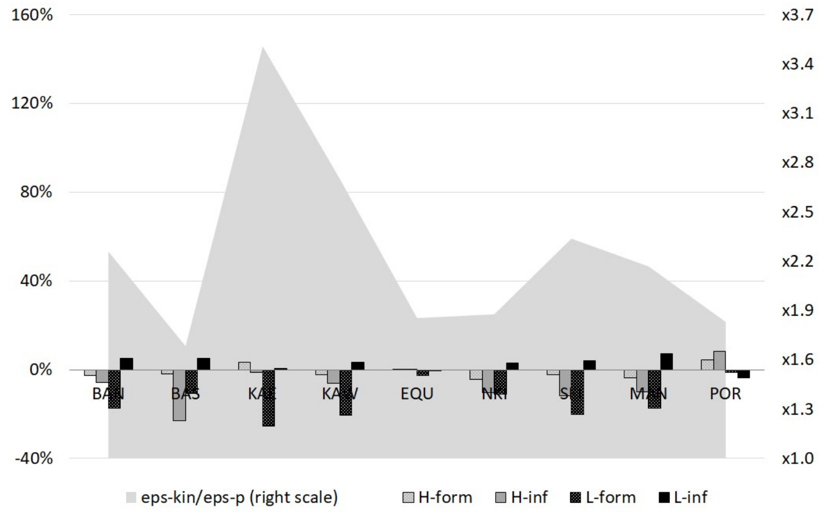
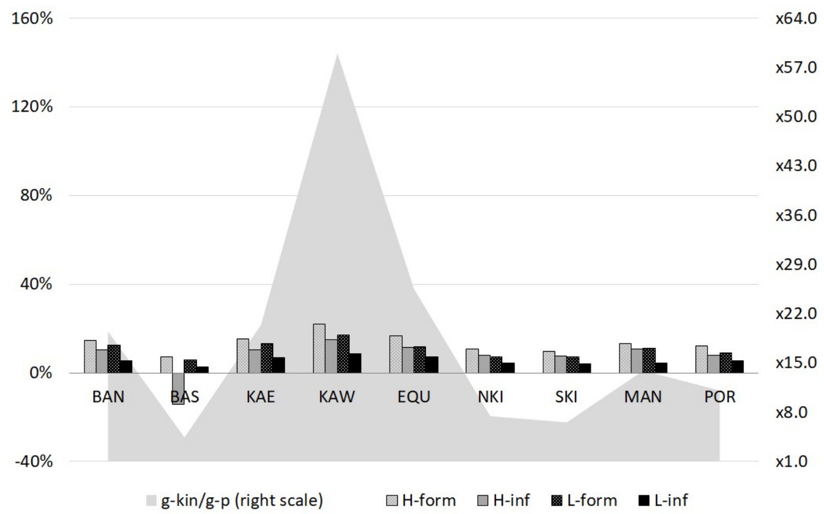


Figure 3.4: Income responses ($\frac{dw_p^S}{w_p^S}, \frac{d\omega_p^S}{\omega_p^S}$) to policy reforms (cont'd)

(d) Labor market frictions (X_p^L)



(e) Public infrastructure (X_p^G)



3.4.2 Complementarities Between Policies: O-Ring Patterns

The main findings of the previous section are that (i) each policy taken in isolation has moderate effects on spatial inequality and potential undesirable effects on the distribution of income (and in turn, on within-province inequalities and extreme poverty); and (ii) most of these undesirable effects are linked to the friction-driven, imperfect mobility of workers across sectors. By construction, if all province-specific parameters were equalized with those of Kinshasa, spatial inequality would disappear. This clearly suggest that reducing spatial inequality is multifaceted challenge requiring a combination of favorable conditions. Below, we illustrate the strong interactions between policies in two ways.

First, the role of complementarities between policy actions appears clearly when summing up the effect of the five policies taken in isolation, and quantifying the role of the interactions (capturing complementarities) between them. Results by province are shown in Figure 3.5. In general, although the role of technological parameters is important, each individual effect (represented by the black and gray areas) accounts for a small portion of the total income gap with Kinshasa, captured by the total length of the bar. The residual interaction term (represented by the red area) accounts for about one fifth of the total effect in three provinces (BAN, NKI and SKI), one third of the total in three provinces (KAE, KAW and EQU). These are Provinces where TFP in the formal sector is low. The residual term reaches its maximum of 41% of the observed income gap with Kinshasa in the highly productive province of Bas-Congo. Smaller interactions are found in Maniema and Province Orientale.

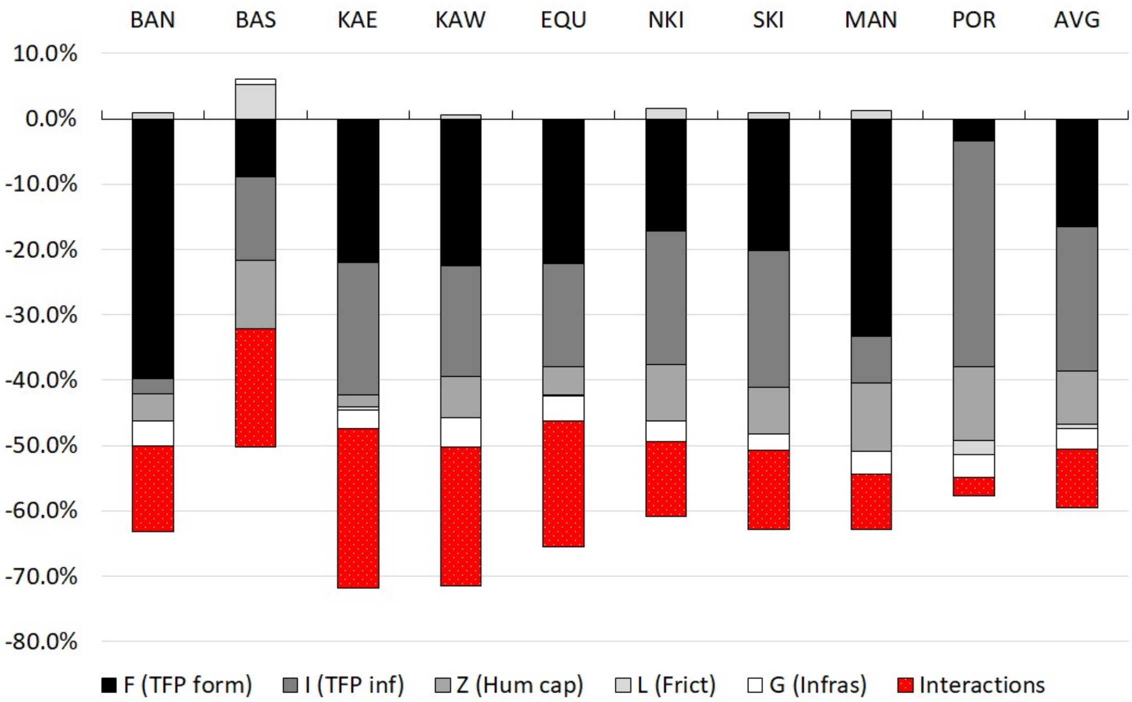
Overall, our results are compatible with the o-ring theory of development (Kremer, 1993). The o-ring theory implies that there are strategic complementarities between ingredients of the development process. These complementarities are important in explaining the development gap between provinces.[14]

Second, we get back to Figure 3.3 but instead of considering one policy at a time (as depicted by the blue curves), we now turn to the opposite exercise, which consists of combining four policies at a time. Results are depicted by the red curves, which represent the effect of leaving aside one policy at a time,[15] and expressing the income change as percentage of the observed gap with Kinshasa. Effects are larger than 100% when one characteristics is on average more detrimental to growth in Kinshasa than in other provinces. Policies targeting the technology parameters of the formal sector are key to increase the average level of income. On average, only half of the gap can be filled

[14]The original o-ring theory focuses more on the micro aspects of production such as workers' skills, type of capital and nature of tasks in a firm's production process. We extend the o-ring concept to the macro ingredients of the development machinery.

[15]Remember that, by construction, combining the five policies would lead to the same equilibrium as in Kinshasa.

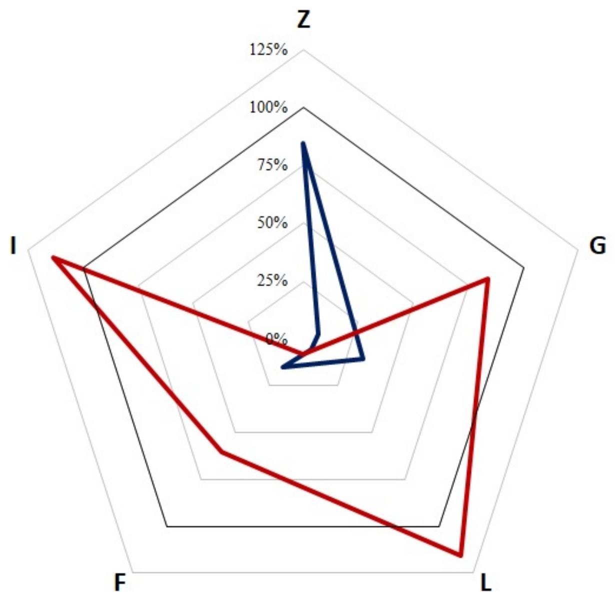
Figure 3.5: Isolated policies and interactions between them ($\overline{w}_p$)



if the TFP of the formal sector is not affected (with bigger losses in BAN, KAE, KAW, EQU, SKI and MAN). In isolation TFP of the final sector only fills 25% of the income gap with Kinshasa. This reveals the interaction of TFP with other factors is of utmost importance. Recall that the TFP scale factor in the formal sector is also associated with governance and institutional stability. In the same vein, only 70% of the gap can be addressed if the technology of the informal sector is not improved (with bigger effects in NKI, SKI and POR). Finally, a gap of 15 to 20% subsists if the levels of education and infrastructure are left unchanged. Yet, working on each of these factors separately and ignoring complementarities between them cannot entirely boost the development process and eradicate poverty.

Before we proceed to next section, let us briefly discuss the effects of one-at-a-time and quadruples of policies on the average informality rate in DRC. Figure 3.6 shows that improving human capital reduces the average informality gap between Kinshasa and other provinces by more than 80%. Our results are in line with (Gong and van Soest (2012); Gong et al. (2004); Mondragon and Pena (2008); Quiroga-Martínez and Fernández-Vázquez (2021)), who find a negative association between the size of the informal sector and human capital. Quiroga-Martínez and Fernández-Vázquez (2021) further deduce that by reducing informality, human capital will also reduce spatial inequalities in Argentina. Their conclusion regarding the size of informality and spatial inequality is based on the paper by Binelli (2016), which provides empirical evidence from Mexico that inequality and informality move together. However, the latter does not link human capital to the informality rate. Our findings reveal that though human capital reduces the size of informality, it does not reduce spatial inequalities (recall Figure 3.3). This suggests that policies that focus on reducing the size of informal sector without implementing policies that increase labour demand in the formal sector for the unskilled have little or no effect on spatial inequalities. Such policies would rather reduce the size of informality at the cost of reduction in the wage for some skill groups. Our findings are consistent with the dual view of informality (Harris and Todaro., 1970; La Porta and Shleifer, 2014b; Lewis, 1954; Rauch, 1991), which suggests that informality is a subsistence sector, and that development comes from the growth of formal employment for both types of workers.

Figure 3.6: Effect of one-at-a-time and quadruple policy changes on informality ($\bar{i}_p$)



Notes: The effect if expressed as a fraction of the gap in the average informality rate with Kinshasa.

3.4.3 Most Effective Policy Pairs

In line with the o-ring theory of development, working on the single source of underdevelopment is poorly effective on the aggregate, and can even be detrimental when focusing on poverty and inequality. We now consider pairs of policies, identify the most effective ones, and identify pairs that can be detrimental for the economy. To this purpose, for each province, we set one pair of policy targets at a time at the Kinshasa levels while leaving other subsets of parameters at their baseline level. We only consider pairs involving the technological parameters of the formal sector (X_p^F).

Results are depicted in Figure 3.7. Panel 3.7a gives the effect on the province-wide average level of income ($\overline{w}_p$), while Panel 3.7b gives the effect on the average income/welfare of unskilled workers, defined as $\overline{w}_p^L = (1 - i_p^L)w_p^L(1 - \tau) + i_p^L\omega_p^L$. The first bar in black depicts the observed income gap with Kinshasa. The second bar in pixelated black gives the income gap obtained under the X_p^F counterfactual (referred to as 'With Ao only' in the legend). Then, the bars in dark red, red, green and yellow show the effect of combining X_p^F with X_p^G, X_p^Z, X_p^L and X_p^I, respectively. The order of these policy pairs is correlated with their average effectiveness when focusing on $\overline{w}_p^L$.

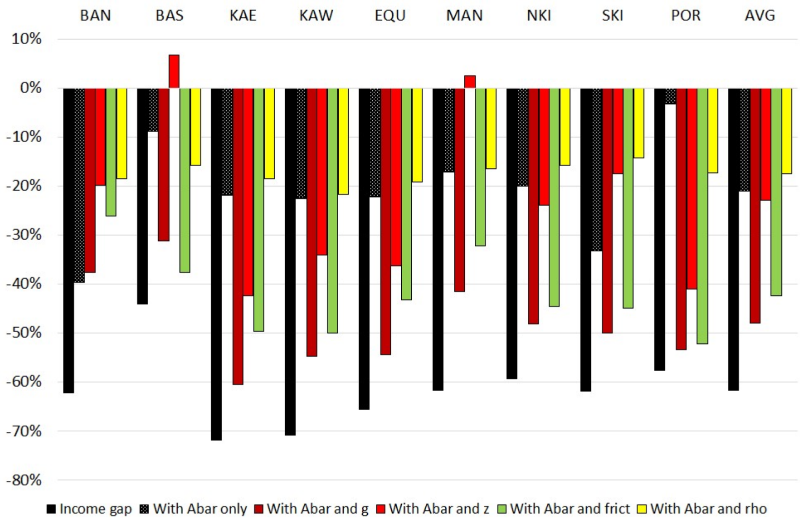
The first key result is that policy pairs targeting X_p^F and X_p^I jointly are always the most effective ones. They drastically reduce the income gap with Kinshasa for the unskilled in the provinces of Kasai, Kivu and Equateur. In Province Orientale, it even brings $\overline{w}_p^L$ at a higher level than in Kinshasa. A significant gap persists in Bandundu and Maniema, where the informal sector is (relatively) highly productive. The effect on the province-wide average income $\overline{w}_p$ is smaller because improving productivity in the informal sector attracts skilled workers in this sector (as entrepreneurs), where income levels are lower. In a very poor country like DRC, a development policy that disregards the situation of the informal sector has low or even detrimental effects on inequality and extreme poverty.

The second key result is that, whatever the outcome variable (province-wide average income or average income of the unskilled), combining X_p^F with policies targeting infrastructure (dark red) or education (red) can be counterproductive. These policy pairs are less effective than improving the technology of the formal sector alone. In line with Figure 3.4, this is due to the fact that infrastructure and education policies stimulate the attractiveness of the formal sector, and labor market frictions are larger for unskilled workers. Hence, these policies reduce the skill ratio in the informal sector, where most of the unskilled are employed. Combining X_p^F with policies targeting labor market frictions has similar effects on $\overline{w}_p^L$ than X_p^F alone, except in Bandundu and Maniema. However, they deteriorate the income of the skilled and the province-wide average income in most provinces. In a poor economy with huge labor market frictions, traditional development policies can induce a smoother transition to formality for the richest part of the labor force, and generate detrimental effect for the overwhelming part of the unskilled population trapped in informality.[16]

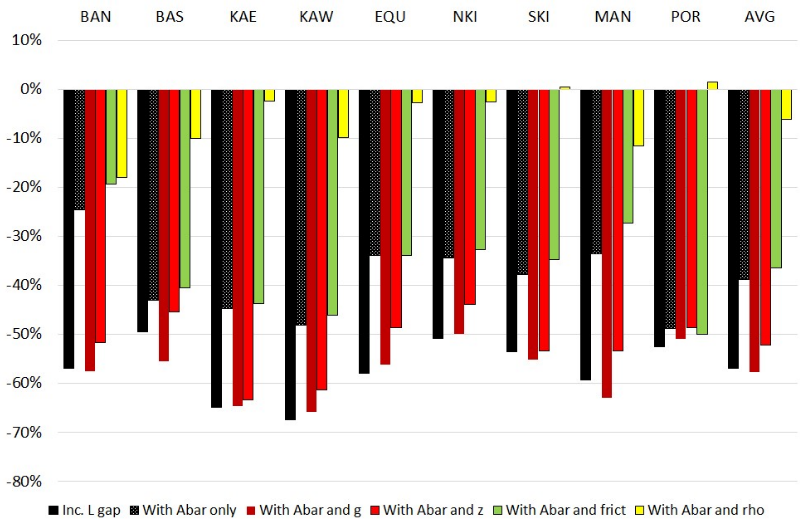
[16] Several effects are at work when policy pairs change which makes it difficult to disentangle different channels and explain why X_p^F alone is more effective than a policy pair. For example consider the policy pair combining X_p^F and X_p^G, improving X_p^F improves both marginal product and outside options of the worker however, combining X_p^F and X_p^G increases the productivity in both sectors but also increases the productivity gap between the formal and informal sector ($\varphi < \phi$) thus deteriorating the outside options of workers. This effect attenuates the positive effects of improving X_p^F and X_p^G on the wages. We thus have a smaller effect of this policy pair on inequality than improving X_p^F alone. Similarly, improving X_p^F and X_p^L has a smaller effect on the average per capita income than X_p^F alone, but for the low-skilled workers this policy pairs out performs improving X_p^F alone. This is because removing frictions improves

Figure 3.7: Effectiveness of policy pairs

(a) Effect on income per capita



(b) Average effect on unskilled workers



3.4.4 Are Labor Market Frictions Irrelevant?

Although labor market frictions prevent unskilled people from moving massively to formality when this sector becomes more attractive, policies targeting frictions alone (see Figure 3.4d) or targeting the technology and frictions jointly (see Figure 3.7b) have little or negative effects on the average income of unskilled workers. Does it mean that frictions are irrelevant? The answer is negative. Remember that our X_p^L counterfactual mostly consists of equalizing the levels of ϵ_p^S, the scale factor in the skill-specific matching func-

the skill ratio in the formal sector. This attenuates the positive effects of X_p^F on high-skilled workers. However, the low skilled workers gain from this rise in the skill ratio in the formal sector. This is why this policy pair is more effective on low-skilled workers than improving X_p^F alone. Similar mechanisms are also involved in other policy pair experiments. The key message is that different policies affect different skill groups and production sectors in heterogeneous ways.

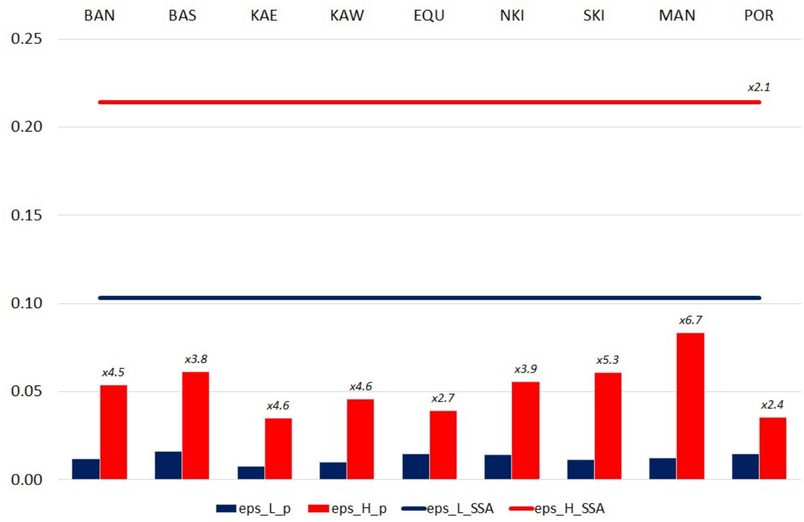
tions, with those observed in Kinshasa (i.e., $\epsilon_p^L = 0.027$ and $\epsilon_p^H = 0.056$). These levels are greater than in the rest of DRC, but are still very low compared to the levels estimated in other countries. For example, Docquier and Iftikhar (2019) obtain average levels of $\epsilon_p^L = 0.103$ and $\epsilon_p^H = 0.214$ in a sample of 34 sub-Saharan African countries which, by mere chance, implies a ratio $\epsilon_p^H/\epsilon_p^L$ equal to that of Kinshasa.

In Figure 3.8, we simulate the effect of applying the average sub-Saharan levels of ϵ_p^S to all provinces. Panel 3.8a shows the calibrated levels of ϵ_p^L (blue bar) and ϵ_p^H (red bar); the number above the red bar gives the ratio $\epsilon_p^H/\epsilon_p^L$ calibrated in the province, which ranges from 2.4 in Province Orientale to 6.7 in Maniema. The blue and red lines are the average levels observed in sub-Saharan Africa, taken from Docquier and Iftikhar (2019). Panel 3.8b presents the effect of such a drastic policy on the average income of the unskilled ($\overline{w}_p^L$ in black) and on the province-wide average income ($\overline{w}_p$ in gray). This Panel shows that a dramatic reduction in labor market frictions has small but positive effects on $\overline{w}_p^L$ in 5 out of the 9 provinces. The largest gains are obtained in Bas Congo, Kasai Oriental, South Kivu. An adverse effect is obtained in Equateur, North Kivu and Province Orientale where the current levels of ratio ϵ_p^L are the greatest (albeit small). When focusing on $\overline{w}_p$, the effect is positive and large in all provinces, evidencing large gains for the skilled.

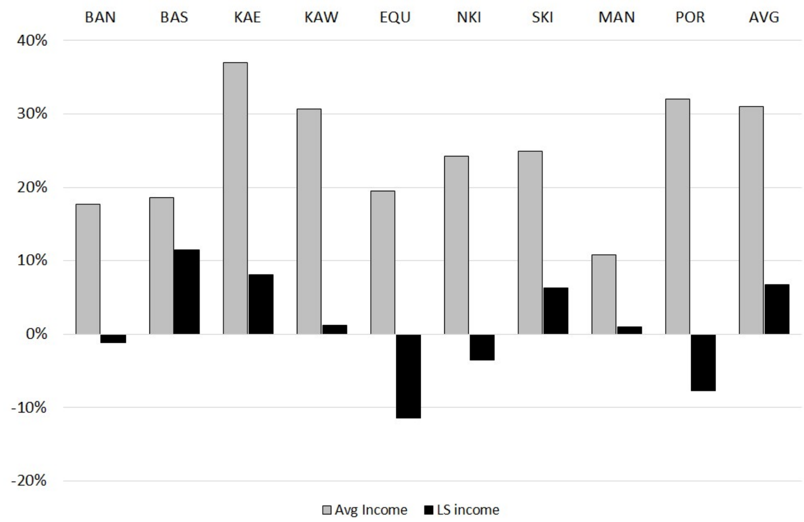
Figure 3.8b shows that the average income increases for both skilled and unskilled workers in DRC. This implies that reducing labour market frictions could reduce poverty and be welfare enhancing. However, looking at the wages of the workers in the two sectors reveals interesting information. Figure 3.9 shows that wages increases for skilled workers in all provinces including Kinshasa in both sectors. By contrast, income levels decrease in both sectors for the unskilled workers. How do we reconcile this result with one presented in Figure 3.8b? Reducing frictions allows mobility of workers from informal to the formal sector, thereby reducing informality rate in all provinces for both skill groups. Average wages increase in most of the provinces due to a decline in the informality rate in all provinces. But the opportunities in the formal sector do not respond much to the changes in labor market frictions. Hence, the mobility of workers reduces the skill ratio in both sectors (on average by 70% in the formal sector while by 36% in the informal sector) and leads to a decline in the wage of unskilled workers and a rise in the wage of skilled workers in both sectors. This brings us to an interesting conclusion that combatting informality by reducing frictions and without expanding opportunities in the formal sector reduces poverty along the extensive margin but increases it along the intensive margin. Once more, this provides additional evidence that reducing informality alone is not the cure to poverty and inequality if the formal sector is not attractive enough.

Figure 3.8: Effect of a dramatic decrease in labor market frictions

(a) Effect on income per capita



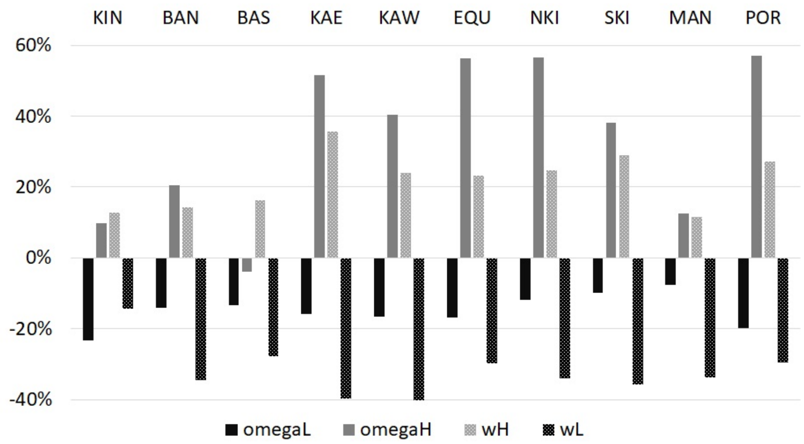
(b) Average effect on unskilled workers



3.4.5 Robustness Checks

In this section, we check the robustness of our results to the value of parameters. We first consider a monthly job destruction rate of 0.04 (instead of 0.06 in the baseline). Second, there is bunch of studies suggesting that quantitatively large aggregate schooling externalities are unlikely to exist in developing countries. We consider a variant with $\eta = 0.0$ (Acemoglu and Angrist, 2000), instead of 0.10. Third, Angrist (1995) recommends a value of σ above 2 to explain the trends in the college premium in developing countries. Our third variant assumes $\sigma = 3.0$, instead of 2.0. Finally, we consider the upper bound of the range of elasticity of TFP to public infrastructure, $\varphi = 0.10$ instead of 0.05, as

Figure 3.9: Effect of a dramatic decrease in labor market frictions on wages



suggested by Calderon and Serven (2014).[17]

Our variable of interest is the province-wide average level of income under the five policy experiments. Focusing on the sum of isolated policies and interactions terms as in Figure 3.5, Figure 3.10 compares the baseline results with those obtained after changing these elasticities.

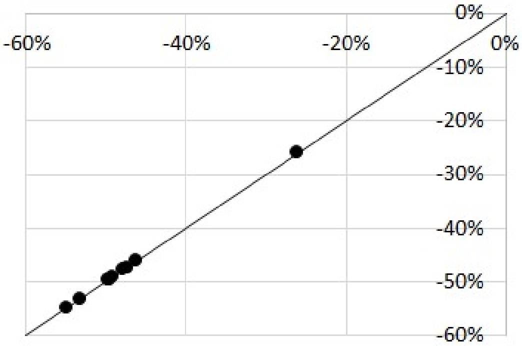
The results are robust to the parameters δ, σ and η ensuring our results in the baseline scenario are not driven by the values of these parameters. However, the results appear to be sensitive to the value of φ. The sum of isolated effects are smaller while the interaction term is much bigger with a higher value of φ. Nonetheless, our measure of public infrastructure is uncorrelated with $\bar{A}_p$ and ρ_p which means we are able to isolate the effect of g_p on income from the scale factors of TFP in the two sectors. Hence, the baseline level of φ does a good job at capturing the effects of g_p on productivity. However, this implies, with the higher elasticity of TFP to public infrastructure, the stronger the o-ring patterns of economic development. The value of φ governs the size of the complementaries between policies.

3.5 Conclusion

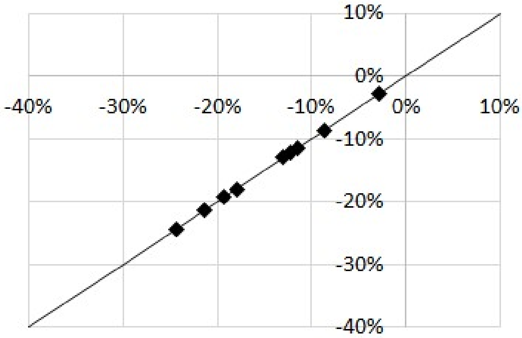
This paper focuses on the causes of spatial and within-province inequalities in DRC, one of the world's poorest countries. We build a two-sector model with labor market frictions to explain income disparities between provinces, sectors (formal vs. informal) and skill groups. We parameterize it to exactly match the observed labor allocation of workers and distribution of income. The calibration reveals large differences across provinces, both in observed characteristics and identified parameters. We then conduct a set of

[17]The elasticity of TFP to public infrastructure in the informal sector ϕ is always set at the half of φ.

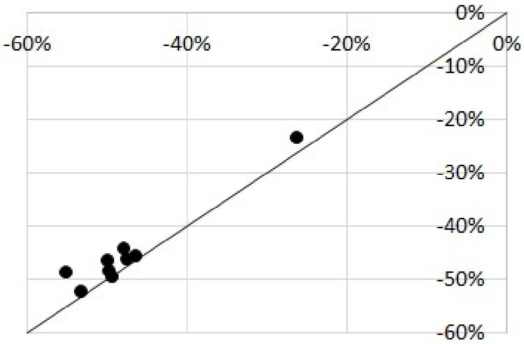
Figure 3.10: Robustness checks - Sum of isolated effects and residual interaction term



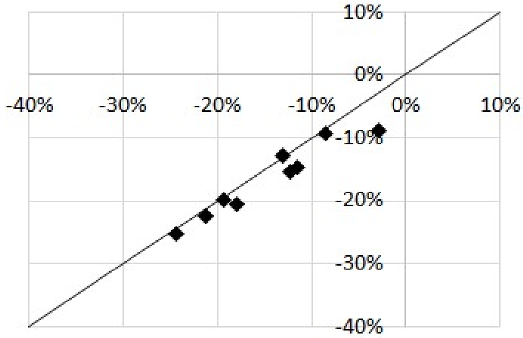
(a) Sum of isolated effects ($\delta = 0.04$)



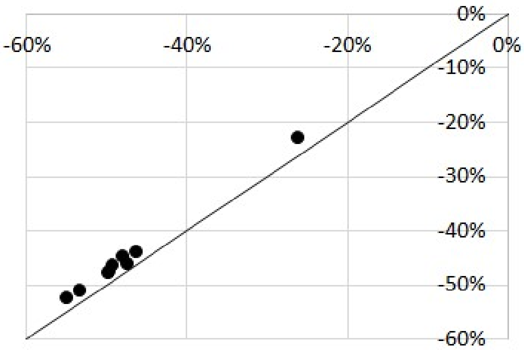
(b) Interaction term ($\delta = 0.04$)



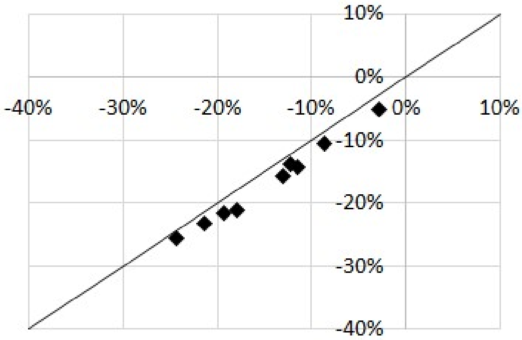
(c) Sum of isolated effects ($\eta = 0$)



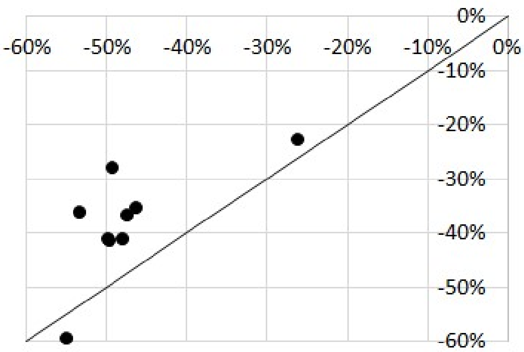
(d) Interaction term ($\eta = 0$)



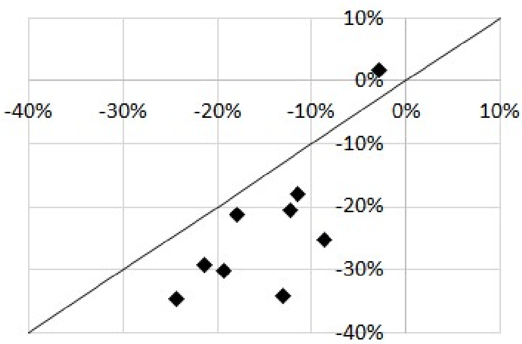
(e) Sum of isolated effects ($\sigma = 3.0$)



(f) Interaction term ($\sigma = 3.0$)



(g) Sum of isolated effects ($\varphi = 0.1$)



(h) Interaction term ($\varphi = 0.1$)

policy experiments to analyze the role of technologies in the formal and informal sectors, human capital, public infrastructure and labor market frictions in explaining spatial and within-province inequalities.

We highlight the high level of complementarity between policies, identify strong o-ring patterns of spatial inequality, and shed light on the role of labor market frictions. Income disparities are mostly determined by the technological characteristics, reflecting both endowment in mineral resources, geographic position and institutional quality. A development policy that disregards the situation of the informal sector has low or even detrimental effects on inequality and extreme poverty. In particular, policies targeting education, labor market frictions, or public infrastructure in isolation have little effects as they mostly impact productivity in the formal sector, and reduces the skill ratio and productivity in informality, where many unskilled workers are trapped.

Our paper sheds light on what are the most relevant policy areas in terms of achieving higher income in a province and combating extreme poverty. We also highlight how different policies have heterogeneous effects across provinces and skill groups. The follow-up research question is to identify which of these policy actions are more feasible in term of cost effectiveness. The answer to this question is greatly hindered by the availability of data on the levels of public investment in the drivers of growth and human development.

Bibliography

Acemoglu, D. and Angrist, J. (2000). How large are human-capital externalities? evidence from compulsory schooling laws. *NBER Macroeconomics Annual*, 15:9—59.

Acemoglu, D., Johnson, S., and J.A., R. (2005). Institutions as a fundamental cause of long-run growth. *Handbook of Economic Growth. Ph. Aghion and S. N/ Durlauf (eds.)*, Chapter 6:385—472.

Adoho, F. and Doumbia, D. (2018). Informal sector heterogeneity and income inequality: Evidence from the democratic republic of congo. *The World Bank*.

Alemán-Castilla, B. (2006). The effect of trade liberalization on informality and wages: Evidence from mexico. *Discussion Paper N°. 763*.

Allingham, M. and Sandmo, A. (1972). Income tax evasion: A theoretical analysis. *Journal of Public Economics*, 1:323–338.

Amuedo-Dorantes, C. (2004). Determinants and poverty implications of informal sector work in chile. *Economic Development and Cultural Change*, 52(2):349–368.

Angrist, J. (1995). The economic returns to schooling in the west bank and gaza strip. *American Economic Review*, 85(5):1065–1087.

Asiedu, E. and Lien, D. (2011). Democracy, foreign direct investment and natural resources. *Journal of International Economics*, 84(1):99—111.

Bajada, C. and Schneider, F. (2005). The shadow economies of the asia-pacific. *Pacific Economic Review*, 10(3):379–401.

Banerjee, A., Duflo, E., Goldberg, N., Karlan, K., Osei, R., Parienté, W., Shapiro, J., Thuysbaert, B., and Christopher, C. (2015). A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science*, 348(6236):1260799.

Barro, R. (1989). conomic growth in a cross section of countries. *NBER WorkingPapers*, 3120.

Barro, R. and Sala-i Martin, X. (2003). *Economic Growth*. The MIT Pres, Cambridge London.

Beaudry, P. and David, G. (2003). Wages and employment in the united states and germany: What explains the differences? *American Economic Review*, 93:573–603.

Becker, G. (1964). *Human Capital*. Columbia University Press: New York.

Bernhardt, I. (1994). Comparative advantage in self-employment and paid work. *Canadian Journal of Economics*, 27:273–289.

Berry, A. and Sabot, R. (1978). Labour market performance in developing countries: a survey. *World Development*, 6(12):1199–1242.

Besley, T. and Persson, T. (2014). Why do developing countries tax so little? *Journal of Economic perspectives*, 28(4):99–120.

Bhattacharya, A. and Kharas, H. (2011). Raising the bar on infrastructure development. *G-24, mimeo.*

Bhattacharyya, S. and Hodler, R. (2010). Natural resources, democracy and corruption. *European Economic Review*, 54(4):608–621.

Binelli, C. (2016). Wage inequality and informality:evidence from mexico. *IZA Journal of Labor and Developmet*, 5(5).

Blinder, A. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources*, 8(4):436–455.

Bloom, D., DavidCanning, D., and Sevilla, J. (2004). The effect of health on economic growth: A production function approach. *World Development*, 32(1):1–13.

Bohn, H. and Deacon, R. (2000). Ownership risk, investment, and the use of natural resources. *American Economic Review*, 90(3):526–549.

Bom, P. R. and Ligthart, J. (2014). What have we learned from three decades of research on the productivity of public capital? *Journal of Economic Surveys*, 28(5):889–916.

Bose, N., Capasso, S., and Wurm, M. (2012). The impact of banking development on the size of shadow economies. *Journal of Economic Studies*, 32(6):620–638.

Brenton, P., Bashinge, C., Hossein, C., Shiho, N., and Ntagoma, J.-B. (2011). Facilitating cross-border trade between the drc and neighbors in the great lakes region of africa: Improving conditions for poor traders. Technical report, 62992-AFR, World Bank, Washington, DC.

Breusch, T. (2005). Estimating the underground economy using mimic models. *Working Paper, Canberra, ACT 0200.*

Buehn, A. and Schneider, F. (2008). Mimic models, cointegration and error correction: An application to the french shadow economy. *Discussion Paper N° 3306.*

Calderon, C. and Serven, L. (2010). Infrastructure and economic development in sub-saharan africa. *Journal of African Economies*, 19(1):13–87.

Calderon, C. and Serven, L. (2014). Infrastructure, growth, and inequality: an overview. *Policy Research Working Paper Series (World Bank)*, 7034.

Caselli, F. and Ciccone, A. (2013). The contribution of schooling in development accounting: Results from a nonparametric upper bound. *Journal of Development Economics*, 104:199–211.

Christofides, Louis, N. and Panos, P. (2020). Self/paid-employment. public/private sector selection and wage differentials. *Labor Economics*, 9(6):737–762.

De Soto, H. (1989). *The Other Path: The Invisible Revolution in the Third World*. Harper and Row, New York.

De Soto, H. (2000). *The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else*. Bantam Press, New York.

Deakin, S. (2013). Addressing labour market segmentation: The role of labour law' governance. *Centre for Business Research: Working Paper*, 52.

Del Boca, D. (1981). Parallel economy and allocation of time. *Micros-Quarterly Journal of Microeconomics*, 4(2):13–18.

Dell'Anno, R. (2003). Estimating the shadow economy in italy: a structural equation approach. *Working Paper No. 2003-07*.

Dell'Anno, R. (2008). What is the relationship between unofficial and official economy? an analysis in latin american countries. *European Journal of Economics, Issue 12*.

Dell'Anno, R. (2010). Institutions and human development in the latin american informal economy. *Constitutional Political Economy*, 21(3):207–230.

Dell'Anno, R. and Schneider, F. (2003). The shadow economy of italy and other oecd countries: What do we know? *Journal of Public Finance and Public Choice*, 21(2-3):97–120.

Dell'Anno, R. and Schneider, F. (2006). Estimating the underground economy by using mimic models: A response to t. breusch's critique. *Working Paper No. 0607*.

Dickens, W.and Lang, K. (1985). A test of dual labour market theory. *American Economic Review*, 4(75):792–805.

Dikhanov, Y. and Ward, M. (2001). Evolution of the global distribution of income in 1970–99. *In Proceedings of the 53rd Session of the International Statistical Institute held in Republic of Korea*.

Dimova, R., Gang, I., and Landon-Lane, J. (2011). Revealed informal activity. *IZA Discussion Papers*, 5607.

Docquier, F. and Iftikhar, Z. (2019). Brain drain, informality and inequality: A search-and-matching model for sub-saharan africa. *Journal of International Economics*, 120:109–125.

Docquier, F., Naval, J., and Muller, T. (2017). Informality and long-run growth. *Scandinavian Journal of economics*, 119(4):1040–1085.

Dufflo, E. and Pande, R. (2007). Dams. *Quarterly Journal of Economics*, 122:601–646.

Elgin, C. and Oztunali, O. (2012). Shadow economies around the world: Model based estimates. *Working Paper 2012/05*.

Elgin, C. and Oztunali, O. (2014). Institutions, informal economy, and economic development. *Journal of Emerging Markets Finance and Trade*, 50(4):145–162.

Elgin, C. and Schneider, F. (2016). Shadow economies in oecd countries: Dge vs mimic approaches. *Review of Social, Economic Administrative Studies*, 30(1):1–32.

Estache, A. (2010). Infrastructure finance in developing countries: An overview. *EIB Papers, European Investment Bank, Economics Department.*

European_Commission (1998). Communication of the commission on undeclared work. Technical report, Available at https://eur-lex.europa.eu/legal-content/EN/LSU/?uri=CELEX%3A51998DC0219, New York.

Falco, P., Kerr, A., Rankin, N., Sandefur, J., and Teal, F. (2010). The returns to formality and informality in urban africa. *CSAE Working Paper 3, Oxford*.

Fay, M., Michael, T., Daniel, B., and Csordas, S. (2011). Infrastructure and sustainable development. *In Post-Crisis Growth and Development, edited by S. Fardoust, K. Yongbeom, and C. Sepúlveda. Washington, DC: World Bank.*

Feld, L. and Schneider, F. (2010). Survey on the shadow economy and undeclared earnings in oecd countries. *German Economic Review*, 11(2):109–149.

Fields, G. (1990). Labour market modelling and the urban informal sector: Theory and evidence. In *In D. Thurnham, B. Salome, and A. Schwarz : The Informal Sector Revisited, (eds), Paris: OECD*. OECD.

Fields, G. (2009). Segmented labor market models in developing countries. In *The Oxford Handbook of Philosophy of Economics*. Oxford University Press, New York.

Firpo, S., Fortin, N., and Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, 77:953–973.

Garcıa, G. A. (2017). Labor informality: Choice or sign of segmentation? a quantile regression approach at the regional level for colombia. *Review of Development Economics*, 21(4):985–1017.

Gennaioli, N., LaPorta, R., Lopez-de Silanes, F., and Shleifer, A. (2014). Growth in regions. *Journal of Economic Growth*, 19(3):259–309.

Giles, D. (1999). Modeling the hidden economy and the tax-gap in new zealand. *Empirical Economics*, 24:621–640.

Giles, D. and Lindsay, T. (2012). Taxes and the canadian underground economy. *Canadian Tax Paper, No. 106.*

Gong, X. and van Soest, A. (2012). Wage differentials and mobility in the urban labour market: A panel data analysis for mexico. *Labour Economics*, 9(4):513–529.

Gong, X., van Soest, A., and Villagomez, E. (2004). Mobility in the urban labor market: A panel data analysis for mexico. *Economic Development and Cultural Change*, 53(1):1–36.

Gunderson, M. (1979). Earnings differentials between the public and private sectors. *Canadian Journal of Economics*, 12:228–242.

Gutiérrez_Romero, K., Holder, C., and Mayers, S. (2006). Estimating the size of underground economy in barbados. *Journal of Business, Finance and Economics in Emerging Economies*, 4(1):196–227.

Gutiérrez_Romero, R. (2010). The dynamics of the informal economy. *Working Paper 2010-07*.

Gyimah-Brempong, K. (2011). Education and economic development in africa. *African Development Review*, 23(2):219–236.

Gérard, G. (2014). Rd congo : l'échec des pouvoirs provinciaux, une nouvelle étape dans la déconstruction de la troisième république. *Les Cahiers d'Afrique de l'Est*, (48).

Günther, I. and Launov, A. (2017). Informal employment in developing countries. *Journal of Development Economics*, 97:88–98.

Hanuschek, E. (2013). Economic growth in developing countries: The role of human capital. *Economics of Education Review*, 37:204–212.

Harris, J. R. and Todaro., M. P. (1970). Migration, unemployment, and development: A two-sector analysis. *American Economic Review*, 60(1):126–42.

Heckman, J. J. (1974). Shadow prices, market wages and labor supply. *Econometrica*, 42:679–694.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47:153–161.

Helberger, C. and Knepel, H. (1988). How big is the shadow economy? a re-analysis of the unobserved variable approach of b. s. frey and h. weck hannemann. *European Economic Review*, 32(4):965–976.

Henderson, J., Storeygard, A., and Weil, D. (2012). Measuring economic growth from outer space. *American Economic Review*, 102(2):994–1028.

Herderschee, J., Kaiser, K., and Samba, D. (2012). Resilience of an african giant: boosting growth and development inthe democratic republic of congo. *Washington, DC: World Bank*.

IMF (2014). Recovery strengthens, remains uneven. *World Economic Outlook, International Monetary Fund: Washington DC*.

IMF (2015). Democratic republic of the congo, country report. Technical Report 15/281, International Monetary Fund.

Ingram, G. and Kessides, C. (1994). Infrastructure for development. *Finance and Development*, 31(3):18–21.

Irmen, A. and Kuehnel, J. (2009). Productive government expenditure and economic growth. *Journal of Economic Surveys*, 23(4):692–733.

Isachsen, A. and Strøm, S. (1985). The size and growth of the hidden economy in norway. *Review of Income and Wealth*, 31(1):21–38.

Johnson, S abd Kaufmann, D. and Zoido-Lobatón, P. (1998). Regulatory discretion and the unofficial economy. *American Economic Review*, 88(2):387–392.

Joreskog, K. and Goldberger, A. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, 70:631–639.

Jütting, J. and De Laiglesia, J. (2009). Is informal normal? towards more and better jobs in developing countries. *Paris: OECD Development Centre*.

Koenker, R. and Bassett, Jr., G. (1978). Regression quantiles. *Econometrica*, 46:33–50.

Kremer, M. (1993). The o-ring theory of economic development. *Quarterly Journal of Economics*, 108(3):551–575.

La Porta, R. and Shleifer, A. (2014a). Informality and development. *Journal of Economic Perspectives*, 28(3):109–126.

La Porta, R. and Shleifer, A. (2014b). Informality and development. *Journal of Economic Perspectives*, 28(3):109–126.

Lee, L. (1983). Generalized econometric models with selectivity. *Econometrica*, 51(3):545–61.

Levy, S. (2008). *Good Intentions, Bad Outcomes: Social Policy, Informality, and Economic Growth in Mexico*. Brookings Institution Press, Washington.

Lewis, W. (1954). Economic development with unlimited supplies of labor. *The Manchester School*, 22(2):139–191.

Makabu, T., Nkenda, M., and Mba, M. (2006). Le secteur informel en milieu urbain en république démocratique du congo: Performances, insertion, perspectives. principaux résultats de la phase 2 de l'enquête 1-2-3, 2004-2005. *DIAL*.

Maloney, W. (1999). Informality imply segmentation in urban labor markets? evidence from sectoral transitions in mexico. *The World Bank Economic Review*, 13:275–302.

Maloney, W. (2004). Informality revisited. *World Development*, 32:1159–1178.

Mankiw, N., Romer, D., and Weil, D. (1992). A contribution to the empirics ofeconomic growth. *TheQuarterly Journal of Economics*.

Medina, L. and Schneider, F. (2018). Shadow economies around the world: What did we learn over the last 20 years? *Working Paper 18/17*.

Milanovic, B. (2002). True world income distribution, 1988 and 1993: First calculation based on household surveys alone. *Economic Journal*, CXII(2):51–92.

Milanovic, B. (2016). *Global Inequality: A New Approach for the Age of Globalization*. Harvard UniversityPress, Cambridge, Massachusetts.

Mincer, J. (1974). *Schooling, Experience, and Earnings*. NBER and Columbian University Press: New York.

Mogensen, G., Kvist, H., Kfrmendi, E., and Pedersen, S. (1995). The shadow economy in denmark 1994: Measurement and results. Technical report, The Rockwool Foundation Research Unit, Copenhagen.

Mohammad, A. (2014a). Comparing micro and informal firms in drc. *unpublished draft, World Bank, Washington, DC.*

Mohammad, A. (2014b). Comparing micro and informal firms in drc. *unpublished draft, World Bank, Washington, DC.*

Mondragon, C. and Pena, X. (2008). Business ownership and self-employment in developing economies: the colombian case. *CEDE Working Paper, Universidad de los Andes*, 2008(3).

Moretti, E. (2004). Workers' education, spillovers, and productivity: Evidence from plant-level production functions. *American Economic Review*, 94(3):656–690.

Mueller, B. (2020). Why public policies fail: Policymaking under complexity. *EconomiA*, 21(2):311–323.

Mushagalusa-Mudinga, E. (2014). L'accaparement des terres dans la province du sud-kivu: expériences paysannes. *Ressources.*

Mushagalusa-Mudinga, E., Ansoms, A., and Claessens, K. (2014). Competition over soil and subsoil. land grabbing by local elites in south kivu, drc. *In: A. Ansoms and T. Hilhorst, Losing your land. Dispossession in the Great Lakes. Indiana University Press*, 7:68.

Méon, P.-G., Weill, L., and Schneider, F. (2011). Does taking the shadow economy into account matter when measuring aggregate efficiency? *Applied Economics*, 43(18):2303–2311.

Nkuku, A. and Titeca, K. (2018). Market governance in kinshasa: the competition for informal revenue through connections (branchement). *Institue of Development Policy - Working Paper*, 2018.03.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3):693–709.

Olson, O. and Congdon Fors, H. (2004). Congo: The prize of predation. *Journal of Peace Research*, 41(3):321–336.

Ottaviano, G. and Peri, G. (2012). Rethinking the effect of immigration on wages. *Journal of the European Economic Association*, 10(1):152–197.

Park, T. (1979). Reconciliation between personal income and taxable income. Technical report, Washington D.C.: Bureau of Economic Analysis, Washington D.C.

Petrongolo, B. and Pissarides, C. A. (2001). Looking into the black box: A survey of the matching function. *Journal of Economic Literature*, 39(2):390–431.

Pocius, A. (2015). Evaluation of the shadow economy and informal employment trends in lithuania. *Lithuanian Journal of Statistics*, 54(1):18–32.

Polachek, S. and Xiang, J. (2009). The gender pay gap across countries: A human capital approach. *SOEP papers on Multidisciplinary Panel Data Research*, 227.

Proville, J., Zavala-Araiza, D., and Wagner, G. (2017). Nighttime lights: A global, long-term look at links to socio-economic trends. *PLoS ONE*, 12(3):1–12.

Quiroga-Martínez, F. and Fernández-Vázquez, E. (2021). Education as a key to reduce spatial inequalities and informality in argentinean regional labour markets. *Regio nal Science Policy and Practice*, 13:177–189.

Rauch, J. (1991). Modeling the informal sector formally. *Journal of Development Economics*, 35(1):33–47.

Reyes, A., Alvaro, G., Dino, M., Carly, P., and Javier, S.-R. (2017). Democratic republic of congo: Jobs diagnostic. Technical report, World Bank CC BY 3.0 IGO.

Sala-i Martin, X. (2006). The world distribution of income: Falling poverty and . . . convergence. period. *Quarterly Journal of Economics*, 121(2):351–397.

Sara Geenen, S. and Iragi-Mukotanyi, F. (2013). Les grands poissons mangent les petits: multiples aspects d'un conflit autour d'une concession minière au sud-kivu. *Politique Africaine*, 3(131):121–141.

Satchi, M. and Temple, J. (2009). Online appendix for labor markets and productivity in developing countries. *Review of Economic Dynamics*, 06-167.

Schermelleh-Engel, K., Moosbrugger, H., and Müller, H. (2003). Evaluating the fit of structural equation models: test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8(2):23– 74.

Schneider, F. (2005). Shadow economies around the world: What do we really know? *European Journal of Political Economy*, 21(3):598–642.

Schneider, F. (2007). Shadow economies and corruption all over the world: New estimates for 145 countries. *Economics*.

Schneider, F. (2012). The shadow economy and work in the shadow: what do we (not) know? *IZA Discussion Paper*, (6423).

Schneider, F. (2017). Shadow economies around the world: New results for 158 countries over 1991-2015. *Working paper N° 46*.

Schneider, F. and Buehn, A. (2016). Estimating the size of the shadow economy: methods, problems and open questions. *Turkish Economic Review*, 3(2):256–280.

Schneider, F., Buehn, A., and Montenegro, C. (2010). New estimates for the shadow economies all over the world. *International Economic Journal*, 24(4):443–461.

Schneider, F. and Enste, D. (2000). Shadow economies: Size, causes, and consequences. *Journal of Economic Literature*, 38(1):77–114.

Schneider, F., Raczkowski, K., and Mróz, B. (2015). Shadow economy and tax evasion in the eu. *Journal of Money Laundering Control*, 18(1):34–51.

Schultz, T. (1990). Testing the neoclassical model of family labor supply and fertility. *Journal of Human Resources*, 25:599–634.

Singh, A., Jain-Chandra, S., and Mohommad, A. (2012). Inclusive growth, institutions, and the underground economy. *Working Paper N°12/47*.

Stapleton, D. (1978). Analyzing political participation data with a mimic model. *Sociological Methodology*, 15:52–74.

Tansel, A. (2005). Public private employment choice. wage differentials. and gender in turkey. *Economic Development and Cultural Change*, 53(2):453–477.

Tanz, V. (1983). The underground economy in the united states: Annual estimates, 1930-80. *Staff Papers (International Monetary Fund)*, 30(2):283–305.

Tanzi, V. (1999). Uses and abuses of estimates of the underground economy. *The Economic Journal*, 109:338–347.

Theil, H. and Seale JL, J. (1994). The geographic distribution of world income, 1950-1990. *De Economist)*, 142:387–419. Dedicated to the memory of Jan Tinbergen.

Thomas, J. (1992). *Informal economic activity. LSE, Handbooks in Economics*. Harvester Wheatsheaf, London.

Trebicka, B. (2014). Mimic model: A tool to estimate the shadow economy. *Academic Journal of Interdisciplinary Studies*, 3(6):295–300.

Trost, Robert, P. and Lee, L.-F. (1984). Technical training and earnings: A polytomous model with selectivity. *Review of Economics and Statistics*, 66(2):151–56.

Tsui, K. (2014). More oil, less democracy: evidence from worldwide crude oil discoveries. *Economic Journal*, 121:89–115.

Töpfer, M. (2017). Detailed RIF decomposition with selection: The gender pay gap in italy. *Working Paper*, 26.

Verick, S. (2008). The impact of globalization on the informal sector in africa. Economic and Social Policy Division, UNECA, Addis Ababa.

Vuletin, G. (2008). Measuring the informal economy in latin america and the caribbean. *Working paper 08/102*.

Wacziarg, R. (2002). Review of easterly's the elusive quest for growth. *Journal of Economic Literature*, XL:907–918.

Wang, Y. and Wu, B. (2015). Railways and the local economy: evidence from qingzang railway. *Economic Development and Cultural Change*, 63(3):551–588.

Williams, C. (2010). Out of the shadows: explaining the undeclared economy in baltic countries. *Journal of Baltic Studies*, 41(1):3–22.

Williams, C. (2012). Tackling entrepreneurship in the informal economy: evaluating the policy options. *Journal of Entrepreneurship and Public Policy*, 1(2):111–124.

Williams, C. and Round, J. (2008). Re-theorising the nature of informal employment: some lessons from ukraine. *International Sociology*, 3(23):367–88.

World-Bank (2005). Enterprise survey. Technical report, World Bank. http://www.enterprisesurveys.org/Data/ExploreTopics/informality.

Zhou, F. and Oostendorp, R. (2014). Measuring true sales and underreporting with matched firm-level survey and tax office data. *The Review of Economics and Statistics*, (96):563–576.