Chapter 3 Theoretical reflections and empirical evidence on informal employment: An application using panel data with small within country-variation

Capítulo 3 Reflecciones teóricas y evidencia empírica sobre la informalidad laboral: Una aplicación usando datos panel con variación pequeña por país

GUILLERMO-PEÓN, Sylvia Beatriz* & ESTRADA-QUIROZ, Liliana

Benemérita Universidad Autónoma de Puebla

ID 1st Author: *Sylvia Beatriz, Guillermo-Peón* / **ORC ID:** 0000-0002-0510-3645, **CVU CONACYT ID**: 30484

ID 1st Co-author: *Liliana, Estrada-Quiroz* / **ORC ID:** 0000-0002-0168-602X, **CVU CONACYT ID**: 87000

DOI: 10.35429/H.2022.7.26.45

S. Guillermo & L. Estrada

*silvia.guillermo@correo.buap.mx

A. Marroquín, A. Álvarez, M. Morado and P. Gómez (Coord) Social Sciences and Humanities. Handbooks-©ECORFAN-México, Querétaro, 2022.

Abstract

This research paper presents some theoretical reflections on the economic sources of the informal economy phenomenon. The discussion is presented under a new approach called the *non-retributed factors approach*. Under this approach, we consider informal economy, all economic activities in which at least one factor of production is non-retributed or under-retributed. Additionally, using panel data for Latin American countries and other developing countries as well as data for developed countries, we present empirical evidence regarding the impact of physical capital and human capital on informal employment rates. Because these variables show very small variation over time, the *small within-group variation* characteristic must be considered when choosing the appropriate model estimation technique with panel data. Our findings show that the scarcer the physical and human capital, the higher the informal employment rates will be.

Informal Employment, Informal Economy, Panel data model, Small within-group variation

Resumen

Este artículo de investigación presenta algunas reflexiones teóricas sobre las fuentes económicas del fenómeno de la economía informal. La discusión se presenta bajo un nuevo enfoque llamado "*enfoque de factores no retribuidos*". Bajo este enfoque, consideramos economía informal, todas las actividades en las cuales al menos un factor de producción no es retribuido o es sub-retribuido. Adicionalmente, usando un panel de datos para países Latinoamericanos y otros países en desarrollo, así como datos para países desarrollados, presentamos evidencia empírica acerca del impacto del capital físico y humano sobre las tasas de empleo informal. Debido a que estas variables presentan una pequeña variación en el tiempo, la característica de "*pequeña variación dentro de grupo*" debe ser considerada cuando se elige la técnica de estimación apropiada con un panel de datos para el modelo. Nuestros hallazgos muestran que mientras más escaso sean el capital físico y el capital humano, más elevadas serán las tasas de empleo informal.

Empleo Informal, Economía Informal, Modelo con datos panel, Variación pequeña dentro de grupo

1. Introduction

In Latin America and many developing countries worldwide, the informal sector and informal employment remain substantial. Countries like Bolivia, Honduras, and Paraguay, show informal employment rates above 70 percent, while in Peru, El Salvador, and Nicaragua, this rate is above 60 percent. On the other hand, the GDP share of the informal economy in these countries has been relevant; as an example, the GDP share of the informal economy has been 23.3 percent on average for Mexico during the period 2003-2017[1]. These data are just a few examples of how important and persistent the phenomenon of informality continues to be. The economic activity under informality conditions is relevant as a source of employment and for producing goods and services. More importantly, knowing that informality is associated with poverty and precarity, these data tell a story about countries' difficulties in improving economic growth and social welfare.

International organizations have made vigorous efforts to measure the informal economy and informal labor in order to make cross-country comparisons, and we may find vast literature with theoretical explanations of this phenomenon. However, there is scarce literature showing evidence regarding factors influencing informality. In this paper, we use panel data for developing and developed countries to analyze the influence of human and physical capital, output growth, institutional functioning, and the cost of starting a formal business, on the country´s informal employment rate. One of the characteristics of these variables is the small variation they show over time for each country. Data are not sufficiently rich in information when there is little within-country variation. Therefore, to estimate the model, the appropriate econometric technique must be chosen; otherwise, estimates will be poor and unreliable.

---

[1] Own calculation based on data from the National Survey of Occupation and Employment (ENOE), 2003-2017 II trimester, INEGI http://en.www.inegi.org.mx/programas/enoe/14ymas/

Our results show that education and physical capital are key factors that influence the informal employment rate. In particular, the empirical evidence provided by the estimated model shows that increasing human capital will reduce the informal employment rate.

The paper is organized as follows: Section 2 explains the informal economy, informal sector, and informal employment concepts with a brief theoretical framework. Additionally, this section shows some stylized facts regarding human and physical capital in developing and developed countries that help us understand the relationship between the variables used in the model. Also, as an example, we present data and a brief analysis of informal employment and the GDP share of the informal activity for the particular case of Mexico and some sociodemographic characteristics of informal workers contrasted with those of the formal workers. Section 3 presents the model and a detailed explanation of the estimation methodology considering the small within-country variation characteristic of the variables. This section also explains the methodology implemented to solve collinearity (the sequential regression method suggested by Graham, 1997 and Dorman et al., 2013) and heteroskedasticity problems. Estimation results are interpreted and analyzed in Section 4. And finally, concluding remarks are presented in Section 5.

2 Conceptual framework and facts

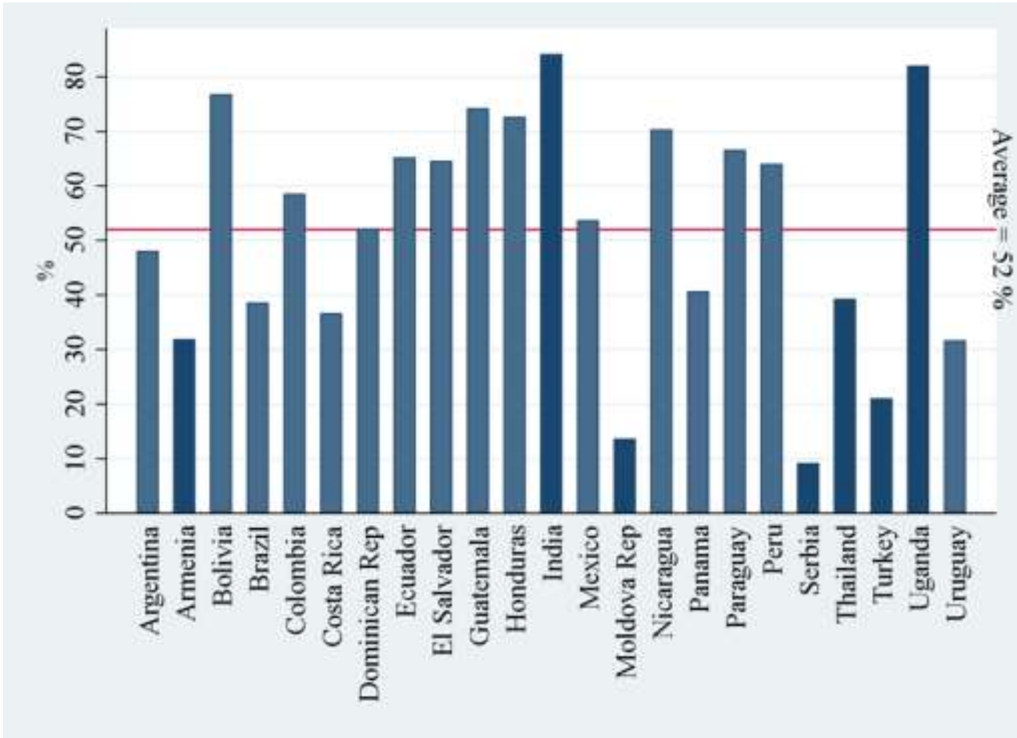
The International Labour Organization (ILO) refers to the informal economy as "all economic activities by workers and economic units that are –in law or in practice– not covered or insufficiently covered by formal arrangements" (ILO, 2002). Accordingly, the informal economy emerges as a result of two kinds of economic activities: (a) economic activities in the informal sector and (b) informal economic activities in the formal sector (outside the informal sector). Concerning the first group of activities, the 15th International Conference of Labour Statisticians (ICLS) Resolution states that the informal sector is "…characterized as consisting of units engaged in the production of goods or services with the primary objective of generating employment and income to the persons concerned. These units typically operate at a low level of organization, with little or no division between labor and capital as factors of production and on a small scale. "Labor relations –where they exist– are based mostly on casual employment, kinship or personal and social relations rather than contractual arrangements with formal guarantees." (ILO, 1993). Regarding the second group of activities, named "informal economic activities outside the informal sector" and also known as other forms of informality, we may say that "[…] although they are operating within the formal reach of the law, the law is not applied or not enforced; or the law discourages compliance because it is inappropriate, burdensome, or imposes excessive costs." (ILO, 2002).

On the other hand, employment in the informal economy comprises two components: (a) employment in the informal sector and (b) informal employment outside the informal sector (ILO, 2013). Informal employment[2] "encompasses persons in employment who, by law or in practice, are not subject to national labor legislation and income tax or entitled to social protection and employment benefits. Informal employment can exist in both the informal and the formal sector of the economy." (ILO, 2013, p. 4). "Employment in the informal sector and informal employment are concepts, which refer to different aspects of the 'informalization' of employment […]" (Hussmanns, 2005), and the difference between these two concepts is a consequence of the existence of informal employment outside the informal sector (ILO, 2013). Therefore, informal employment outside the informal sector refers to informal jobs in the formal sector.

Despite the implementation of economic policies aimed at reducing poverty and transforming traditional economies into dynamic and modern economies, informal employment accounts for a major proportion of employment for many poor and developing countries around the world. In Latin America, informal employment rates are particularly high in Bolivia, Guatemala, and Honduras, countries showing an average informality rate above 70 percent from 2008-to 2017 (see graph 1). Ecuador, El Salvador, Paraguay, and Peru report average informal employment rates above 60 percent, while Colombia, Dominican Republic, and Mexico show average rates above 50 percent. Why is this problem so persistent in developing countries? Is it the case that researchers, labor institutions, policymakers, and analysts have paid too much attention to characteristics and measurement and set aside the causes of the phenomenon?

[2] The 17th International Conference of Labour Statisticians defined informal employment as the total number of informal jobs, whether carried out in formal sector enterprises, informal sector enterprises, or households, during a given reference period (Hussmanns, 2005).

Graph 1 Informal Employment Rate-Developing Countries Country Average (2008-2017)



Source: Own Calculations based on ILOSTAT Database.[3] *Averages are computed with the data available for each country using STATA software 16.1*

The ILO definitions of informal economy and informal employment provide very clear concepts that help us understand the characteristics of informal economic activities and informal jobs. Indeed, the ILO definitions of informality might have their foundations in understanding the phenomenon; however, they were mainly designed to meet measurement objectives. Measuring and generating a system of statistics and a database for information on informal economy and informal employment is a key task for macroeconomic analysis, policy formulation, and evaluation. Additionally, informality measurement is essential for "…the formulation and implementation of policies for economic and social development, including employment creation, production, income generation, human capital formation and the mobilization of financial resources;" (ILO, 2013, p. 6). There is no doubt about the importance of having access to a definition of economic informality based on a statistical approach. Such a definition must be internationally accepted and may allow us to measure and carry out cross-country comparisons of this phenomenon. However, a measurement-oriented definition might not help us analyze the economic causes and roots of the problem.

To provide an economic definition of informal economy, we follow Guillermo & Angulo (2016), which presents a new approach called *non-retributed factors approach*. Under this approach, we consider informal economy all those economic activities in which at least one factor of production is non-retributed or under-retributed. This definition makes a critical difference from other general conceptualizations of the informal economy; it captures an essential part of the problem: non-retribution or under-retribution of production factors, characteristics which, as will be explained, are related to scarcity. In Mexico and other Latin American countries like Bolivia, Colombia, Nicaragua, etc., *non-retribution* is most frequently observed in the case of payment to physical capital. In contrast, the under-retribution characteristic is mainly observed in the case of payment to human capital. The scarcity of physical capital –with the corresponding high price of this factor– generates an invasion of public spaces, spaces that are essential for most of the informal economic units to carry out the production process of goods and services. But the invasion of public spaces is not restricted to informal economic units only. During the last decade, the appropriation of public spaces by formal economic units in Mexico has increased and become very common every day.

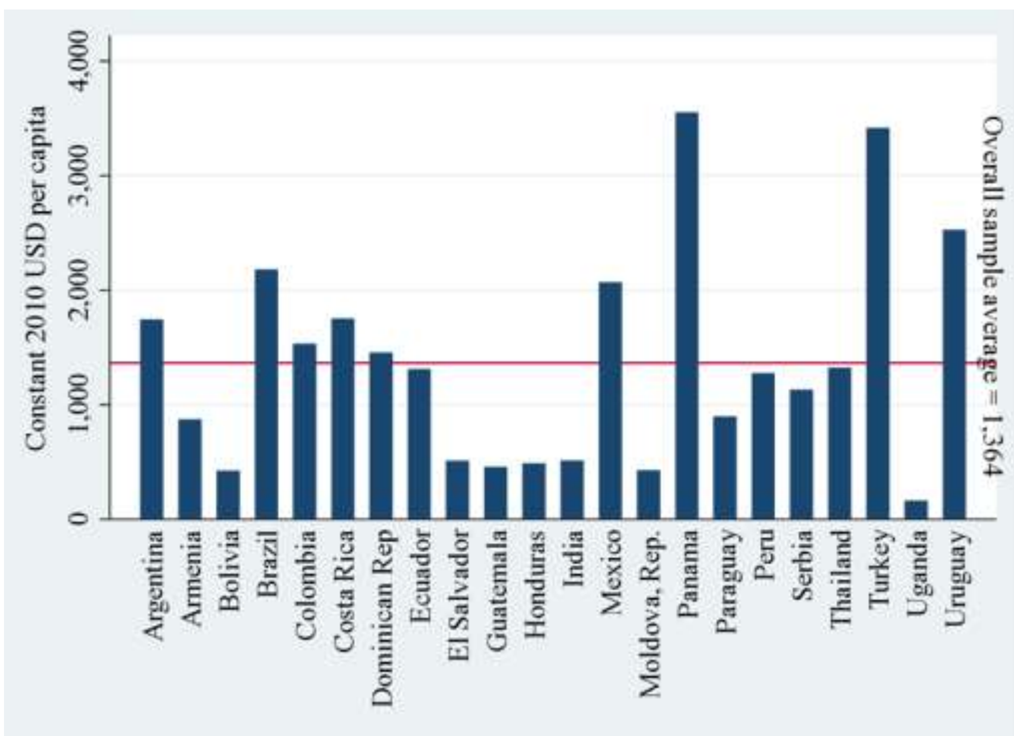
[3]Available at https://ilostat.ilo.org/topics/informality/

As an example, formal businesses –usually small– without enough customer parking space or enough space for tables and seats use public spaces to carry out their activities. We refer here to those formal businesses that use the surrounding streets and sidewalks to expand their grounds to provide parking or a dining table for their clients that otherwise would not go to those businesses.

The scarcity of physical capital in developing countries is evident when comparing the data with developed countries. Although we do not have data on capital stock, we may compare the gross fixed capital formation per capita for developing countries with that of developed countries. Graphs 2 and 3 show this variable's 2008-2017 country average for developing and developed countries, respectively. Comparing the overall sample average, we may observe that gross fixed capital formation per capita in developed countries is 6.5 times as much as in developing countries. Human capital is also very scarce in developing countries relative to developed countries. Graphs 4 and 5 show the 2008-2017 average educational attainment rates for developing and developed countries, respectively. The overall sample average educational attainment rate for developed countries is almost twice (1.8 times) the rate in developed countries.

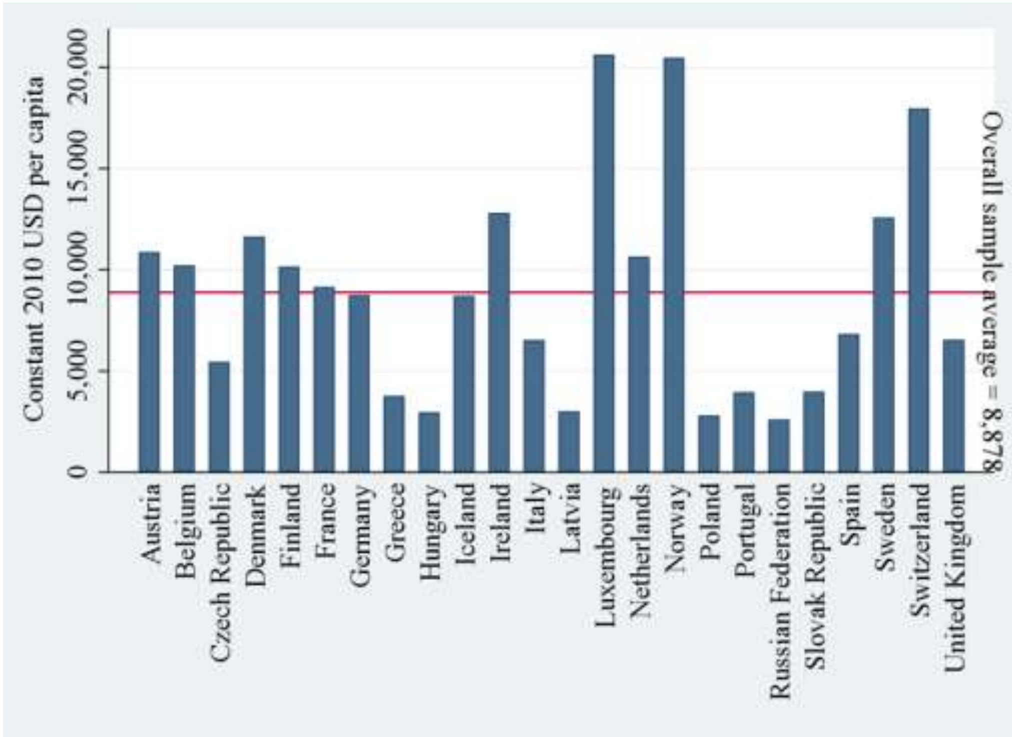
Education has been widely perceived as one of the crucial determinants of an individual's decision to participate in the informal economy. Education is also a very important characteristic of a country´s population related to other sociodemographic characteristics that might help us to describe and identify differences between formal and informal workers. As shown in section 3, education and physical capital are key factors influencing the informal employment rate. In particular, the empirical evidence provided by the estimated model shows that increasing human capital will reduce the informal employment rate, and the effect is stronger in developing countries.

Graph 2 Gross Fixed Capital Formation Per Capita- Developing Countries Constant 2010 USD (2008-2017 average by country)



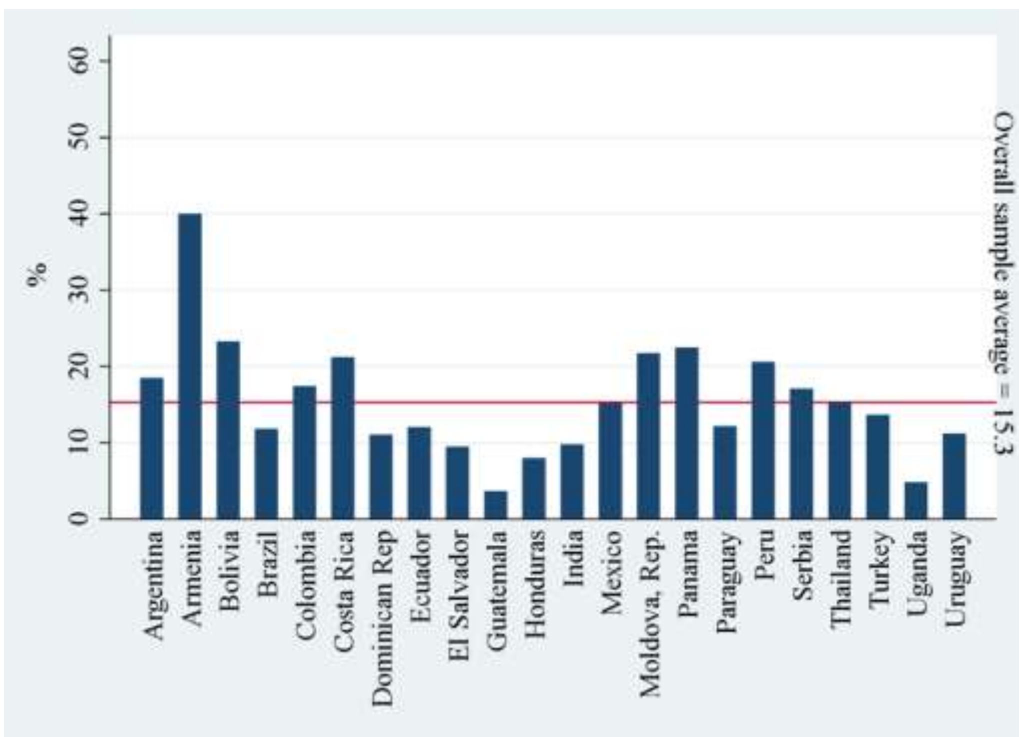
Source: Own Elaboration using STATA software 16.1 and based on data from The World Bank, World Development Indicators

Graph 3 Gross Fixed Capital Formation Per Capita- Developed Countries Constant 2010 USD (2008-2017 average by country)



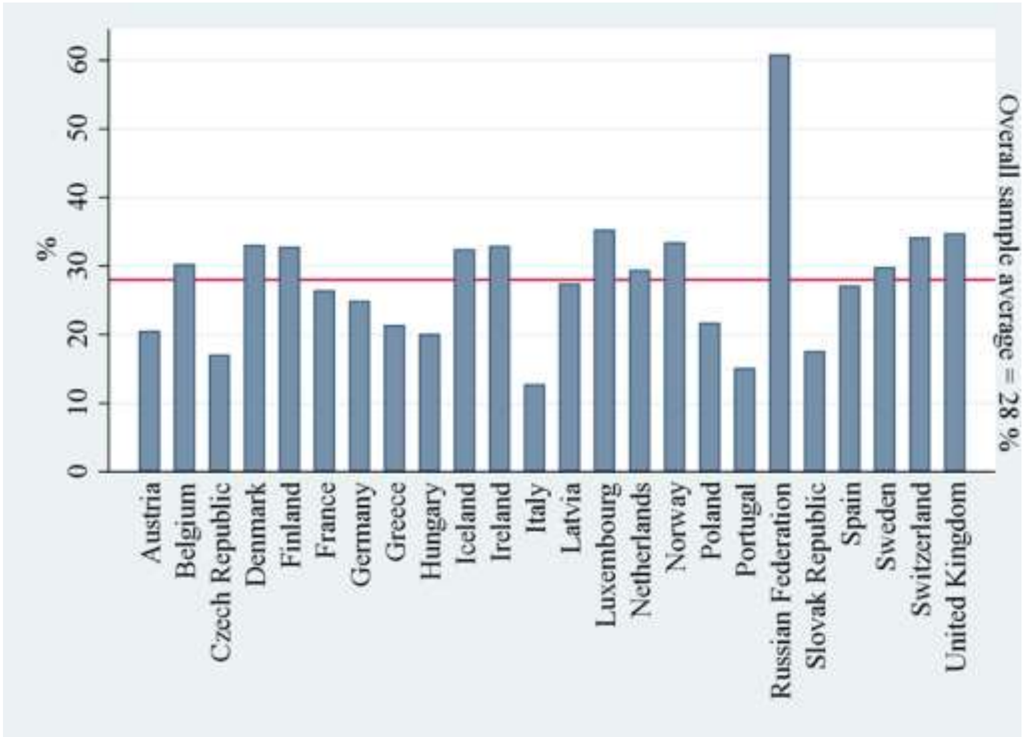
Source: Own Elaboration using STATA software 16.1 and based on data from The World Bank, World Development Indicators

Graph 4 Educational Attainment – Developing Countries (2008-2016 average by country) % of Population 25+years that at least completed short-cycle tertiary education



Source: own elaboration using STATA software 16.1 and based on data from UNESCO

Graph 5 Educational Attainment – **Developed Countries** (2008-2016 average by country) % of Population 25+years that at least completed short-cycle tertiary education



Source: Own Elaboration using STATA software 16.1 and based on data from UNESCO

2.1 Informal Economy and some Sociodemographic characteristics of Employed Workers in Mexico

This section describes some sociodemographic characteristics of employed workers in Mexico. Why analyze this for Mexico? Because of data accessibility, special interest in our country, and very likely, employed workers in other Latin American countries have similar sociodemographic characteristics to Mexican workers (we keep this last issue for further research).

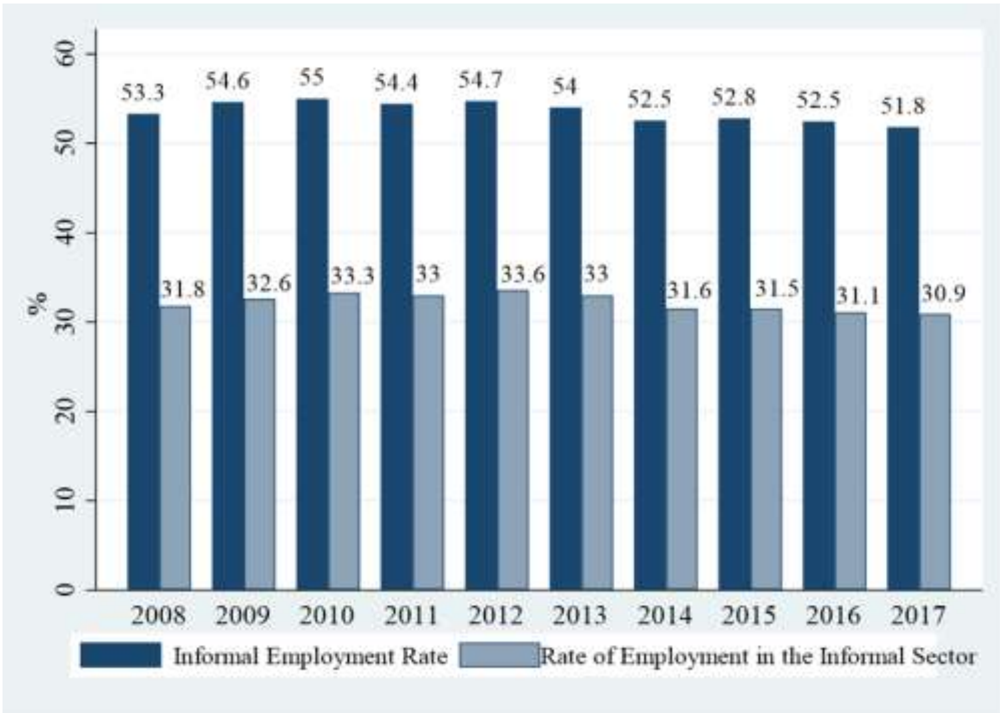
The objective is to review some basic sociodemographic indicators of employed workers and identify the differences between informal and formal employment workers. First of all, it is important to point out that more than half of the non-agricultural working population in Mexico has an informal job (see Graph 6). On the other hand, the informal sector's employment rate is close to 30 percent considering non-agricultural activities. This fact tells us that the informal sector is an essential source of employment in Mexico and that half of the informal jobs are outside of the informal sector, accounting for a substantial proportion of employment in this country. Therefore, informal employment is the largest component of the workforce.

In Mexico and many Latin American countries, the informal economy is not only the largest source of employment but also has essential participation in the production of goods and services. The contribution of the informal economy (inside and outside the informal sector) is substantial. Graph 7 shows that, on average, during the period 2003-2017, the informal economy's contribution to Mexico's GDP has been around 26 percent, reflecting the importance of informal economic activities in income generation and hence poverty alleviation.

To identify specific sociodemographic characteristics of informal workers, table 1 compares some basic indicators of employed workers (aged 15 years or older) by type of employment: informal or formal. During the analyzed period (2008-2017), the sociodemographic characteristics of employed informal workers have not significantly changed. The particularities of informal workers that mark a difference from formal workers are the following: First, we see in table 1 that among the workers with informal employment, 40.3 percent live in urban localities, while 67 percent of formal workers live in urban areas. The distribution of workers by sex is very similar for both informal and formal workers since women represent 38.2 and 36.6 percent of informal and formal employed workers, respectively.

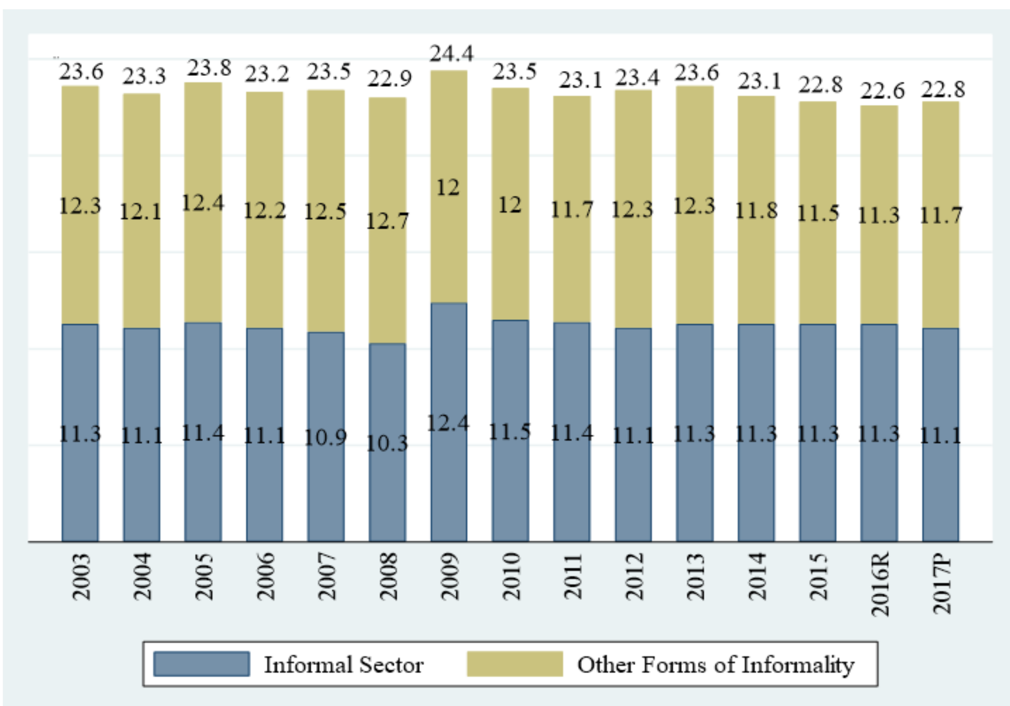
The data tell us that women and men are almost equally likely to be engaged in informal economic activities or informal employment. One particularity of informal workers is their age distribution. Informal workers have greater proportions of young (aged 15 to 24) and elderly (aged 65 and older) relative to formal workers. In particular, 23 percent of informal workers –on average– are school-aged workers, while only 14 percent of formal workers –on average– are under this condition. Regarding marital status, differences are practically indistinguishable. Also, the proportion of informal workers being the head of household is around six percentage points lower than that of the formal workers.

Graph 6 Mexico: Informal Employment and Employment in the Informal Sector (Share of working population in Non-agricultural activities)



Source: Own Elaboration using STATA software 16.1 and with data from the National Survey of Occupation and Employment (ENOE). INEGI, Strategic Indicators 2008-2017 (Q2). Rates are calculated as the share of the working population in non-agricultural activities.

Graph 7 Mexico: Contribution of Informal Economy (% of GDP)



Source: Own elaboration with data from INEGI. R: Revised estimate; P: Preliminary
https://www.inegi.org.mx/temas/pibmed/default.html#Informacion_general (Abril, 2022)

Table 1 México: Sociodemographic characteristics of employed workers by type of employment 2008-2017 (workers 12 years and older)

Indicator	Type of employment	Year									
		2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
% living in urban localities [a]	Informal	40.3	40.2	39.9	39.9	39.9	39.7	38.9	38.9	38.3	38.4
	Formal	67.2	67.3	67.4	66.5	66.7	66.6	65.7	65.7	66.0	65.9
% of women	Informal	38.2	37.4	38	30.8	39	39.1	38.2	38.1	38.4	38.0
	Formal	36.6	37.4	37.2	36.8	37.2	37.6	37.3	37.1	37.7	37.8
% aged 12 to 24 years	Informal	23.7	23	23.6	22.9	22.6	21.8	22.6	21.7	21.4	21.0
	Formal	15.3	13.5	14.0	13.9	14.0	13.6	13.2	12.7	12.7	12.8
% aged 65 years and older	Informal	6.3	6.3	6.4	6.4	6.3	6.4	6.5	6.8	6.8	6.9
	Formal	2.5	2.3	2.4	2.3	2.3	2.2	2.5	2.5	2.5	2.7
% married or living with partner	Informal	60.7	60.6	60.3	60.5	60.4	61.1	60.0	60.0	60.4	60.5
	Formal	64.5	64.4	64.2	64.3	63.8	63.9	65.7	64.7	63.5	63.7
% single	Informal	30.6	30.6	30.7	30.3	30	29.9	30.8	30.7	30.1	29.9
	Formal	28.0	27.6	27.7	28.1	28.3	28.2	26.5	27.2	28.0	27.9
% head of household	Informal	45.1	45.9	45.6	45.9	45.6	45.6	44.9	44.9	45.5	45.4
	Formal	52.3	52.7	52.0	52.3	51.4	51.3	51.5	51.5	51.0	50.2
% working children of the household head	Informal	28.6	28.5	28.1	27.5	27.5	27.3	28.2	28.2	27.5	27.6
	Formal	24.4	24.1	24.8	24.5	25.0	24.7	23.8	24.2	24.6	24.8

Source: Own Elaboration with data from the National Survey of Occupation and Employment (ENOE). INEGI, 2008-2017
[a]Urban localities with more than 100 thousand inhabitants

Concerning education, the differences between informal and formal workers are an issue to highlight. Table 2 shows that although illiteracy rates are low, the proportion of informal workers considered illiterate is seven times higher –on average– than that of formal workers.

Table 2 México: Selected indicators of Education for employed workers, 2008-2017 (workers 12 years and older)

Indicator	Type of Employment	Year									
		2008	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**
% of illiterate	Informal	8.7	8.0	7.9	7.7	7.0	6.6	5.9	5.6	5.4	5.3
	Formal	1.4	1.1	1.2	1.0	0.9	0.9	0.7	0.6	0.6	0.6
% with primary school (complete or incomplete)	Informal	40.7	39.4	38.5	37.2	36.2	35.5	35.3	34.8	33.9	32.7
	Formal	16.3	14.7	13.8	13.5	12.6	11.7	12.0	11.7	10.9	10.8
% with middle school (complete or incomplete)	Informal	28.5	28.9	29.2	30.1	30.7	30.8	31.1	32.0	32.1	32.4
	Formal	25.0	24.0	24.5	24.4	24.3	24.3	24.8	24.9	24.5	24.1
% with high school or post-secondary (non-tertiary) education (complete or incomplete)	Informal	15.3	16.2	17.0	17.6	18.1	18.5	18.6	18.5	19.1	11.6
	Formal	29.1	29.9	29.9	30.2	29.9	30.3	29.1	28.7	29.7	29.2
% at least completed short-cycle tertiary education	Informal	6.8	7.5	7.4	7.5	8.0	8.5	8.3	8.4	8.8	9.1
	Formal	28.1	30.1	30.6	30.9	32.1	32.7	32.9	33.8	34.1	35.0
% workers with school attendance	Informal	8.6	8.2	8.3	8.0	8.6	8.1	7.8	7.6	7.5	7.5
	Formal	4.8	4.4	4.5	4.5	5.1	4.9	4.5	4.4	4.4	4.2

Source: Own Elaboration with data from the National Survey of Occupation and Employment (ENOE). INEGI, 2008-2017

We observe an important proportion of informal workers with educational attainment below the middle (secondary) school compared to formal workers. In general, the educational gap between informal and formal workers gets wider the higher the educational level. For example, the gap between informal and formal workers reaches 24 percentage points on average for employed workers with- some- tertiary educational attainment. This gap shows an increasing trend during the period 2008-2017. In particular, for 2017, the percentage of informal workers that attained or completed some tertiary level of education (short-cycle tertiary, bachelor´s or equivalent, master's or doctoral degree) was only 9.1 percent, while for formal workers was 35 percent. This fact provides some evidence for our hypothesis that human capital is an essential factor in explaining and reducing informal employment.

Another aspect that must be highlighted in this regard is that the proportion of informal workers combining school with work is almost twice that of formal workers in this situation. On average, 8 percent of informal workers attend school, while only 4.6 percent of formal workers do so.

3. The Model and Estimation Methodology

This section aims to show some empirical evidence regarding the influence of physical capital and human capital on informal employment rates and the influence of other macroeconomic variables like GDP growth and corruption. The majority of studies that explore the relationship between education and informality are mainly focused on how an individual´s decision to participate in the informal economy is influenced by the level of education (Jimenez *et al.*, 2015; Angel-Urdinola & Tanabe, 2012; Günsel, 2012; Yamasaki, 2012; Gërxhani & Van de Werfhost, 2011; Bucheli & Ceni, 2010; Pisani & Pagan, 2004). Using microdata for specific countries, the studies in the literature show evidence of a negative relationship between education and informal economy participation.

Besides physical and human capital, other factors may influence the informal employment rate; omitting them from the model may yield biased and inconsistent parameter estimators. This is the case of the GDP per capita growth rate, an explanatory variable that must be included to capture the impact of the economic expansion (and economic development over time) on the informality rate. Although there has been a widespread assumption of a negative relationship between economic growth and the informal economy, the empirical evidence is unclear. On the one hand, some studies have shown a negative relationship between the size of the informal economy and the level of per capita GDP. For example, La Porta (2008) uses different proxies for the informal economy's size on data for 96 countries to show that the size of the informal economy declines as one moves from poor to rich countries. Loayza & Rigolini found similar evidence (2006); these authors showed that, in the long run, informality (measured as the self-employment rate) is larger in countries that have lower GDP per capita. Galli & Kucera (2003) also explored the relationship between the share of informal employment and the GDP. Using panel data for fourteen Latin American countries in the 1990-1997 period, these authors found evidence of a negative GDP elasticity of informal employment share, indicating a countercyclical behavior of informal employment.

On the other hand, some empirical studies have shown that high informality rates can co-exist with high economic growth rates. According to Castells and Portes (1989, pp. 16-17), between 1950 and 1980, Latin American countries grew at a weighted average of 5.5 percent, while informal employment declined only from 46 percent to 42 percent of the Latin American labor force. Heintz & Polling (2003) found that increasing economic growth will reduce the rate at which informalization is increasing in developing countries. Still, economic growth does not produce an absolute decline in the informal employment rate. More recently, Loayza (2016) modeled, calibrated, and simulated the behavior of the informal sector. As a result, this author presents projections of the percentage of the informal labor force for several developing and developed countries and concludes that the TFP growth will lead to a reduction in informality in the long run. Perhaps this is why the influence of GDP growth on the informal employment rate is unclear. We must remember that the TFP is only one component of the GDP growth rate; hence, even though an economy might be experiencing an important positive GDP growth rate, its TFP growth rate may be zero or negative.

As we know, one of the main problems faced in research studies is data availability. Our research on informal employment is not an exception. Even for a particular country (developing or developed) is very difficult to have data with an appropriate time length to perform a time series analysis. In this sense, the data don´t allow one to set up a model to analyze the factors that influence informal employment and estimate the impact of those factors on the country´s informal employment rate. But data availability becomes even more problematic when we talk about developing countries. Because economic activities under informality and informal employment are particularly –but not exclusively– found in developing countries, the study of this phenomenon becomes complicated.

Considering data availability on the variables included in the model, we started setting a model based on a panel data set. We have data for 46 countries (22 developing countries, among which 15 are Latin American countries and 24 developed countries) with a time span from 2008 to 2013. However, the panel is not balanced because the number of time observations is different across countries; hence, the sample shows some missing observations for some countries. In particular, for some developing countries, we have the entire period of observations (six). In contrast, for some other developing countries, we have less than six observations (see table A1 in the Appendix for the list of countries included in the sample).

The limitation of data coverage across countries is an obvious disadvantage for our study for adequately identifying the factors that may influence the informal employment rate and the magnitude and direction of those influences. Moreover, the limitation of data coverage across countries may also translate into *small within-country variation*, data characteristics that must be considered when choosing the model estimation technique.

Based on the data characteristics, we initially set up the model to estimate the informal employment rate as a function of GDP per capita growth rate, Gross Fixed Capital Formation per capita, Educational Attainment (as a proxy of Human Capital), the Corruption Index (as a proxy of a country´s institutional functioning) and the number of days it takes to start a business (as a proxy of the cost to start a formal business). It should be mentioned here that capital stock data is unavailable, which is why we use Gross Fixed Capital Formation instead (change in Gross Fixed Capital Stock). Additionally, the effect of a country's development is captured by a binary variable taking the value one if the observation in question corresponds to a developed country and zero otherwise, and by the interaction terms of this binary variable with the other explanatory variables. The model's functional form was chosen after graphically exploring the relationship between each explanatory and the dependent variable. Therefore, the baseline model for our panel data set is expressed as follows:

$$\begin{aligned} infe_{it} = {} & \alpha_i + \beta_2 dc_{it} + \beta_3 gdpg\_pc_{it} + \beta_4 (dc_{it} \times gdpg\_pc_{it}) + \beta_5 lgfcf\_pc_{it-1} \\ & + \beta_6 (dc_{it} \times lgfcf\_pc_{it-1}) + \beta_7 educ_{it} + \beta_8 (dc_{it} \times educ_{it}) + \beta_9 ci_{it} \\ & + \beta_{10} (dc_{it} \times ci_{it}) + \beta_{11} tsb_{it} + \beta_{12} (dc_{it} \times tsb_{it}) + e_{it} \end{aligned} \tag{1}$$

Where:

$infe_{it} =$ Informal employment rate for country i at time t
dc_{it}= binary variable taking the value one if country i is a developed country and zero otherwise (time-invariant for the sample period).
$gdpg\_pc_{it} =$ GDP per capita growth (annual rate) for country i at time t
$lgfcf\_pc_{it-1} =$ Natural log of the Gross Fixed Capital Formation per capita for country i at time t-1
$educ_{it}$ = Educational Attainment for country i at time t. This variable is measured as the percentage of population 25+ years old that at least completed short-cycle tertiary education[4]
ci_{it} = Corruption Index for country i at time t.
tsb_{it}= Time to start a business (days) for country i at time t
e_{it} = Idiosyncratic error term

As we may observe, equation (1) includes interaction terms to capture differences in coefficients between developing and developed countries. The commonly used and suitable model set up for short and wide panel data sets specifies the term α_i to capture all the *unobserved heterogeneity* across countries. Under the fixed effects approach, the term α_i is a country-specific constant term in the regression model embodying some unobserved elements correlated with the explanatory variables. Under the random effects approach, however, the unobserved heterogeneity term α_i does not embody any elements correlated with the explanatory variables of the model (Geene, 2018). On the other hand, if α_i is constant for all countries –meaning that there are no behavioral differences across countries–the model estimation procedure reduces to pooled ordinary least squares (pooled OLS).

The appropriateness of each model or estimation technique strongly depends on the data characteristics. In particular, the appropriate estimation technique depends on the assumptions about the unobserved heterogeneity effects. As is well known in panel data econometrics, if α_i is uncorrelated with the explanatory variables, random effects will produce consistent estimators and efficiency gain over fixed effects estimation. However, if the effects of heterogeneity across countries α_i, are correlated with the explanatory variables, implying an endogeneity problem, only the fixed effects estimation technique will produce unbiased and consistent parameter estimates. In order to solve the estimation dilemma of random versus fixed effects estimation, the Hausman test provides a suitable methodology to find evidence of the correlation between the unobserved cross-country heterogeneity and the model's explanatory variables.

[4] UNESCO. http://data.uis.unesco.org/

Nonetheless, when carrying out empirical work, other *sample* characteristics of the data may also influence the choice of estimation technique. In this regard, we have previously mentioned an important limitation on our data coverage across countries and an unbalanced panel. These two characteristics of the sample, particularly for the variables we use in our model, translate into a very *small within-country variation* for each variable in the model. Hence, the question now is how the *small within-country variation* of the model's variables affects the choice of the estimation technique under a panel data context.

Following Hahn *et al.* (2011), "if the within variation is small, the fixed effect estimates may not be asymptotically normal, potentially invalidating the basic premise of the Hausman test." Consequently, the conventional Hausman test may not be reliable (ibid., 294). The authors provide a valid version of the Hausman test for between effects versus fixed effects[5]. However, we have not found econometric software[6] that provides a command to implement Hahn's valid version of the Hausman test, which requires a bootstrap algorithm to generate the corresponding valid critical values. Therefore, we discarded the use of the conventional Hausman test as a criterion for choosing the estimation procedure (fixed effects versus random effects).

Now, considering that our sample data shows little within-country variation, working with country average data (the so-called between-effects estimator) could be much better than fixed effects specification. This is so because the fixed effects transformation (transforming the variables in deviations with respect to their country means) restricts sample information –for estimation purposes– to within-country action only. That is, fixed effects models rely on within-group variation, which is why we need a reasonable amount of variation of key explanatory variables within each group (Dranove, 2012). Thus, data are not sufficiently rich in information if we have very small within-country action due to the nature of the variables and short time series length. In such a case, the fixed effects model will produce poor and unreliable estimates because most of the variation in the model will come from across-country variation. Fixed effects estimation washes out all across-country variation, which explains why fixed effects estimates will be poor and unreliable when having little within-group variation. To summarize this issue, we must point out that an important limitation of fixed effects models is that we cannot assess the effect of those variables that have small within-group variation (ibid), which is the case in our data set.

Once we have explained the problems associated with fixed effects using small within variation conditions, we must consider other estimation procedures. Is it the random effects option? We must consider that random effects estimation will produce inconsistent estimates if the unobserved country-heterogeneity is correlated with the explanatory variables; however, the conventional Hausman test to discard this possibility in our model cannot be implemented. Additionally, we must consider that random effects estimation through the generalized least squares procedure transforms the data by partially demeaning each variable. That is, instead of subtracting the entire country-specific mean (for each variable), only a fraction θ of the mean is subtracted, so we estimate the model with *quasi-demeaned data* (Wooldridge, 2003). The estimated fraction used to partially demean each variable is between zero and one and is a function of the number of time-series observations T, the estimated variance of the idiosyncratic error component $\hat{\sigma}_e^2$ showed in equation (1) and the estimated variance of the individual (country) error component[7] $\hat{\sigma}_u^2$.

In particular, in the case of unbalanced panels, the fraction $\hat{\theta}_i$ is computed for each cross-section unit (e.g., country-specific) in the following way (Wooldridge, 2010)[8].

$$\hat{\theta}_i = 1 - \sqrt{\frac{\hat{\sigma}_e^2}{T_i\,\hat{\sigma}_u^2 + \hat{\sigma}_e^2}} \tag{2}$$

[5] The authors also show that a version of the bootstrap provides valid critical values for this test.

[6] Developing a program (for example with STATA) to implement Hahn´s (2011) version of the Hausman test with the corresponding valid critical values, goes beyond the scope of this paper.

[7] Recall that, under the random effects model, the individual (country) heterogeneity term α_i is defined as $\alpha_i = \bar{\alpha} + u_i$, where $\bar{\alpha}$ is the population average (common intercept) and the $u_i{'}s$ are the unobserved random individual differences from the population average (Hill *et all* (2011)). Therefore, $\hat{\sigma}_u^2$ is the estimated variance of the unobserved individual heterogeneity term α_i.

[8] See also Stata 13. Longitudinal-Data/ Panel-Data Reference Manual, pp. 384

From this expression (2), we can easily observe that if cross-sectional variation $\hat{\sigma}_u^2$ explains almost all model variation given by $\hat{\sigma}_u^2 + \hat{\sigma}_e^2$, then $\hat{\theta}_i$ will be close to one. In this case, random effects and fixed effects estimates will be very similar[9] (Wooldridge, 2003, pp. 471).

But why is this explanation relevant for our data and model estimation choice? The explanation takes relevance when we again consider the small within-country variation of our data set. Under such conditions, we may expect values for $\hat{\theta}_i$ very close to one (and the median of the $\hat{\theta}_i$´s close to one, too), implying that fixed and random effects estimates will be similar. The problem with this expected similarity between fixed and random effects estimates is that –as we have previously explained– fixed-effects estimates are poor and unreliable when there is *small within variation*.

4. Estimation Results

Given the small variation of variables in the data set, the first estimation attempt for the model specified in equation (1) was under the random effects approach. As expected, the estimation results show a very high fraction of total variation due to cross-sectional variation as $\hat{\sigma}_u^2$ was 0.9738, meaning that almost all variation in the model is explained by cross-country variation. As a consequence, the estimation results also report the distribution[10] of $\hat{\theta}_i$ with a median equal to 0.9333, which is a value very close to one. This result implies that random and fixed effects estimates are very similar; hence, the random-effects estimates are as poor as those we get with the fixed-effects model. We conclude that random and fixed effects models are not suitable estimation techniques for our data set.

Considering all mentioned characteristics of our data set and all the caveats associated with alternative estimation procedures under such characteristics, we estimated the model using two approaches. The first one was generalized least squares on the pooled data set, implying that the α_i coefficient is constant across countries. The estimated model can be expressed as:

$$infe_{it} = \alpha + \boldsymbol{x}_{it}\boldsymbol{\beta} + e_{it} \tag{3}$$

Where $\boldsymbol{x}_{it}$ is the raw vector of all the eleven explanatory variables shown in equation (1) including interaction variables, and $\boldsymbol{\beta}$ is the (11×1) column vector of coefficients.

The feasible generalized least squares (FGLS) procedure was necessary to account for the presence of heteroskedasticity. A heteroskedastic partition was found associated with the country development effect dc_{it}. The estimated residuals for developing countries showed greater dispersion than those for developed countries.

The Breusch-Pagan (Lagrange Multiplier) and Goldfeld-Quandt tests were carried out to test the null hypothesis of homoscedasticity. Both tests provided evidence to reject the null hypothesis, as the sample values of the test statistics were 38.43 and 14.74, respectively, with p-values of zero in both cases. Therefore, FGLS estimation was implemented assuming the following variance function[11].

$$\sigma_{it}^2 = \begin{cases} \sigma_1^2 & \text{if } dc_{it} = 0 \\ \sigma_2^2 & \text{if } dc_{it} = 1 \end{cases} \tag{4}$$

---

[9] Recall that the fixed effects estimation procedure transforms the observations into time-demined data (deviations with respect to group means). This is the reason why the fixed effects transformation is also called within transformation (see Wooldridge 2002 and Wooldridge 2003 for a comprehensive explanation of the fixed effects and random effects models).

[10] Summary of the sample distribution of $\hat{\theta}_i$ (factor used to partially demean observations under the random effects model):

$\hat{\theta}_i$				
min	5%	median	95%	max
0.8850	0.8850	0.9333	0.9333	0.9333

[11] The Breusch-Pagan statistic has a χ^2_{p-1} distribution where p is the number of parameters included in the auxiliary regression, which were two in this case (as dc_{it} was the only one explanatory variable). On the other hand, the Goldfeld-Quandt statistic obtained from the estimated partitioned regression into two subsamples (one for developing countries and the other for developed countries) has an $F_{\upsilon_1,\upsilon_2}$ distribution where $\upsilon_1 = 91$ and $\upsilon_2 = 138$ degrees of freedom respectively.

Collinearity was another problem found during the estimation procedure. In particular, the Gross Fixed Capital Formation per capita is highly linearly correlated with the corruption index (r = 0.86). Besides obtaining wide standard errors, the consequence of this very close linear relationship between two explanatory variables in the model is that we cannot properly isolate the individual effect that each explanatory variable has on the dependent variable. To deal with this problem –and because we are particularly interested in estimating the effect of fixed capital and human capital on the informal employment rate[12] – we follow the sequential regression method proposed by Graham (1997, 2003) and also suggested by Dorman *et al* (2013). With this method, we can isolate the effect of the corruption index on the informal rate from that of the gross fixed capital formation. Therefore, be careful when interpreting the meaning of the estimated coefficient for the corruption index[13].

The estimation results under the pooled FGLS approach[14] are presented in table 3. We may observe that, in general, the effects of the explanatory variables on the informal employment rate are bigger for developing countries (e.g. magnitude of coefficients is bigger for developing countries). Also, except for the GDP per capita growth rate in developing countries and the time to start a business variable, all coefficients are statistically significant (at 1% and 5% significance levels). While informal employment in developing countries is not affected by the GDP growth rate, for developed countries, a one-percentage-point increase in the GDP per capita growth rate will reduce the informal employment rate –vulnerable employment– by 0.7 percentage points, on average.

Our estimation results show that, taking into account the logarithmic transformation of the gross fixed capital formation per capita, a one-percentage-point increase that took place in the previous year (*t-1*) in this variable will (on average) decrease the informal employment rate by approximately 0.15 and 0.013 percentage-points (at time *t*) in developing and developed countries respectively. The linear-log relationship between the informal employment rate and the gross fixed capital formation also implies that the marginal effect of the physical capital on the informality rate is smaller at higher levels of gross fixed capital formation[15]. For example, for the 2008-2013 period, the gross fixed capital formation, on average, was 225 and 1,826 (constant 2005) USD per capita in Bolivia and México, respectively[16]. With these gross fixed capital formation levels, our results estimate that a 100 dollar increase in this variable in year *t-1* will decrease the informal employment rate in year *t* by 6.7 percentage points in Bolivia, while in Mexico, informal employment will only fall by 0.82 percentage points. This is so because a 100 dollars increase in gross fixed capital formation per capita means increasing this variable by 44.4% in Bolivia (which is a considerable increase in capital formation!); while for Mexico, increasing 100 dollars the gross fixed capital formation per capita, only represents a 5.5% increase in this variable.

---

[12] Eliminating the corruption index from the model must not be considered as a choice to solve the collinearity problem, because this will generate bias due to omission of a relevant variable. Omitting the corruption index as explanatory variable will make the coefficient of the gross fixed capital formation to explain part of the effect of the omitted one. Inferences will be incorrect and prediction may be compromised (Dorman *et al*, 2013).

[13] Sequential regression is a method that creates new explanatory variables in the sense that they have been *purified* from the effect of other explanatory variables. That is we extract the unique contribution of an explanatory variable, from its shared contribution with other explanatory variables. When two explanatory variables are collinear, "This can be done by regressing the less important variable against the other, and replacing the less important variable with the residuals from the regression […]" (Graham, 2003 pp. 2810). In our model, we consider the corruption index as the less important variable which is highly collinear with the gross fixed capital formation. Hence, the first step in applying the sequential regression method was to regress the corruption index against the logarithm of gross fixed capital formation per capita, and obtain the estimated residuals from this regression. This residuals –by construction– are orthogonal to the log of gross fixed capital formation per capita. In this sense, they represent that part of the corruption index which is not explained by the gross fixed capital formation. As second step, we use the estimated residuals (which we named *ehat_ci*) as explanatory variable in our pooled FGLS regression.

[14] Because the model in equation (1) allows for different coefficients between developing and developed countries, and there is a heteroscedastic partition due to development condition (dc), then estimation with FGLS is actually estimating two separate regressions (developing and developed countries).

[15] Recall that if two variables y and x have a linear-log relationship like $y = \beta_1 + \beta_2 ln(x)$, the slope of this function will be $\beta_2 \frac{1}{x}$. This is the marginal effect of x on y.

[16] Own calculations based on World Bank data.

Table 3 Estimation Results, Feasible Generalized Least Squares on Pooled Data

Dependent Variable: $infe_{it}$			
Variable	Coefficient	Std. Error	P-val
_cons	162.9782***	12.8534	0.000
dc_{it}	-135.0791***	15.2729	0.000
$gdpg\_pc_{it}$	0.4208	0.3060	0.170
$dc_{it} \times gdpg\_pc_{it}$	-0.6982**	0.3427	0.043
$lgfcf\_pc_{it-1}$	-15.0382***	1.9843	0.000
$dc_{it} \times lgfc\_pc_{it-1}$	13.7366***	2.1872	0.000
$educ_{it}$	-1.1246***	0.1672	0.000
$dc_{it} \times educ_{it}$	0.8897***	0.1752	0.000
$ehat\_ci_{it}$	-0.5786***	0.1215	0.000
$dc_{it} \times ehat\_ci_{it}$	0.4503***	0.1313	0.001
tsb_{it}	0.0533	0.0461	0.249
$dc_{it} \times tsb_{it}$	-0.0289	0.0690	0.675
Number of obs = 241			
F(11, 229) = 104.69			
Prob > F = 0.0000			
Adj R-squared = 0.8262			

Source: Own estimations. (***), (**) statistically significant coefficient at 1 and 5 percent respectively

Regarding the effect of education, our estimation results show that a one-percentage-point increase in the educational attainment rate will bring a 1.125 and 0.24 percentage-points fall in the informal employment rate in developing and developed countries, respectively. Therefore, based on our results, increasing human capital has a more considerable effect on reducing informality than increasing physical capital. This result particularly makes sense for developing countries. Why? Because even if the increase in capital investment takes place in the formal sector, an important proportion of the newly hired workers will likely be employees holding informal jobs; that is, workers not covered by social security and not entitled to other formal employment benefits. A poor labor regulation system, lack of law enforcement, and corruption in general are –unfortunate and prevailing–conditions that create incentives for informal jobs in formal enterprises. On the other hand, for developed countries, the increase in physical capital investment may go to own-account workers, which are classified as vulnerable employment workers.

Concerning the impact of corruption, we might expect a unit increase in the corruption index, –implying that a country becomes less corrupt– will bring a 0.58 and 0.128 percentage-point fall in the informal employment rate in developing and developed countries, respectively. As previously explained, this is the unique –independent from gross fixed capital formation– influence of corruption on informality. Regarding the *time to start a business* variable that we included to capture the time-costs of opening a formal business, we can see that the corresponding coefficients have the expected sign; however, the variable has no significant impact on the informal employment rate. Finally, the estimated intercepts are 162.98 and 27.9 for developing and developed countries, respectively.

The second estimation approach implemented was the between-effects model, which estimates regression on country-average data. Although this model is not commonly used, the small within variation of our data set suggests that this approach may be a suitable estimation technique in this case. The between-effects estimation results are shown in table A2 in the Appendix. In particular, the estimated coefficients associated with the (log of) gross fixed capital formation, educational attainment, and corruption index (*ehat_ci*) are very similar to those obtained with FGLS on pooled data. These results support the unbiasedness and reliability of the estimated coefficients obtained under the first estimation approach. However, the GDP per capita growth rate coefficients are very different in magnitude (and significance for developing countries) compared to those of the first estimation approach. The fact that the GDP per capita growth rate is the explanatory variable that shows more within-country variation may explain the difference between the alternative estimation approaches.

5. Final Remarks and Conclusions

The data shown tell us that informal employment accounts for a major proportion of employment for many poor and developing countries worldwide. Taking the *non-retributed factors approach,* we have explained that, economically, the source of the persistent presence of informal economy and informal employment in developing countries (Latin American countries in particular) is the scarcity of physical capital and the scarcity of human capital. The empirical evidence presented by estimating a model to explain the informal employment rate by country showed that human capital and physical capital are the main factors explaining informal employment. Together with the corruption index and the GDP growth rate, these variables explain 82 percent of the informal employment rate movements around its mean. Except for the "time to start a business" variable, the magnitude of marginal effects of each explanatory variable on the informal employment rate is different for developing countries compared to developed ones.

Regarding the econometric methodology used, we must point out that for panel data sets showing small within-group variation, the estimation approach must be carefully chosen to avoid low reliability and biasness. The application shown in this paper tells us that the little within-country variation, implies that the data are not sufficiently rich in information; therefore, a fixed-effects approach under these conditions is not the appropriate estimation technique. When we have data with small variation over time, the fixed effects transformation of variables in deviations with respect to their group means washes out all across-country variation, and estimates will be poor and unreliable.

Additionally, the fixed effects model with small within-variation conditions has another caveat. Following Hahn *et al.* (op. cit), the fixed effects estimates may not be asymptotically normal. Consequently, the conventional Hausman test used to show evidence of endogeneity of the individual (country) effects may not be reliable.

6. References

Angel-Urdinola D. & Tanabe, K. (2012). Micro-Determinants of Informal Employment in The Middle East and North Africa Region. SP Discussion Paper 1201. Social Protection and Labor. The World Bank. January 2012. https://openknowledge.worldbank.org/handle/10986/26828

Bucheli, M. & Ceni, R. (2010). "Informality Sectoral selection and Earnings in Uruguay". Estudios Económicos, vol. 25, núm. 2, julio-diciembre 2010, pp. 281-307 https://www.redalyc.org/pdf/597/59716168001.pdf

Castells, M. & Portes, A. (1989). World Untherneath: The origins, Dynamics, and Effects of the Informal Economy. In, The Informal Economy: Studies in adavance and less developing countries. Portes, A. (ed). John Hopkins University Press.

Dranove, D. (2012). "Practical Regression: Fixed Effects Models". Technical Note, Kellog School of Management, Northwestern University. https://www.studocu.com/en-us/document/northwestern-university/business-analytics-for-decision-makers/ktn-fixed-effects-practical-regression-from-stylized-factsto-benchmarking/1045173

Dorman C. *et al* (2013). "Collinearity: a review of methods to deal with it and a simulation study evaluating their performance". Ecography 36: 027–046, 2013 https://onlinelibrary.wiley.com/doi/full/10.1111/j.1600-0587.2012.07348.x Last time retrieved on May 25, 2022

Galli, R. & Kucera, D. (2003). "Informal employment in Latin America: Movements over business cycles and the effects of worker rights". Institute for Labour Studies. Decent Work Research Programme. Discussion Paper DP/145/2003. https://www.wiego.org/sites/default/files/publications/files/Kucera_Informal-Sector-Latin-America.pdf

Gërxhani, K. & Van de Werfhost, H.G. (2011). "The Effect of Education in Informal Sector Participation in a Post-Communist Country". European Sociological Review, Vol. 0, Num. 0, 2011 pp. 1–13 https://doi.org/10.1093/esr/jcr087

Graham, M. H. (1997). "Factors determining the upper limit of giant kelp, Macrocystis pyrifera Agardh, along the Monterey Peninsula, central California, USA". Journal of Experimental Marine Biology and Ecology 218, pp. 127–149. https://doi.org/10.1016/S0022-0981(97)00072-5

Graham, M. H. (2003). "Confronting Multicollinearity in Ecological Multiple Regression". Ecology, 84(11), 2003, pp. 2809–2815 https://www.jstor.org/stable/3449952?origin=JSTOR-pdf

Greene, W. (2018). Econometric Analysis. Eighth Edition. Pearson, New York.

Guillermo, S. & Angulo, J. (2016). "Enfoque alternativo sobre el concepto de la economía informal y análisis de la informalidad en México". In Guillermo Peón, Sylvia Beatriz (coordinadora) Ensayos sobre la Economía Informal, un análisis teórico y empírico con aplicación en Municipios de Puebla. Educación y Cultura, 2016 http://cmas.siu.buap.mx/portal_pprd/work/sites/dfe/resources/PDFContent/81/1.%20Econom%C3%ADa_informal_COLOR%20digital.pdf

Günsel D. (2012). "Determinants of Formal and Informal Sector Employment in the Urban Areas of Turkey". International Journal of Social Sciences and Humanity Studies. Vol. 4, No. 2, 2012 http://www.sobiad.org/eJOURNALS/journal_IJSS/arhieves/2012_2/h_gunsel_dogrul.pdf

Hahn J., Ham J. & Moon H.R. (2011). "Test of random versus fixed effects with small within variation". Economic Letters, 112, 293-297. https://doi.org/10.1016/j.econlet.2011.05.018

Heintz, J. & Pollin, R.. (2003). "Informalization, Economic Growth and the Challenge of Creating Viable Labor Standards in Developing Countries." Political Economy Research Institute. Working Papers Series, # 60, University of Massachusetts-Amherst. https://peri.umass.edu/fileadmin/pdf/working_papers/working_papers_51-100/WP60.pdf

Hill C., Griffiths W.E. and Lim G. (2011). Principles of Econometrics, Fourth Edition. Wiley

Hussmanns, R. (2005). "Measuring the informal economy: From employment in the informal sector to informal employment". International Labour Organization. Working Paper 53. https://www.ilo.org/wcmsp5/groups/public/---dgreports/---integration/documents/publication/wcms_079142.pdf

ILO, International Labour Organization (1993). Resolution concerning statistics of employment in the informal sector, adopted by the Fifteenth International Conference of Labour Statisticians (January 1993). https://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/normativeinstrument/wcms_087484.pdf

ILO, International Labour Organization (2002). Resolution concerning decent work and the informal economy. The General Conference of the International Labour Organization, meeting in its 90th Session, 2002. https://www.ilo.org/wcmsp5/groups/public/---ed_norm/---relconf/---reloff/documents/meetingdocument/wcms_080105.pdf

ILO, International Labour Organization (2013). Measuring Informality: A statistical manual on the informal sector and informal employment. https://www.ilo.org/global/publications/ilo-bookstore/order-online/books/WCMS_222979/lang--en/index.htm

Jiménez, A., Palmero-Cámara, C., González-Santos, M., González-Bernal, J. and Jiménez-Eguizábal, J. (2015). "The impact of educational levels on formal and informal entrepreneurship". Business Research Quarterly (2015) 18, pp. 204-212. https://journals.sagepub.com/doi/full/10.1016/j.brq.2015.02.002

La Porta, R. & Sheleifer, A. (2008) "The Unofficial Economy and Economic Development". Brookings Papers on Economic Activity, Fall 2008 https://www.brookings.edu/wp-content/uploads/2008/09/2008b_bpea_laporta.pdf

Loayza, N. & Rigolini, J. (2006). "Informality Trends and Cycles". World Bank Policy Research Working Paper 4078, December 2006. https://openknowledge.worldbank.org/bitstream/handle/10986/8855/wps4078.pdf?sequence=1&isAllowed=y

Loayza, N. (2016). "Informality in the process of development and growth". Policy Research Working Paper 7858. World Bank https://openknowledge.worldbank.org/bitstream/handle/10986/25303/WPS7858.pdf?sequence=1&isAllowed=y

Pisani, M & Pagán, J. (2004). Sectoral Selection and Informality: a Nicaraguan Case Study. Review of Development Economics, 8(4), 541-556 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=610465

Stata, Realase 13. Logitudinal-Data/Panel-Data Reference Manual.

Wooldridge, J. M. (2002). Econometric Analysis of Cross-Section and Panel Data. The MIT Press, Cambridge Massachusetts.

Wooldridge, J. M. (2003). Introductory Econometrics, a Modern Approach. 2d. Ed. Thomson South Western.

Wooldridge, J. M. (2010). "Correlated Random Effects Models with Unbalanced Panels". Working Paper. Michigan State University, Department of Economics. http://econ.msu.edu/faculty/wooldridge/docs/cre1_r4.pdf

Yamasaki, Izumi (2012). The Effect of Education on Earnings and Employment in the Informal Sector in South Africa. Ph.D. Dissertation, Columbia University. https://academiccommons.columbia.edu/doi/10.7916/D8280FV0/download

Appendix

Table A1: Countries and observations included in the sample

Developing Countries			Developed Countries		
Country	Observations Included	Year	Country	Observations Included	Year
Argentina	6	2008 - 2013	Austria	6	2008 - 2013
Bolivia	2	2008, 2009	Belgium	6	2008 - 2013
Brazil	4	2009, 2011 - 2013	Czech Rep	6	2008 - 2013
Colombia	6	2008 - 2013	Denmark	6	2008 - 2013
Costa Rica	5	2009 - 2013	Finland	6	2008 - 2013
Dominican Republic	6	2008 - 2013	France	6	2008 - 2013
Ecuador	3	2008 - 2010	Germany	6	2008 - 2013
El Salvador	6	2008 - 2013	Greece	6	2008 - 2013
Guatemala	4	2010 - 2013	Hungary	6	2008 - 2013
Honduras	6	2008 - 2013	Iceland	6	2008 - 2013
Mexico	6	2008 - 2013	Ireland	6	2008 - 2013
Panama	3	2009, 2012, 2013	Italy	6	2008 - 2013
Paraguay	6	2008 - 2013	Latvia	6	2008 - 2013
Peru	6	2008 - 2013	Luxembourg	6	2008 - 2013
Uruguay	6	2008 - 2013	Netherlands	6	2008 - 2013
Armenia	3	2009, 2012, 2013	Norway	6	2008 - 2013
India	2	2010, 2012	Poland	6	2008 - 2013
Moldova Republic	4	2009, 2011 - 2013	Portugal	6	2008 - 2013
Serbia	4	2010 - 2013	Russian Fed	6	2008 - 2013
Thailand	3	2011 - 2013	Slovak Rep	6	2008 - 2013
Turkey	4	2009, 2011 - 2013	Spain	6	2008 - 2013
Uganda	2	2010, 2013	Sweden	6	2008 - 2013
			Switzerland	6	2008 - 2013
			United Kingdom	6	2008 - 2013

Source: Own Elaboration

Table A2 Estimation Results from the Between Effects Model

Between Effects Estimates			
Dependent Variable: $infe_{it}$			
Variable	Coefficient	Std. Error	P-val
$\_cons$	174.2197***	26.9010	0.000
dc_{it}	-132.1177***	46.6737	0.008
$gdpg\_pc_{it}$	2.9347**	1.4097	0.045
$dc_{it} \times gdpg\_pc_{it}$	-4.9259**	2.2254	0.034
$lgfcf\_pc_{it-1}$	-17.4748***	4.2635	0.000
$dc_{it} \times lgfc\_pc_{it-1}$	14.3251**	6.1039	0.025
$educ_{it}$	-1.2245***	0.3572	0.002
$dc_{it} \times educ_{it}$	0.9922**	0.4167	0.023
$ehat\_ci_{it}$	-0.7455***	0.2629	0.008
$dc_{it} \times ehat\_ci_{it}$	0.7053**	0.3507	0.052
tsb_{it}	0.0668	0.1012	0.514
$dc_{it} \times tsb_{it}$	0.0384	0.2514	0.880
Number of obs = 241		Obs per group: min = 2	
Number of Groups = 46		average = 5.2	
F(11, 34) = 18.74		max = 6	
Prob > F = 0.0000			
Overall R-squared = 0.73.64			

Source: Own estimations. (***), (**) statistically significant coefficient at 1 and 5 percent respectively

Table A3 Data Sources and links

Variable	Data Source	Link	Year the data were retrieved
Informal employment rate for developing countries	International Labor Organization	https://ilostat.ilo.org/topics/informality/	2019
Vulnerable employment rate for developed countries	World Bank	Vulnerable employment, total (% of total employment) (modeled ILO estimate) | Data (worldbank.org)	2019
GDP per capita (constant 2005 US$)	World Bank-World Development Indicators	https://databank.worldbank.org/reports.aspx?source=world-development-indicators	2019
Gross Fixed Capital Formation (constant 2005 US$)	World Bank-World Development Indicators	http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators&Type=TABLE&preview=on#	**2019**
Educational Attainment	UNESCO & World Bank - Barro-Lee Indicators	http://data.uis.unesco.org/Index.aspx?queryid=134# http://databank.worldbank.org/data/reports.aspx?source=education-statistics-~-all-indicators&preview=on	2019
Corruption Perception Index	Tranparency International	https://www.transparency.org/en/cpi	2019
Time to start a business	World Bank -Doing Business	https://databank.worldbank.org/reports.aspx?source=doing-business	2019

Note: The Educational Attainment data was collected from two sources, given the data availability for each country in the sample; we basically used UNESCO and as a secondary source, we used the Barro-Lee Indicators (Barro-Lee: Percentage of population age 25+ with tertiary schooling. Total (Incomplete and Completed Tertiary) from the World Bank site.