2025

UAB
Universitat Autònoma
de Barcelona

The Informal Economy in Comparative Perspective: A Multidisciplinary Analysis

Daniela Contreras Martinez
Tutor: Xavier Vila Carnicero

2024/2025

Contemporary History, Politics and Economics

Table of contents

Author: Daniela Contreras Martinez
Tutor: Xavier Vila Carnicero
Studies: Bachelor's Degree in Contemporary History, Politics and Economics
Academic year: 2024/2025

Keywords

Informal economy; Informal employment; Informality output; Socioeconomic determinants; Multiple linear regression; Human capital; Tax burden; Corruption perception; Cross-country analysis.

Abstract

The informal economy includes all economic activities that are outside formal social-security systems and labor regulations, and it composes great part of many economies. This study offers a comparative, multidisciplinary analysis of informality across 49 countries classified by income level, using data predominantly from 2021–2024. Following the definitions established by the International Labour Organization (ILO) and the International Monetary Fund (IMF). The research differentiates between informal employment (measured as a percentage of total employment) and informal output (measured as a percentage of GDP). Gathered information from harmonized ILOSTAT survey data and supplementary sources, such as the World Bank, Transparency International, and the Human Development Index. Seven explanatory variables gather fiscal, economic, social, and institutional dimensions: tax burden, GDP per capita (PPP), unemployment rate, Gini index, mean years of schooling, HDI rank, and the Corruption Perceptions Index.

Autor: Daniela Contreras Martinez

Tutor: Xavier Vila Carnicero

Estudios: Grado en Historia, Política y Economía Contemporáneas

Curso académico: 2024/2025

Palabras clave

Economía informal; Empleo informal; Producción informal; Determinantes socioeconómicos; Regresión lineal múltiple; Capital humano; Carga impositiva; Percepción de la corrupción; Análisis comparativo internacional.

Resumen

La economía informal incluye todas las actividades económicas que funcionan al margen de los sistemas formales de seguridad social y de las regulaciones laborales y constituye una parte considerable de muchas economías. Este estudio presenta un análisis comparativo y multidisciplinar de la informalidad en 49 países, clasificados según su nivel de renta, utilizando datos principalmente de los años 2021–2024. Siguiendo las definiciones establecidas por la Organización Internacional del Trabajo (OIT) y el Fondo Monetario Internacional (FMI), el estudio distingue entre empleo informal (medido como porcentaje del empleo total) y producción informal (medida como porcentaje del PIB). Se recopilaron datos de las encuestas armonizadas de ILOSTAT y de fuentes complementarias, como el Banco Mundial, Transparencia Internacional y el Índice de Desarrollo Humano. Siete variables explicativas reúnen dimensiones fiscales, económicas, sociales e institucionales: carga impositiva, PIB per cápita (PPA), tasa de desempleo, índice de Gini, años promedio de escolaridad, posición en el IDH y el Índice de Percepción de la Corrupción.

Autor: Daniela Contreras Martinez
Tutor: Xavier Vila Carnicero
Estudis: Grau en Història, Política i Economia Contemporànies
Curs acadèmic: 2024/2025

Paraules clau

Economia informal; Ocupació informal; Producció informal; Determinants socioeconòmics; Regressió lineal múltiple; Capital humà; Càrrega fiscal; Percepció de la corrupció; Anàlisi comparativa internacional.

Resum

L'economia informal inclou totes les activitats econòmiques que operen fora dels sistemes formals de seguretat social i de les regulacions laborals, i constitueix una part important de moltes economies. Aquest estudi ofereix una anàlisi comparativa i multidisciplinària de la informalitat en 49 països, classificats segons el seu nivell de renda, i fa servir dades principalment dels anys 2021–2024. Seguint les definicions establertes per l'Organització Internacional del Treball (OIT) i el Fons Monetari Internacional (FMI), la recerca diferencia entre ocupació informal (mesurada com a percentatge de l'ocupació total) i producció informal (mesurada com a percentatge del PIB). S'han recollit dades de les enquestes harmonitzades d'ILOSTAT i de fonts suplementàries, com el Banc Mundial, Transparency International i l'Índex de Desenvolupament Humà. Set variables explicatives reuneixen dimensions fiscals, econòmiques, socials i institucionals: càrrega fiscal, PIB per càpita (PPA), taxa d'atur, índex de Gini, anys mitjans d'escolarització, posició en l'IDH i l'Índex de Percepció de la Corrupció.

1. Introduction

Informal economy is tricky to define at first glance. The term usually gets mixed up with concepts like underground, shadow, or illegal economy. Informality does not mean illegality. Following the International Labor Organization (ILO) methodology, in this research, the informal economy is any economic activity not covered by formal social-security systems and labor laws (ILO, 2018). That definition includes street vendors, unpaid family helpers, micro-enterprises that keep no formal books, and even big businesses with unregistered, casual workers. The International Monetary Fund (IMF) extends that to include domestic work as well as value produced informally inside formal firms (IMF, 2020). Both institutions emphasize the difficulty of its measurement as there is not a consensus on a universal statistical definition, making its international comparison a challenge.

The ILO usually links informality to poverty and open unemployment in developing economies. Researchers such as Castells and Portes (1989) believe, however, that informality is dynamic and profit-seeking, webbed into global supply chains rather than just a means of survival. Having this in mind, we can divide informality into three different aspects: informality as a means of survival, dependent exploitation (subcontracting of companies) and growth economy (micro-firms seeking to accumulate capital and expand).

Studying informal economy is important because of its magnitude. The IMF (2020) estimates that it accounts for as much as 50 % of GDP in some low-income countries and accumulates millions of workers with very little access to labor rights[2]. Under-reporting biases macro signals, not just that, but it also limits the tax base and hides policy choices. This study aims to unpack the social and economic drivers that contribute to the magnitude of informality.

With 49 countries' harmonized survey data, two dependent variables "proportion of informal employment" and "informal output (% of GDP)" are tested on how they respond to seven explanatory variables from GDP per capita to perception of corruption.

2. State of the art

Throughout literature, authors have identified factors that influence the size of informal economy. For example, Medina & Schneider (2018) point out three different categories as causes:

- Monetary factors: high taxes and labor costs that incentive companies to stay out of the legal framework (Tanzi, 1983).
- Regulatory factors: bureaucracy and excessive regulations driving firms and workers towards informality (Djankov et al., 2002)
- Institutional factors: corruption and lack of trust in governmental institutions that contribute to the growth of the informal sector (Friedman et al., 2000)

According to Schneider and Medina (2018), the estimated size of the informal economy in 158 countries between 1991 and 2015 was around 31.9% of GDP, with values that went from 7.2% in Switzerland to 62.3% in Bolivia. The regions that presented higher levels of informality were those of Latin America and Sub-Saharan Africa with values above 35%. However, there has been a downward trend of informality due to advances in technology and reforms to tributary systems. These results were drawn by using the MIMIC methodology which implies indirect estimates.

Despite multiple approaches in literature there remain many challenges in the comparison and prediction of the size of informal economy. Previous studies used methods such as MIMIC (Multiple Indicators Multiple Causes) or demand for money, which rely deeply on heavy macroeconomic assumptions that can lead to different estimates. This study contributes to literature through a model of multiple regression using Ilo Stat's database which gathers surveys. These direct estimations allow for a higher comparability between countries as they come from a standardized source. Multiple regression can be used to estimate the direct impact of GDP per capita, unemployment, poverty rates and other socio-economic variables. This approach offers the possibility to study significative statistical impacts of different variables in the size of informal economy.

Additionally, this study refers only to the legal informal economy, which does not act in illicit spaces like smuggling or drug trafficking. This distinction enables more specific analyses compared to other studies that lump all forms of legal and illegal informality together. Another important contribution to this work is the comparison of the informality in countries with different levels of development, which facilitates the identification of patterns and the evaluation of the impact of variables such as poverty, education, and corruption on informality. By differentiating between high, middle, and low income economies, the study provides a more detailed view of each socioeconomic context.

3. Object of study

This work will focus on the comparison of informal economy between countries of different levels of economic development, analyzing how socioeconomic conditions play a role in its size. On this note, the following research questions are proposed:

1. How does the proportion of informal employment vary between countries with different levels of economic development?
2. What socioeconomic factors have the greatest influence on the magnitude of the informal economy?
3. Is there a significant relationship between GDP per capita, unemployment rate, poverty rate, education, and the perception of corruption with the informal economy?

Based on these questions, the following working hypotheses have been formulated:

1. The informal economy is expected to be more prevalent in countries with a lower level of economic development. Since informality is often a response to a lack of opportunities in the formal sector, countries with weaker economies are likely to have a higher proportion of informal employment.
2. It is suggested that there are significant relationships between the proportion of informal employment and key variables such as GDP per capita, unemployment rate, poverty rate, education and perception of corruption.

4. Theoretical framework:

4.1. Definitions of informal economy

The definition of informal sector was introduced in 1972 in the International Labor Organization (ILO) report on Kenya, inspired by an earlier article by Keith Hart in 1970 (Botello Peñaloza & Guerrero Rincón, 2022). This contribution highlighted that in less developed countries, employment challenges were less about unemployment and more about underemployment, where workers had jobs but earned insufficient income (Portes and Haller, 2004). The report identified that some individuals survived through jobs that often operated outside regulatory frameworks. These activities ranged from subsistence-based work to profitable firms, yet remained unrecognized, unregistered, unprotected, and unregulated, despite their functional role in the broader economy.

In Latin America, the term informal sector gained relevance through the Regional Employment Programme for Latin America and the Caribbean (PREALC) of the ILO in the 1970s. Studies from PREALC, influenced by structuralist perspectives developed by the Economic Commission for Latin America and the Caribbean (ECLAC). Linking the informal sector to labor market dynamics and national development levels. The informal sector was seen as a result of an excess labor supply and a shortage of adequate employment opportunities, particularly quality jobs. This led to labor heterogeneity, where a small part of workers had high productivity and wages in modern economic sectors, while a much larger group remained in lower-productivity, lower-income occupations (Abramo, 2022).

The informal sector was initially understood as a different economic space separate from the formal economy. It consisted mainly of self-employed individuals (excluding professionals and technicians), unpaid family workers, microenterprise employees (in firms with five or fewer workers), and domestic workers. Data from the late 20th century showed the growing importance of the informal sector in urban employment, rising from about 30% between 1950 and 1980 to 50% by 2000 due to structural adjustment policies. The sector's job creation surpassed that of the formal sector, with informal jobs accounting for 7 out of 10 new jobs in the 1990s. However, by 2015, informal employment had declined to 43.1% (Deléchat & Medina, 2021), reflecting rapid formal employment growth during the commodity boom of the early 21st century.

Over time, the concept started to change. By the 1980s and 1990s, economic transformations and the restructuring of labor markets challenged the differentiation between formal and informal activity. The 2002 International Labor Conference (ILC) broadened the definition to informal economy, recognizing that informal employment existed not only in small-scale enterprises but also within formal firms and rural economies (Abramo, 2022). The 2003 International Conference of Labor Statisticians (ICLS) further refined this by introducing informal employment, covering three categories: employment in the informal sector, informal employment in the formal sector, and informal work in household production.

By 2015, informal employment accounted for 46.8% of urban workers in Latin America. However, new forms of informality have appeared due to technological change. Digital labor platforms, such as crowdsourcing and freelance work, have created employment arrangements that blur traditional relationships between employees and employers. These jobs often lack legal protections, social security, or labor rights, challenging conventional frameworks for regulating work (Martínez & Infante, 2019).

4.2. A Multidisciplinary Perspective on the Informal Economy

This research is not motivated only to illustrate how data correlates to each other to demonstrate the magnitude of informal economy. It aims to shine light on the social implications behind the data set. The goal is to present the informal economy through a multidisciplinary lens. Considering that data reflects real-life experiences and families. As the UNDP exposes, at least 60% of the global population earns their living through informal employment, including through digital platforms that as seen before having become a source of informal employment. With the latest crisis of the COVID-19, informality became even more apparent as a problem that needed to be controlled as many of these workers had no legal framework to be safeguarded against the effects of the lockdown. Therefore, it became clear how it was imperative to create jobs that could offer employees their rights in order to reduce poverty and inequality (UNDP, 2022).

Informal economy can't be studied unilaterally. As complex as it is to define, calculate and measure; it is also complex in each individual case. Each country's experience can be as unique as the one before, considering differences in social norms, cultural bases and economic development. Differentiating how this phenomenon manifests in developing versus underdeveloped countries can illustrate the structural factors that combine to shape informal economies. In developing economies, the informal sector usually appears as a response to rapid urbanization, regulatory gaps, and evolving labor markets. On the other hand, in underdeveloped countries it manifests as a means of survival, deeply rooted in traditional economies with limited opportunities to join the formal economy. Understanding these distinctions allows for a deeper understanding of the challenges and opportunities within each context (Gërxhani, 2005).

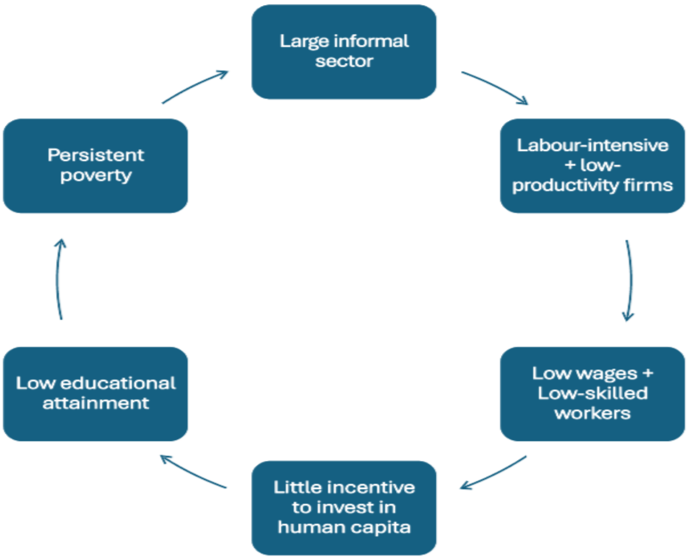
Informality can be seen both as a cause and a consequence of underdevelopment, often creating a vicious cycle that many less developed countries find difficult to break. In fact, there is also a high cost to meeting this goal in countries with higher informality as the UN has estimated that about $400 billion per year from 2019 to 2030 would be needed just for social protection, health, education, and climate change intervention (Ohnsorge et al., 2022). This is seen in the difficulties faced by countries with high levels of informality in achieving the Sustainable Development Goals (SDGs), which also widens the gap between developing and developed countries; affecting directly **Goal 10**, which aims to reduce inequality between and among countries. The SDGs are part of the 2030 Agenda for Sustainable Development, adopted by all UN Member States in 2015. Many of the social effects of informality, such as lack of social protection, precarious working conditions, and limited access to basic services, make the path towards development steeper. These challenges are particularly visible in relation to **Goal 8** (Decent work and economic growth), **Goal 6** (Clean water and sanitation), **Goal 1** (No poverty), **Goal 10** (Reduced inequalities), and **Goal 16** (Peace, justice, and strong institutions).

For countries with higher informality levels, it is harder to obtain a good amount of government revenue. According to the World Bank, government revenues in these countries tend to range between 5 and 12 percent of GDP. These limits to the budget also affect the coverage that the government can provide to citizens, including social protection programs and public services. This reveals a sort of domino effect, as the inability of the government to offer protection to citizens creates distrust among the population and reduces interaction with the state, mainly through maintaining informal activity (Ohnsorge et al., 2022). This whole dynamic reinforces a much deeper-rooted problem, which is institutional weakness. Factors such as poor handling of taxes, corruption, and overregulating governments all affect how the population acts upon the system and can therefore lead to higher informality when institutional weakness is perceived. At the same time, informality creates a viable atmosphere for these institutions to remain weak.

Lacking social protection does not only decrease trust and the government's stability, it also leaves informal workers far more exposed to poverty with external shocks like the COVID-19 crisis showed. More than a billion people worldwide were drawn below the poverty line due to payments made towards health care. *Lagging Behind* (2022) estimates that in high-informality countries those expenses are about 15 % more likely to impoverish households. Families that rely solely on informal earnings usually have no savings and no "safety net," whether from the state or from formal employment, making them even more vulnerable (Ohnsorge et al., 2022). As the World Bank and WHO warn, "the COVID-19 pandemic is likely to halt two decades of global progress toward Universal Health Coverage." Combined with widespread informality, this is a serious effect on development and widens the gap between countries. Although many governments provided targeted support during the pandemic, most aid focused on formally employed workers, leaving households outside legal employment uncovered. Not only this, but as countries have weaker government institutions, it is much harder to correctly target action. (WHO, 2021)

Ohnsorge, Okawa, and Yu (2022) mention in *Lagging Behind: Informality and Development* that informal firms are typically less productive and more labor-intensive than their formal counterparts. Where informality is deep-rooted and persistent, it is hardly surprising to also find persistent poverty. Labor-intensive, low-productivity activities pay lower wages and rely on low-skilled workers who become "trapped" into the job structure created by a large informal sector. Because such production does not require highly skilled labor, it seeks lower-cost workers and there is less incentive to invest in human capital. In this sense, a sizeable informal sector "demands" lower skills, and we can infer that countries with high informality levels also tend to have lower educational attainment.

Figure 1. *The Informality–Poverty Vicious Cycle*



Source: own creation.

5. Methodology

This study employs a quantitative and cross-sectional approach to analyze socioeconomic variables that may influence the magnitude of the informal economy in a sample of 50 countries across different income levels (low, middle, and high) from various regions. The classification that was followed was the World Bank's country classification and the United Nations' list of least developed countries. By analyzing socioeconomic and institutional variables through multiple linear regression across income groups, the study aims to provide a global perspective on informality.

The initial sample was gathered by compiling countries with available informal economy data from ILOSTAT, gathering about 125 countries. To narrow down the sample, countries lacking data for at least the year 2021, such as Cambodia, Malawi, or Kenya, were excluded. The remaining countries were then randomly filtered, resulting in a final sample of 49 countries: 19 high-income, 19 middle-income, and 9 low-income. The limited representation of low-income countries is acknowledged as a limitation; however, it reflects the challenge of lacking available data.

Seven variables were selected from the literature based on their presumed relevance to the size of the informal economy. The analysis focuses on identifying correlations between these variables and statistical significance. The data gathered corresponds to the most recent year available, mostly 2022, with a few exceptions' as for example Egypt and Guinea-Bissau, due to data limitations.

In order to analyze the magnitude of the informal economy, two dependent variables will be used to create different models of analysis. Model 1 will use "Proportion of informal employment as a %" as the observable dependent variable. Data was retrieved from the ILOSTAT database, which uses national labor force surveys, including population census, the Continuous National Survey of Labor Force, the EU Survey on Living Conditions and Income, among others. The data ranges from 2021 to 2024 depending on the country. This is in line with the first working hypothesis in Section 3

Model 2 is based on informal economy output, in other words, the proportion of the informal economy as a share of GDP. The data was gathered from the Explorer of Informal Economy Data by UNDP, ILOSTAT, and the World Bank. The last available data for this variable is mostly from 2018, and in some cases from 2017 (such as Benin, Gambia, and Togo). It is important to note missing values for countries like Peru, Tonga, Türkiye, and Serbia. This will help to investigate the second working hypothesis.

Using two dependent variables allows for a broader analysis of informality by capturing both its employment dimension and its contribution to economic output. This approach helps compare how different socioeconomic factors influence different aspects of informal economy and offers more robust insights across models (Loayza, 2016; ILO, n.d.).

5.1. Data collection

In the first phase, seven explanatory variables were chosen corresponding to the theoretical work mentioned earlier that can in some way or another influence the magnitude of the informal economy in countries. The variables try to captivate institutional, social and economic dimensions for a broader analysis.

The explicative variables that will be used in both models are as follows:

Table 1. *Variable definitions and data sources*

Variable	*Description*
Tax burden	Revenues received by the central government of each country, including personal and corporate income taxes, value added taxes, excise taxes, tariffs, social contributions (social security and hospital insurance), grants, and net revenues from public enterprises. Normalized by GDP, allowing cross-country comparisons. Data was retrieved from CIA.gov in *The World Factbook* using last available data mainly from 2023. Few countries have data from later years, as for example Egypt, with data from 2015.

GDP per capita, PPP (current international US$ 2022)	GDP holds all added value by resident producers, including product taxes but not subsidies. PPP adjustment refers to the application of conversion of prices that adjusts to the purchasing power of countries. The entire population is accounted for, regardless of legal conditions. Variable accounts for a comparable measure of average economic output throughout countries. Data collection through *The World Bank* data portal, all countries' values correspond to 2022.
Unemployment, total (% of total labor force) (modeled ILO estimate) 2022	Percentage of the unactive population out of the total labor force, including unactive youth. The ILO analyses global estimates that enable comparable labor data across the globe. Data gathered from *The World Bank* database.
Gini Index	Gini is used to measure inequality with values from 1 to 0. With value 1 representing perfect inequality. Values were also gathered from *The World Bank,* with data ranging from 2019 to 2022.
Mean years of schooling (2022)	The average years of education (not including repeating years) that have been completed by adults over 25 years of age. Data was retrieved from the Human Development Index.
HDI rank (2022)	The Human Development Index showcases three dimensions: - **Health:** life expectancy - **Education:** mean years of schooling + expected years of schooling - **Standards of living:** GDP per capita.

	Values for each dimension are summed up and composed into the index by geometric mean. Higher values in the index indicate higher development.
Corruption perceptions index (2022)	Index conducted by Transparency International, using a scale of 0 to 100, where 0 represents a highly corrupt public sector.

5.2. Running initial models

During the second phase, the initial estimation was developed through RStudio, using Ordinary Least Square models (OLS). The first model for proportion and output of informality is a full model which includes all explicative variables. To run the model in R and evaluate prediction accuracy it was imperative to use the following libraries:

- library(readxl)
- library(lmtest)
- library(car)
- library(MASS)
- library(sandwich)
- library (stargazer)

The code used for the models are as follows:

- **Model 1:**

```
Fullmodel1 <- lm(`proportion of informal employment` ~ Tax + PIB +
Unemployment + Gini + Schooling + HDI + Corruption, data = datos)
```

- **Model 2:**

```
fullmodel2 <- lm(`Informality by output` ~ Tax + PIB + Unemployment + Gini +
Schooling + HDI + Corruption, data = data)
```

Once the models are created, diagnosis are conducted on the assumptions of multiple linear regression:

Table 2. *Multiple Regression Assumptions*

Multiple Regression Assumptions	***Description***	***R Code***
Normality of errors	Residuals are assumed to be distributed normally (aligned towards the 45° line) for statistical inferences to be accurate. To make this diagnosis for the models used in this study, Normal Q-Q Plots are used.	qqnorm(resid(…)): qqline
Homoscedasticity	Values are meant to follow the same dispersion, which means residuals' values should not change due to explanatory variables. When heteroskedasticity is present, significance can be off which might lead to erroneous conclusions (Long & Ervin, 2000). To study this assumption, the *Breusch-Pagan* test is used. If the p-value is > 0.05 it means, there is no prove of homoscedasticity.	lmtest: :bptest(…)
Error autocorrelation	Errors need to be independent for valid statistical analysis. This can be verified through the *Breusch-Godfrey* test, with p-values > 0.05 rejecting autocorrelation.	lmtest: :bgtest(…)
Multicollinearity	Variables should not be too related to each other as one might undermine the significance of the other.	car: :vif(…)
Influence points	Outliers can dominate the results and therefore reduce robustness.	car: :influencePlot(…)

As **table 2** exposes, when OLS models do not meet the mentioned criteria, it can lead to misleading interpretation of results. Therefore, to correct the model three different actions are taken:

- **HC3:**

In case of heteroskedasticity, OLS models tend to lead to incorrect inferences, therefore the HC3 matrix is used to correct its effects. There are different versions of **heteroskedasticity-consistent estimators** that help make valid inferences even with unequal variances in a much easier manner and if there is no information on the form of heteroscedasticity. The HCO estimator (or White's estimator) is the most basic and as MacKinnon and White (1985) showed, it usually underestimates the true variances in small samples, and this can lead to misleading inferences. This is when the adjusted versions HC1, HC2 and HC3 come into play, allowing for better results for smaller samples. In this study HC3 is the matrix used as Long and Hervin (2000) recommend, this matrix has the stronger correction, and it is best to use when there are less than 250 samples and when influence points.

---

library(sandwich)

library(lmtest)

coeftest((…), vcov = vcovHC((…), type = "HC3"))"

- **Robust Regression:**

When outliers carry high leverage, OLS models turn fragile, as has been previously mentioned, a single "weird" value can distortion the results. Once it is confirmed that these extremes are real observations rather than errors, it is ideal to move to Robust Regression. The goal is not to throw data away but to re-ponder it through M-estimators. Trading the raw sum-of-squares for loss functions so each residual gets down-weighted instead of deleted (Li, 2006).

---

library(MASS)

Rlm_model(..) <- rlm(formula = … data = data)

- **Selective exclusion:**

Selective deletion of highly influential observations consists in identifying and removing those **outliers** that have a high impact on the parameter estimates; by comparing the coefficients before and after each omission, determining how much they shift and therefore how stable they remain when faced with extreme outliers.

```
data_no_outliers <- data[-c(...), ]

Model(...)_no_outliers <- lm(...)
```

5.3. Model selection

Complex models containing many variables are not always the ones with the best explanatory capacity. To achieve the best possible model, we turn to model selection through **Backward elimination** using the Leave-One-Out iterative method, it repeats the same process as many times as needed in order to increase precision and guide the model toward its optimum (Rodrigo, 2021). Starting with the full model, each run keeps every variable but one to validate the fit. However, omitting a single predictor can lead to high-variance error estimates, this is why the cycle is conducted for each variable in turn, calculating the *adjusted* R^2 (to check the prediction capacity) after every scenario. This procedure reduces the variability that appears when predictors are split at random. In this study, the variables left out are Schooling in one model and HDI rank in another, the choice driven by multicollinearity principles.

6. Results

This section gathers and presents empirical findings progressively. It is divided into three blocks:

- ❖ Descriptive statistics
- ❖ Results of the econometric models
- ❖ Relevant implications are synthesized.

6.1. Descriptive statistics

The analysis uses a cross-sectional sample of forty-nine countries with data from 2021-2024. Two metrics of informality (employment and output) and seven explanatory variables that capture fiscal, economic, social, and institutional dimensions are considered. **Table 4** summarizes the descriptive statistics (mean, median, standard deviation, minimum, and maximum) of all variables.

Table 3. *Descriptive statistics*

Variable	**Mean**	**Median**	**SD**	**Min.**	**Max.**
Proportion of informal employment%	43.6%	52.2%	35.9%	0.7%	96.3%
Informal output (% PIB)	28%	27.6%	10.6%	9.1%	53.8%
Tax burden	18.60	18.1	6.5	6.7	39.9
PIB per capita PPP (current international US$ 2022)	$30,462.09	$23,345.40	$27,160.94	$2,525.20	$144,872.70
Unemployment, total (% of total labor force) 2022	5.5%	4.8%	2.8%	0.9%	12.9%
Gini Index	35.4	34.3	7.4	21.2	54.8
Mean years of schooling (2022)	9.7	10.6	3.2	1.6	14.3
HDI rank (2022)	82.1	81	53.9	7	188
Corruption perceptions index (2022)	46.1	42	17.0	21	90

There is a great dispersion of informal employment as the proportion ranges from just 0.7% in Malta to 96.3% in Benin, highlighting the strong structural differences between economies. This wide range, reinforced by a high standard deviation, confirms the heterogeneity of the phenomenon between countries.

In terms of development, GDP per capita averages $30,462, evidencing the coexistence of low-income and very high-income economies when looking at the difference between minimum and maximum values. Social indicators such as the unemployment rate average 5.5, the Gini index 35.4, and average years of schooling 9.7; these indicators show smaller but still significant differences. Finally, the HDI average 82.1 and the Corruption Perception Index average 46.1; confirm notable differences in well-being and governance. Overall, the wide range highlights the structural heterogeneity of the sample.

The results gathered will be illustrated by the five best results and the five worst. The complete information can be found in the (appendices…).

Table 4. *Dependent variables (Top 5 & Bottom 5).*

Variable	**Country with lowest results**	**Value**	**Country with highest results**	**Value**
Proportion of informal employment%	**Malta**	0.7%	**Benin**	96.3%
	Bulgaria	1.2%	**Mali**	95.4%
	Croatia	1.4%	**Senegal**	95.1%
	Cyprus	1.8%	**Guinea-Bissau**	94.8%
	Germany	2%	**Burkina Faso**	93.8%
Informal output (% PIB)	**Luxembourg**	9.1%	**Bolivia**	53.8%
	Viet Nam	11.6%	**Panama**	46%
	Netherlands	12.6%	**Benin**	45.4%
	France	14.1%	**Peru**	44.8%
	Germany	15%	**Republic of Moldova**	42.9%

As can be seen from this summary, the five countries with the lowest informality rates all fall within the high-income group of the sample, with the exception of Bulgaria, classified as upper-middle-income. On the other hand, the five countries with the highest informality levels are uniformly low-income economies.

There is a slight shift when we turn to informal output, the top five performers are from low, middle and high income groups, revealing a more heterogeneous pattern. By contrast, the bottom-five follows the employment trend, composed entirely of low income countries.

Table 5. *Economic variables (Top 5 & Bottom 5).*

Variable	**Country with lowest results**	**Value**	**Country with highest results**	**Value**
Tax burden	**India**	6.7	**Bolivia**	39.9
	Panama	7.5	**Denmark**	31.4
	Bangladesh	7.6	**Luxembourg**	27.8
	Guinea-Bissau	8.6	**Greece**	26.6
	Paraguay	10.2	**Jamaica**	25.7
PIB per capita, PPP (current	**Mali**	$2,525.20	**Luxembourg**	$144,872.70
	Burkina Faso	$2,641.90	**Denmark**	$77,400.30

international US$ 2022)	**Guinea-Bissau**	$2,685.80	**Netherlands**	$77,152.20
	Togo	$2,852.10	**Germany**	$67,589.80
	Gambia	$3,067.30	**Finland**	$61,344.60
Unemployment, total (% of total labor force) (modeled ILO estimate) 2022	**Republic of Moldova**	0.9%	**Spain**	12.9%
	Viet Nam	1.5%	**Greece**	12.4%
	Benin	1.7%	**Costa Rica**	11.3%
	Togo	2%	**Türkiye**	10.5%
	Tonga	2.3%	**Colombia**	10.5%

Tax burden ranges from a low of 6.7% of GDP in India to a high of 39.9% in Bolivia. The five lowest values are concentrated in lower- and middle-income emerging economies like India, Panama, Bangladesh, Guinea-Bissau, and Paraguay. While the highest figures, except for Bolivia and Jamaica, correspond to high-income countries such as Denmark, Luxembourg, and Greece. This suggests that tax collection capacity can be connected to more developed economies.

There is a vast gap in income with the lowest values between $2,600 and $3,100, belonging to countries in Africa, like Mali, Burkina Faso, Guinea-Bissau, Togo, and Gambia. While the highest income is held by Luxembourg $144,872.70, followed by advanced European economies.

The countries with the lowest rates of unemployment are Republic of Moldova (0.9%), Viet Nam (1.5%), Benin (1.7%), Togo (2.0%), and Tonga (2.3%). These different countries and levels of development suggest structural effects (intensive agriculture or hidden underemployment). In contrast, the highest rates are concentrated in Spain (12.9%) and Greece (12.4%), followed by Costa Rica, Türkiye, and Colombia (ranging from10% to 11%), evidencing labor rigidities and recent shocks in middle- and high-income economies (for example, COVID-19).

Table 6. *Social Variables (Top 5 & Bottom 5)*

Variable	Country with lowest results	Value	Country with highest results	Value
Gini index	**Slovak Republic**	21.2	**Colombia**	54.8
	Republic of Moldova	25.7	**Zambia**	51.5
	Netherlands	26.3	**Panama**	50.9
	Poland	26.3	**Costa Rica**	47.2
	Kyrgyzstan	26.4	**Ecuador**	45.5
Mean years of schooling (2022)	**Mali**	1.6	**Germany**	14.3
	Burkina Faso	2.3	**Lithuania**	13.5
	Senegal	2.9	**Latvia**	13.3
	Benin	3.1	**Poland**	13.2
	Guinea-Bissau	3.7	**Luxembourg**	13
HDI rank (2022)	**Germany**	7	**Mali**	188
	Denmark	8	**Burkina Faso**	185

Finland	11	**Guinea-Bissau**	179
Netherlands	11	**Gambia**	174
Luxembourg	19	**Benin**	173

The five most egalitarian countries are the Slovak Republic (21.2), Republic of Moldova (25.7), Netherlands (26.3), Poland (26.3) and Kyrgyzstan (26.4). All belonging to high-income EU members and middle-income transition economies (Kyrgyzstan). While, the greatest inequality is observed in Colombia (54.8), Zambia (51.5), Panama (50.9), Costa Rica (47.2) and Ecuador (45.5), pointing to structural inequalities typical of middle-income Latin American and sub-Saharan countries.

Mean years of schooling show even a stronger difference between income levels. The lowest values belong to Mali (1.6), Burkina Faso (2.3), Senegal (2.9), Benin (3.1) and Guinea-Bissau (3.7). The highest being Germany (14.3), Lithuania (13.5), Latvia (13.3), Poland (13.2) and Luxembourg (13.0), belonging once again to high-income European countries.

The developmental division is confirmed by the HDI rank, as the top positions go to Germany (7) and Luxembourg (19), all high-income, while the bottom ranking goes to Mali (188), Burkina Faso (185), Guinea-Bissau (179), Gambia (174) and Benin (173), these values align to the countries that also show the weakest education scores.

***Table** 7. Institutional Variables (Top 5 & Bottom 5)*

Variable	Country with lowest results	Value	Country with highest results	Value
Corruption perceptions index (2022)	**Guinea-Bissau**	21	**Denmark**	90
	Bangladesh	25	**Finland**	87
	Kyrgyzstan	27	**Netherlands**	80
	Mali	28	**Germany**	79
	Paraguay	28	**Luxembourg**	77

The CPI widens the differences already seen in the previous indicators. Scores range from 21 points in Guinea-Bissau to 90 points in Denmark, a spread of almost 70 points on a 0-100 scale where higher values mean cleaner public sectors. The five countries at the bottom, Guinea-Bissau, Bangladesh, Kyrgyzstan, Mali, and Paraguay are all low- to middle-income economies struggling with weak institutions and limited fiscal capacity. On the other hand, the top ranked, Denmark, Finland, the Netherlands, Germany, and Luxembourg are all high-income democracies. This contrast highlights governance quality as another structural dimension that could shape the informal sector dynamics and tax performance across the sample.

6.2. Results of the econometric models

The results are organized into two blocks (**Model 1**) informal employment proportion and (**Model 2**) informal output (% GDP).

- ❖ **Model 1:**

Table 8. *FullModel1*

	Dependent variable:
	Informality
Tax	-0.455 (0.317)
`PIB per capita`	-0.0001 (0.0001)
Unemployment	-1.708** (0.837)
`Gini index`	0.891*** (0.300)
Schooling	-0.257 (1.681)
HDI	0.466*** (0.137)
Corruption	-0.058 (0.204)
Constant	-1.149 (36.677)
Observations	48
R2	0.901
Adjusted R2	0.884
Residual Std. Error	12.353 (df = 40)
F Statistic	51.964*** (df = 7; 40)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 9. *No_HDI1*

	Dependent variable:
	Informality
Tax	-0.204 (0.345)
PIB	-0.0003** (0.0001)
Unemployment	-3.177*** (0.803)
Gini	0.794** (0.334)
Schooling	-5.261*** (0.906)
Corruption	-0.337 (0.209)
Constant	111.536*** (17.471)
Observations	48
	0.872
	0.854
Residual Std. Error	13.847 (df = 41)
F Statistic	46.718*** (df = 6; 41)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 9. *No_HDI1*

Table 8 illustrates the results for the full model 1, which sets proportion of informal employment as the dependent variable with the complete set of explanatory variables. This model holds a high explanatory capacity as can be seen by the high adjusted R^2 (0.884). Gini index comes up as significant with a positive sign, which indicates that if inequality increases, informality would have an increase of 0.891 points. Another variable that shows significance is unemployment, this time with a negative value, implying that once unemployment increases, informality does too. This might seem counterintuitive; however, it can be due to the informal sector not directly absorbing formal unemployment. At last, the HDI ranks shows significance with a positive value, countries that are worst off in the ranking increase informality by 0.466 points. By contrast, tax burden, per-capita income, average schooling, and perceived corruption show the expected signs but are not statistically significant, this can be due to multicollinearity that diffuses the individual effects.

The exclusion of HDI from the model mitigates the effects of multicollinearity and changes the hierarchy of predictors. Average schooling now ranks highest as each additional year of education reduces by a 5.26 point reduction in informality, highlighting the significance of human capital in incentivizing individuals toward formal employment. The adverse impact of unemployment increases while the Gini index loses its significance, suggesting that the impact seen before was, in part, occasioned by developmental variables held by HDI. Lastly, per-capita income shows weak significance, although persistent indication that high-income economies tend to have lower levels of informality when controlling for development conditions. The explanatory capacity (Adjusted R^2 = 0.854) of the model excluding HDI is slightly lower than the full model.

By looking at the model without schooling (annex 2), the main correlations stay almost unchanged. Inequality is still the key driver as in the full model, every extra point in the Gini index raises informality by 0.90. The HDI rank keeps a positive and significant coefficient, which shows that countries located farther down the human-development rank have higher informal sectors. Unemployment keeps its negative sign. In contrast, tax burden, PIB per capita and the corruption index keep the expected signs but remain statistically non-significant. The model still explains a large portion of the variance (Adjusted R^2 = 0.886).

- **Model 2**

Table 10. *FullModel2*

	Dependent variable:
	`Informality by output`
Tax	0.466**
	(0.223)
PIB	-0.0001
	(0.0001)
Unemployment	-0.256
	(0.579)
Gini	0.292
	(0.208)
Schooling	0.311
	(1.224)
HDI	0.037
	(0.100)
Corruption	-0.208
	(0.151)
Constant	16.565
	(26.377)
Observations	46
R2	0.463
Adjusted R2	0.364
Residual Std. Error	8.436 (df = 38)
F Statistic	4.672*** (df = 7; 38)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 11. *No_HDI2*

	Dependent variable:
	`Informality by output`
Tax	0.489**
	(0.212)
PIB	-0.0001
	(0.0001)
Unemployment	-0.353
	(0.510)
Gini	0.286
	(0.205)
Schooling	-0.091
	(0.551)
Corruption	-0.234*
	(0.131)
Constant	25.458**
	(10.559)
Observations	46
R2	0.461
Adjusted R2	0.378
Residual Std. Error	8.342 (df = 39)
F Statistic	5.551*** (df = 6; 39)
Note:	*p<0.1; **p<0.05; ***p<0.01

Model 2, holds informality by output (% GDP) as the dependent variable. Once again, we observe the full model and two versions of the initial model, one excluding HDI and the other excluding schooling. The model containing all explanatory variables only shows statistical significance for tax burden, in this case with a positive sign, contrasting with the results for model 1. Increasing tax burden also increases output of informality by 0.466 percentual points. Schooling also presents a positive sign, however this is most likely due to collinearity with HDI, as the model that excludes this variable presents the expected negative sign for the schooling variable. As for PIB per capita, unemployment, Gini index, HDI and corruption maintain the expected signs that model 1 predicted, however they don't show statistical significance. The explanatory capacity of full model 2 is below satisfactory with a 0.364 adjusted R^2.

When excluding HDI, tax burden maintains as the key driver, while corruption gains significance ($p<0.1$). When the CPI shows higher values (a cleaner public sector), output of informality decreases by 0.234 percentual points. The R^2 sees a slight increased performance with a value of 0.378. On the other hand, when schooling is excluded (annex 3), tax burden stays as the only significant variable. R^2 for this model maintains almost similar values.

6.3. Diagnosis

Diagnostic checks based on the assumptions of multilinear regression for model 1 show results that are broadly consistent (annex 4 & 5). In the full model the residuals deviate in the left tail, however there is no heteroscedasticity detected (Breusch–Pagan $p = 0.57$) and no autocorrelation (Breusch–Godfrey $p = 0.09$). The only warning light is multicollinearity around HDI (VIF > 9), so we fall back on HC3 errors, a robust-linear fit (RLM) and, removing outliers (14 and 41).

Once HDI is removed, the No-HDI variant keeps the heavy-tailed pattern but still shows homoscedastic errors ($p = 0.08$); however, first-order autocorrelation appears (Breusch–Godfrey $p = 0.04$). Multicollinearity falls to safe levels (all VIF < 5), and the same influential points resurface, so we apply the identical HC3 + RLM + selective-exclusion corrective actions.

Finally, the No-Schooling run again shows a left-tail skew, but it remains free of heteroscedasticity ($p = 0.70$) and autocorrelation ($p = 0.10$), with acceptable collinearity (all VIF < 5). Influence point diagnosis points out the same observations 14 and 41, and we address them with the same robust corrections.

To conclude, we compare every correction to the initial full model.

Table 12. Robustness specifications of the FullModel1

Full Model 1 Proportion of informal employment	**Initial model**	**HC3**	**RLM**	**Excluding outliers**
Tax burden	-0.455	-0.46	** -0.619	*** -1.028
PIB per capita	-0.0001	-0.00008	-0.0001	** -0.001
Unemployment	** -1.708	* -1.71	** -1.851	** -1.953
Gini index	*** 0.891	* 0.89	*** 0.921	** 0.570
Schooling	-0.257	-0.26	-0.074	-2.331
HDI rank	*** 0.466	. 0.47	*** 0.467	0.180
Corruption	-1.149	-0.06	-0.035	0.191
Adj. R^2	0.844	-	-	0.908

The three corrective lenses offer a clearer picture of informality across variables. When turning to HC3, estimates are almost unchanged, as inequality remains the strongest while unemployment and HDI lose significance. The robust model introduces negative significance to tax burden by down weighting the effects of influence points. Finally, by getting rid of the outliers altogether, PIB per capita gains relevance and tax burden is even more significant than observed in RLM. However, HDI is no longer significant by cutting off outliers which suggests that its earlier significance is held by those influence points. Model fit increases when excluding outliers which reinforce results.

For **model 2**, the same diagnosis is carried out (annex 9, 13 & 15), revealing comparable results to model 1. Heavy tails and asymmetry are corrected through HC3 tests. While there is no evidence of autocorrelation nor heteroscedasticity, multicollinearity persists in the full model 2 with VIF > 10, while this distortion is corrected in the No_HDI and No_Schooling observations. Influence points are also prevalent in model 2, this time with 14, 41 and 27.

Table 13. Robustness specifications of the FullModel2

Full Model 2 Informality by output (%GDP)	**Initial model**	**HC3**	**RLM**	**Excluding outliers**
Tax burden	** 0.466	0.47	*** 0.596	0.405
PIB per capita	-0.0001	-0.00008	-0.0001	0.0001
Unemployment	-0.256	-0.26	-0.375	-0.104
Gini index	0.292	0.29	0.210	0.159
Schooling	0.311	0.31	0.196	0.600
HDI rank	0.037	0.04	0.026	0.118
Corruption	-0.208	-0.21	* -0.230	-0.186
Adj. R^2	0.364	-	-	0.291

When the corrections are conducted for Model 2 (informality by output) the results are far less stable than in Model 1. In the initial model, tax burden is the only variable that shows statistical relevance, with a positive coefficient of 0.466, however, the rest of the variables remain non-significant. After running the HC3 robust errors, the results are practically the same. Tax burden keeps its magnitude and significance, and no other variable gains explanatory power. Moving to the RLM correction, tax burden strengthens even more (0.596), and corruption becomes negatively significant (-0.230), pointing to a possible link between better governance and lower informal output once the influence of outliers is downweighed. Finally, excluding the outliers (14, 27 and 41) weakens the tax burden to a value of 0.405, which is no longer significant. Unemployment, Gini, Schooling and HDI continue without statistical significance under any correction. Model fit is also reduced, with Adj. R^2 falling from 0.364 to 0.291 after removing the outliers, confirming that those extreme observations were carrying a good share of the explanatory power for informal output.

With these results we can draw the conclusion that the best explanatory power on proportion of informal employment is held by the full model when excluding outliers when we are guided purely by the Adjusted R^2 value. However, this is with leaving out influence points which is a trade-off. As for Model 2, there is no model that can fully predict its magnitude. The initial model holds the highest value for Adjusted R^2; however, it does not meet the assumptions of multilinear regression making its results unreliable.

7. Conclusion

This research sets out to identify how socio-economic variables influence the magnitude of informal economy by combining a cross-sectional dataset of 49 countries.

First, informality is far from a uniform phenomenon. Median informal employment in low-income countries is more than twenty times higher than in high-income economies. However, when informal output is analyzed, the dispersion among countries narrows down. Emphasizing that what matters is not only how many people work informally but how much value those activities create.

Second, the drivers of informality change depending on the dependent variable analyzed. For the proportion of informal employment, educational attainment, and unemployment figure as the strongest and most stable predictors. Each extra year of schooling lowers informality by about five percentage points, while increasing unemployment is associated with about 1.7 percentage points less of informal employment, perhaps reflecting that both formal and informal job opportunities are affected by externalities. Therefore, for further research it would be beneficial to include a variable that might explain this counter-intuitive result. As for example, unemployment aid which might capture unemployment. Inequality (captured by the Gini index) also correlates positively with informality; however, its significance mitigates when multicollinearity with development indicators is corrected. On the other hand, when the dependent variable is informal output, tax pressure and the perceived cleanliness of public institutions become the key indicators, indicating that informality from this point of view is seen from a cost-benefit perspective based on evasion vs. compliance. It is important to note that output is limited by the year of its last available data. This may, to an extent, explain why Model 2's explanatory capability is comparatively low. With more recent output data result could differ.

Third, the robustness of the models does matter. OLS models alone gave apparent statistically clean results; however, these results were not completely reliable. Once corrections on HC3, robustness, and the selective removal of outliers were conducted, clearer and more consistent results were drawn. The final model for proportion of informal employment (Full Model 1 – excluding outliers) explained up to 90 percent of the cross-country variance. However, results for informal output were less clear (about 30 percent), which highlights how this variable is harder to analyze with available data.

Policy recommendations based on these results would be essentially to invest in human capital. Restructuring tax systems and implementing credible anti-corruption efforts, while not the complete solution, can shrink the output side of informality by leaning cost-benefit decisions toward compliance. Finally, any agenda aimed at "formalizing" jobs must be aware of context, in low-income settings, informality is often a survival option. While in middle-income economies it may coexist with complex, subcontracted segments of global supply chains (Castells & Portes, 1989).

Future research should take analysis in three directions. Analyzing data over a period of time would allow for a follow-up on the dynamic side of informality. With a wider set of institutional variables, the relative significance of perceived corruption and tax burden could be reduced. Finally, qualitative work could have a deeper take on the lived reality behind the regression lines. To conclude, it is not so much a question of addressing one policy to reduce informality but of working up the social, educational, and institutional matters that enable workers and firms to move willingly into the formal sector.

8. Bibliography

- Abramo, L. (2022). *The evolution of the concept of informality in Latin America and new challenges for social and labor protection*. *Brazilian Journal of Social and Labour Economics*. https://doi.org/10.20396/rbest.v4i00.17311
- Bacchetta, M., Ernst, E., & Bustamante, J. P. (2009). *Globalization and informal jobs in developing countries*. International Labour Office & World Trade Organization. https://www.ilo.org/publications/globalization-and-informal-jobs-developing-countries-joint-study
- Barón López, F. J. (2022, November 15). Capítulo 7: Regresión lineal múltiple—Requisitos y limitaciones. In *Apuntes de estadística*. Universidad de Málaga. https://www.bioestadistica.uma.es/apuntesMaster/regresión-lineal-múltiple.html#requisitos-y-limitaciones
- Bonnet, F., Vanek, J., & Chen, M. A. (2019, January 31). *Women and men in the informal economy: A statistical brief*. Women in Informal Employment: Globalizing and Organizing (WIEGO). https://www.wiego.org/research-library-publications/women-and-men-informal-economy-statistical-brief/
- Botello Peñaloza, H. A., & Guerrero Rincón, I. (2022). Informal economy and income distribution in Ecuador during the COVID-19 pandemic. Revista Facultad de Ciencias Económicas: Investigación y Reflexión, 30(1), 11–28. https://doi.org/10.18359/rfce.5206
- Castells, M., & Portes, A. (1989). *World underneath: The origins, dynamics, and effects of the informal economy*. In A. Portes, M. Castells, & L. A. Benton (Eds.), *The informal economy: Studies in advanced and less developed countries* (pp. 11–37). Johns Hopkins University Press.
- Central Intelligence Agency. (n.d.). *Field listing: Taxes and other revenues*. *The World Factbook*. https://www.who.int/news/item/12-12-2021-More-than-half-a-billion-people-pushed-or-pushed-further-into-extreme-poverty-due-to-health-care-costs
- Datosmacro. (n.d.). *Índice de Gini 2022*. Expansión. https://datosmacro.expansion.com/demografia/indice-gini?anio=2022
- Deléchat, C., & Medina, L. (2020). What is the informal economy? *Finance & Development*, Quarterly Publication of the International Monetary Fund and the World Bank. https://www.imf.org/external/pubs/ft/fandd/spa/2020/12/pdf/what-is-the-informal-economy-basics.pdf
- Diez, D. M., Çetinkaya-Rundel, M., & Barr, C. D. (2019). Introduction to linear regression (Chap. 8). In *OpenIntro statistics* (4th ed.). OpenIntro. https://www.openintro.org/book/os/#ch-8-introduction-to-linear-regression
- Diez, D. M., Çetinkaya-Rundel, M., & Barr, C. D. (2019). Multiple and logistic regression (Chap. 9). In *OpenIntro statistics* (4th ed.). OpenIntro. https://www.openintro.org/book/os/#ch-9-multiple-and-logistic-regression
- Djankov, S., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2002). The regulation of entry. *The Quarterly Journal of Economics, 117*(1), 1–37. https://doi.org/10.1162/003355302753399436
- Elgin, C., Erturk, F. (2019). Informal economies around the world: Measures, determinants and consequences. *Eurasian Economic Review, 9*, 221–237. https://doi.org/10.1007/s40822-018-0105-5
- Elgin, C., Kose, M. A., Ohnsorge, F., & Yu, S. (2019, October). *Shades of grey: Measuring the informal economy business cycles* [Conference paper]. 7th IMF Statistical Forum. https://www.imf.org/-/media/Files/Conferences/2019/7th-statistics-forum/session-ii-yu.ashx

❖ Friedman, E., Johnson, S., Kaufmann, D., & Zoido-Lobatón, P. (2000). Dodging the grabbing hand: The determinants of unofficial activity in 69 countries. *Journal of Public Economics, 76*(3), 459–493. https://doi.org/10.1016/S0047-2727(99)00093-6

❖ Gërxhani, K. (2004). The informal sector in developed and less developed countries: A literature survey. *Public Choice, 120*, 267–300. https://doi.org/10.1023/B:PUCH.0000044287.88147.5e

❖ Hlavac, M. (2022). *stargazer: Well-formatted regression and summary statistics tables* (Version 5.2.3) [R package]. https://CRAN.R-project.org/package=stargazer

❖ International Labour Organization. (2023, December 20). *Statistical Methodology Series 11: Identification of informality through labour force surveys* (ILO Department of Statistics Methodology Report). International Labour Office. https://www.ilo.org/sites/default/files/wcmsp5/groups/public/%40dgreports/%40stat/documents/publication/wcms_907238.pdf

❖ International Labour Organization. (2025). *ILO modelled estimates: Methodological overview*. https://webapps.ilo.org/ilostat-files/Documents/TEM.pdf

❖ International Labour Organization. (n.d.). *Gini index (or Gini coefficient)*. Probabilidad y Estadística. https://www.probabilidadyestadistica.net/indice-de-gini-o-coeficiente-de-gini/

❖ International Monetary Fund. (2024). *Government Finance Statistics (GFS) database* [Data set]. https://legacydata.imf.org/?sk=a0867067-d23c-4ebc-ad23-d3b015045405

❖ International Monetary Fund. (2025, February). *Public Finances in Modern History (FPP) database – Government revenue, % of GDP* [Data set]. IMF Data Mapper. https://www.imf.org/external/datamapper/rev@FPP/USA/FRA/JPN/GBR/SWE/ESP/ITA/ZAF/IND

❖ Li, G. (2006). Robust regression. In D. C. Hoaglin, F. Mosteller, & J. W. Tukey (Eds.), *Exploring data tables, trends, and shapes* (pp. 281–343). John Wiley & Sons. https://doi.org/10.1002/9781118150702.ch8

❖ Long, J. S., & Ervin, L. H. (2000). Using heteroscedasticity-consistent standard errors in the linear regression model. *The American Statistician, 54*(3), 217–224. https://doi.org/10.2307/2685594

❖ Martínez, D., & Infante, R. (2019, November 22). *Informality in the ILO's vision: Evolution and prospects for Latin America*. https://www.ilo.org/es/resource/article/la-informalidad-en-la-vision-de-la-oit-evolucion-y-perspectivas-para

❖ Medina, L., & Schneider, F. (2018). *Shadow economies around the world: What did we learn over the last 20 years?* (IMF Working Paper No. 18/17). International Monetary Fund. https://www.imf.org/en/Publications/WP/Issues/2018/01/25/Shadow-Economies-Around-the-World-What-Did-We-Learn-Over-the-Last-20-Years-45583

❖ Ohnsorge, F., Okawa, Y., & Yu, S. (2022). Lagging behind: Informality and development. In F. Ohnsorge & S. Yu (Eds.), *The long shadow of informality: Challenges and policies* (pp. 125–192). World Bank. https://doi.org/10.1596/978-1-4648-1753-3

❖ Organisation for Economic Co-operation and Development. (2024, December 6). *Global Revenue Statistics Database* [Data set]. https://www.oecd.org/en/data/datasets/global-revenue-statistics-database.html

❖ Portes, A., & Haller, W. (2015). The informal economy. *ResearchGate*. https://doi.org/10.13140/RG.2.1.3661.5763

❖ Quiros-Romero, G., Alexander, T. F., & Ribarsky, J. (2021, February 2). *Measuring the informal economy* (IMF Policy Paper No. 2021/002). International Monetary Fund. https://www.imf.org/en/Publications/Policy-Papers/Issues/2021/02/02/Measuring-the-Informal-Economy-50057

- Ripley, B., Venables, W., Bates, D. M., Hornik, K., Gebhardt, A., & Firth, D. (2025, February 19). *rlm: Robust fitting of linear models* (Version 7.3-65) [R package function documentation]. In *MASS: Support functions and datasets for Venables and Ripley's MASS*. R Core Team. https://stat.ethz.ch/R-manual/R-patched/library/MASS/html/rlm.html
- Rodrigo, J. A. (2020). Validation of predictive models: Cross-validation, OneLeaveOut, Bootstrapping. *Ciencia de Datos*. https://cienciadedatos.net/documentos/30_cross-validation_oneleaveout_bootstrap
- Sachs, J. D., Lafortune, G., & Fuller, G. (2024). *Sustainable Development Report 2024: The SDGs and the UN Summit of the Future*. Sustainable Development Solutions Network & Dublin University Press. https://doi.org/10.25546/108572
- Tanzi, V. (1983). The underground economy in the United States: Annual estimates, 1930–1980. IMF Staff Papers, 30(2), 283–305. https://doi.org/10.2307/3867001
- Trading Economics. (2025). *Gambia – Tax revenue (% of GDP): 2025 data, 2026 forecast, 1989-2009 historical* [Data set]. Based on World Bank indicators. https://tradingeconomics.com/gambia/tax-revenue-percent-of-gdp-wb-data.html
- Transparency International. (n.d.). *Corruption Perception Index*. https://www.transparency.org/en/cpi/2024
- United Nations. (n.d.). *The 17 Goals*. https://sdgs.un.org/goals
- United Nations Conference on Trade and Development. (2024, December). *UN list of least developed countries*. https://unctad.org/topic/least-developed-countries/list
- United Nations Development Programme. (n.d.). *Human Development Index (HDI)*. *Human Development Reports*. https://hdr.undp.org/data-center/human-development-index#/indicies/HDI
- United Nations Development Programme. (2022, June 13). *A data explorer on the informal economy*. https://data.undp.org/es/insights/informal-economy
- United Nations University World Institute for Development Economics Research. (2024). *Government Revenue Dataset (GRD)* [Data set]. https://www.wider.unu.edu/database/data-and-resources-grd
- World Bank. (n.d.). *GDP per capita, PPP (current international $) by country*. Trading Economics. https://tradingeconomics.com/country-list/gdp-per-capita-ppp-us-dollar-wb-data.html
- World Bank. (n.d.). *Mean years of schooling of the population age 25+ (total) by country*. Trading Economics. https://tradingeconomics.com/country-list/uis-mean-years-of-schooling-of-the-population-age-25-total-wb-data.html
- World Bank. (2023, May 24). *Benin Economic Governance for Service Delivery Program for Results (P176763): Technical assessment* [Program appraisal document]. https://documents1.worldbank.org/curated/en/099052423104516832/pdf/P17676301da8ed0908e2f00a71b089fb24.pdf
- World Bank. (2025). *Poverty headcount ratio at $3.65 a day (2017 PPP) (% of population)* [Data set]. UNdata, World Development Indicators. https://data.un.org/Data.aspx?d=WDI&f=Indicator_Code%3ASI.POV.LMIC
- World Bank. (2025). *World Development Indicators: United Arab Emirates* [Data set]. World Bank Group. https://databank.worldbank.org/reports.aspx?source=2&country=ARE
- World Health Organization. (2021, December 12). *More than half a billion people pushed or pushed further into extreme poverty due to health-care costs*. https://www.who.int/news/item/12-12-2021-More-than-half-a-billion-people-pushed-or-pushed-further-into-extreme-poverty-due-to-health-care-costs
- World Population Review. (2025). *Highest taxed countries 2025*. https://worldpopulationreview.com/country-rankings/highest-taxed-countries

- (n.d.). *What is: Iterative method*. Statistics Easily. https://es.statisticseasily.com/glosario/%C2%BFQu%C3%A9-es-el-m%C3%A9todo-iterativo-en-la-ciencia-de-datos%3F/
- Zeileis, A. (2004, November 29). Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software, 11*(10), 1–17. https://doi.org/10.18637/jss.v011.i10
- Zeileis, A., Köll, S., & Graham, N. (2020). Various versatile variances: An object-oriented implementation of clustered covariances in R. *Journal of Statistical Software, 95*(1), 1–36. https://doi.org/10.18637/jss.v095.i01

9. Annexes

❖ **Annex 1.** Socio-economic indicators at country-level of the Informal Economy.

Countries	***% of informal economy***	***informal output (% of GDP)***	***Tax burden (2022)***	***PIB per capita (2022)***	***Unemployment (2022)***	***Gini Index***	***Mean years of schooling (2022)***	***HDI rank (2022)***	***Corruption perceptions index (2022)***
Bangladesh	*84.9*	*27.0*	*7.6*	*$8,450.50*	*4.6*	*33.4*	*7.4*	*130*	*25*
Burkina Faso	*93.8*	*31.2*	*17.7*	*$2,641.90*	*5.4*	*37.4*	*2.3*	*185*	*42*
Togo	*92.3*	*31.2*	*14.8*	*$2,852.10*	*2.0*	*37.9*	*5.6*	*160*	*30*
Mali	*95.4*	*33.8*	*14.2*	*$2,525.20*	*2.4*	*35.7*	*1.6*	*188*	*28*
Senegal	*95.1*	*37.9*	*19.4*	*$4,529.10*	*2.8*	*36.2*	*2.9*	*170*	*43*
Gambia	*84.1*	*39.4*	*20.3*	*$3,067.30*	*6.1*	*38.8*	*4.5*	*174*	*34*
Zambia	*83.8*	*39.8*	*16.8*	*$3,840.70*	*6.0*	*51.5*	*7.3*	*154*	*33*
Guinea-Bissau	*94.8*	*42.7*	*8.6*	*$2,685.80*	*2.7*	*33.4*	*3.7*	*179*	*21*
Benin	*96.3*	*45.4*	*17.1*	*$3,844.20*	*1.7*	*34.4*	*3.1*	*173*	*43*
Viet Nam	*67.7*	*11.6*	*24.8*	*$13,852.10*	*1.5*	*36.1*	*8.5*	*108*	*42*
Indonesia	*81.2*	*15.3*	*11.6*	*$14,285.00*	*3.5*	*35.5*	*8.6*	*113*	*34*
Mongolia	*41.9*	*15.4*	*16.9*	*$16,402.00*	*6.2*	*31.4*	*9.4*	*99*	*33*
Chile	*27.4*	*16.1*	*17.7*	*$30,911.40*	*8.2*	*43.0*	*11.1*	*42*	*67*
India	*88.8*	*16.7*	*6.7*	*$9,153.90*	*4.8*	*32.8*	*6.6*	*135*	*40*

Costa Rica	*37.1*	*21.1*	*13.9*	*$25,924.70*	*11.3*	*47.2*	*8.8*	*60*	*54*
Argentina	*50.4*	*21.2*	*11.1*	*$29,597.70*	*6.8*	*40.7*	*11.1*	*47*	*38*
Mexico	*56.3*	*27.4*	*14.3*	*$23,345.40*	*3.3*	*43.5*	*9.2*	*83*	*31*
Bulgaria	*1.2*	*27.9*	*20.5*	*$34,855.90*	*4.3*	*38.4*	*11.4*	*70*	*43*
Ecuador	*68.1*	*28.1*	*13.0*	*$15,280.10*	*3.8*	*45.5*	*9*	*90*	*36*
Egypt	*71.3*	*29.3*	*12.5*	*$17,526.80*	*7.3*	*31.9*	*9.8*	*103*	*30*
Colombia	*56.5*	*30.2*	*17.7*	*$20,677.60*	*10.5*	*54.8*	*8.9*	*89*	*39*
Jamaica	*54.6*	*31.7*	*25.7*	*$10,790.10*	*4.1*	*40.2*	*9.2*	*114*	*44*
Kyrgyzstan	*58.8*	*35.5*	*21.3*	*$6,577.50*	*4.0*	*26.4*	*12*	*116*	*27*
Paraguay	*66.8*	*35.8*	*10.2*	*$16,347.00*	*6.8*	*45.1*	*8.9*	*99*	*28*
El Salvador	*66.5*	*40.8*	*20.6*	*$11,858.10*	*3.0*	*38.8*	*7.2*	*127*	*33*
Peru	*71.6*	*44.8*	*15.9*	*$16,657.50*	*3.9*	*40.3*	*10*	*86*	*36*
Panama	*56.1*	*46.0*	*7.5*	*$36,245.10*	*8.1*	*50.9*	*10.7*	*57*	*36*
Bolivia	*84.5*	*53.8*	*39.9*	*$10,372.00*	*3.6*	*40.9*	*9.8*	*119*	*31*
Turkiye	*27.3*		*18.5*	*$39,101.10*	*10.5*	*45.3*	*8.8*	*48*	*36*
Tonga	*59.7*		*23.8*	*$7,393.80*	*2.3*	*27.1*	*10.9*	*95*	
Luxembourg	*3.5*	*9.1*	*27.8*	*$144,872.70*	*4.6*	*29.1*	*13*	*19*	*77*
Netherlands	*2.4*	*12.6*	*24.8*	*$77,152.20*	*3.5*	*26.3*	*12.6*	*11*	*80*

France	*3.3*	*14.1*	*23.1*	*$56,181 .00*	*7.3*	*29.8*	*11.7*	*27*	*72*
Germany	*2*	*15.0*	*11.0*	*$67,589 .80*	*3.1*	*29.0*	*14.3*	*7*	*79*
Slovak Republic	*5.1*	*16.1*	*19.4*	*$41,111 .70*	*6.1*	*21.2*	*13*	*45*	*53*
Denmark	*2.8*	*16.4*	*31.4*	*$77,400 .30*	*4.4*	*27.7*	*13*	*8*	*90*
Spain	*2*	*20.7*	*15.0*	*$50,496 .60*	*12.9*	*32.0*	*10.6*	*28*	*60*
Poland	*7.7*	*23.3*	*18.1*	*$46,077 .40*	*2.8*	*26.3*	*13.2*	*35*	*55*
Malta	*0.7*	*24.9*	*21.8*	*$60,352 .50*	*2.9*	*31.1*	*12.2*	*21*	*51*
Cyprus	*1.8*	*25.2*	*24.1*	*$56,019 .40*	*6.8*	*29.4*	*12.4*	*29*	*52*
Latvia	*2.4*	*26.0*	*16.8*	*$39,965 .30*	*6.8*	*34.3*	*13.3*	*39*	*59*
Italy	*3.5*	*26.1*	*24.9*	*$56,218 .80*	*8.1*	*32.7*	*10.7*	*30*	*56*
Greece	*4.1*	*26.2*	*26.6*	*$38,969 .00*	*12.4*	*31.4*	*11.4*	*33*	*52*
Romania	*2.5*	*26.6*	*16.2*	*$42,218 .00*	*5.6*	*32.0*	*11.4*	*52*	*46*
Lithuania	*3.1*	*27.9*	*21.4*	*$50,498 .40*	*6.0*	*36.2*	*13.5*	*36*	*62*
Croatia	*1.4*	*28.7*	*21.5*	*$41,959 .50*	*7.0*	*28.5*	*12.3*	*37*	*50*
Finland	*2.1*	*28.7*	*25.3*	*$61,344 .60*	*6.7*	*26.6*	*12.9*	*11*	*87*
Republic of Moldova	*52.2*	*42.9*	*18.8*	*$16,381 .30*	*0.9*	*25.7*	*11.8*	*81*	*39*
Serbia	*17.8*		*23.9*	*$26,242 .30*	*8.4*	*32.0*	*11.5*	*60*	*36*

❖ **Annex 2.** No_Schooling1

	Dependent variable:
	Informality
Tax	-0.467
	(0.304)
PIB	-0.0001
	(0.0001)
Unemployment	-1.661**
	(0.770)
Gini	0.899***
	(0.292)
HDI	0.484***
	(0.065)
Corruption	-0.047
	(0.188)
Constant	-6.100
	(17.101)
Observations	48
R2	0.901
Adjusted R2	0.886
Residual Std. Error	12.205 (df = 41)
F Statistic	62.100*** (df = 6; 41)
Note:	*p<0.1; **p<0.05; ***p<0.01

❖ **Annex 3.** No_Schooling2

	Dependent variable:
	`Informality by output`
Tax	0.482**
	(0.211)
PIB	-0.0001
	(0.0001)
Unemployment	-0.303
	(0.543)
Gini	0.284
	(0.203)
HDI	0.014
	(0.045)
Corruption	-0.224
	(0.135)
Constant	22.540*
	(11.834)
Observations	46
R2	0.462
Adjusted R2	0.379
Residual Std. Error	8.334 (df = 39)
F Statistic	5.573*** (df = 6; 39)
Note:	*p<0.1; **p<0.05; ***p<0.01

❖ **Annex 4.** Diagnosis model 1

Model 1	**Normality of errors**	**Homoscedasticity**	**Error autocorrelation**	**Multicollinearity**	**Influence points**	**Corrective action**
Full model	Strong deviation in the left tail	**Breusch-Pagan** p = 0,57 > 0,05 No heteroscedasticity	**Breusch-Godfrey** p = 0,09 > 0,05 No autocorr.	High collinearity in HDI (Schooling = 9)	14 and 41	HC3, RLM and selective exclusion
No_HDI	Heavy tails and light asymmetry	**Breusch-Pagan** p = 0.08 > 0.05 No heteroscedasticity	**Breusch-Godfrey** p = 0,04 < 0,05 Autocorr. detected	Values below 5, acceptable collinearity.	14 and 41	HC3, RLM and selective exclusion
No_Schooling	Deviation in the left tail	**Breusch-Pagan** p = 0.70 > 0.05 No heteroscedasticity	**Breusch-Godfrey** p = 0,10 > 0,05 No autocorr.	Values below 5, acceptable collinearity.	14 and 41	HC3, RLM and selective exclusion

❖ **Annex 5.** Diagnosis model 2

Model	**Normality of errors**	**Homoscedasticity**	**Error autocorrelation**	**Multicollinearity**	**Influence points**	**Corrective action**
Full model 2	Heavy tails and light asymmetry	**Breusch-Pagan** p = 0,69> 0,05 No heteroscedasticity	**Breusch-Godfrey** p = 0,54 > 0,05 No autocorr.	High collinearity in HDI (Schooling = 10.2)	14, 41 and 27	HC3, RLM and selective exclusion
No_HDI 2	Heavy tails and light asymmetry	**Breusch-Pagan** p = 0.64 > 0.05 No heteroscedasticity	**Breusch-Godfrey** p = 0,56 > 0,05 No autocorr.	Values below 5, acceptable collinearity.	14, 41, 27 and 34	HC3, RLM and selective exclusion
No_Schooling 2	Heavy tails and light asymmetry	**Breusch-Pagan** p = 0.58 > 0.05 No heteroscedasticity	**Breusch-Godfrey** p = 0,55 > 0,05 No autocorr.	Values below 5, acceptable collinearity.	14, 41, 27 and 34	HC3, RLM and selective exclusion

❖ **Annex 6.** HC3 (Full model1)

```
> coeftest(fullmodel, vcov = vcovHC(fullmodel, type = "HC3"))

t test of coefficients:

                  Estimate  Std. Error t value Pr(>|t|)  
(Intercept)     -1.1493e+00  6.1592e+01 -0.0187  0.98521  
Tax             -4.5482e-01  6.1248e-01 -0.7426  0.46207  
`PIB per capita` -7.9667e-05  5.5849e-04 -0.1426  0.88728  
Unemployment    -1.7078e+00  7.8508e-01 -2.1753  0.03558 *
`Gini index`     8.9083e-01  3.3800e-01  2.6356  0.01190 *
Schooling       -2.5739e-01  2.1202e+00 -0.1214  0.90398  
HDI              4.6606e-01  2.7165e-01  1.7156  0.09397 .
Corruption      -5.7525e-02  2.3743e-01 -0.2423  0.80980  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

❖ **Annex 7.** HC3 (No_HDI1)

```
> coeftest(No_HDI, vcov = vcovHC(No_HDI, type = "HC3"))

t test of coefficients:

              Estimate  Std. Error t value  Pr(>|t|)    
(Intercept)  1.1154e+02  2.4040e+01  4.6397 3.545e-05 ***
Tax         -2.0408e-01  6.8709e-01 -0.2970 0.7679445    
PIB         -2.9229e-04  6.3534e-04 -0.4601 0.6479092    
Unemployment -3.1774e+00  8.4994e-01 -3.7383 0.0005664 ***
Gini         7.9412e-01  3.5693e-01  2.2248 0.0316548 *  
Schooling   -5.2613e+00  1.8842e+00 -2.7923 0.0079146 ** 
Corruption  -3.3656e-01  4.6456e-01 -0.7245 0.4728882    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

❖ **Annex 8.** HC3 (No_Schooling1)

```
> coeftest(No_Schooling, vcov = vcovHC(No_Schooling, type = "HC3"))

t test of coefficients:

              Estimate  Std. Error t value  Pr(>|t|)    
(Intercept)  -6.0996e+00  2.3814e+01 -0.2561 0.7991297    
Tax          -4.6660e-01  6.1773e-01 -0.7554 0.4543552    
PIB          -7.5219e-05  4.3034e-04 -0.1748 0.8621048    
Unemployment -1.6613e+00  6.4381e-01 -2.5804 0.0135491 *  
Gini          8.9884e-01  2.9183e-01  3.0800 0.0036870 ** 
HDI           4.8449e-01  1.2575e-01  3.8528 0.0004028 ***
Corruption   -4.6635e-02  2.4758e-01 -0.1884 0.8515190    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

❖ **Annex 9.** HC3 (Full model2)

```
> coeftest(fullmodel2, vcov = vcovHC(fullmodel2, type = "HC3"))

t test of coefficients:

               Estimate  Std. Error t value Pr(>|t|)
(Intercept)   1.6565e+01  2.9589e+01  0.5598   0.5789
Tax           4.6644e-01  2.9594e-01  1.5761   0.1233
PIB          -8.1102e-05  2.8934e-04 -0.2803   0.7808
Unemployment -2.5627e-01  6.1240e-01 -0.4185   0.6780
Gini          2.9162e-01  2.4470e-01  1.1918   0.2407
Schooling     3.1122e-01  1.0243e+00  0.3038   0.7629
HDI           3.6773e-02  1.3594e-01  0.2705   0.7882
Corruption   -2.0774e-01  1.6113e-01 -1.2892   0.2051
```

❖ **Annex 10.** HC3 (No_HDI2)

```
> coeftest(No_HDI2, vcov = vcovHC(No_HDI2, type = "HC3"))

t test of coefficients:

               Estimate  Std. Error t value Pr(>|t|)  
(Intercept)   2.5458e+01  1.1361e+01  2.2408  0.03081 *
Tax           4.8939e-01  2.8680e-01  1.7064  0.09590 .
PIB          -9.5416e-05  1.2473e-04 -0.7650  0.44889  
Unemployment -3.5347e-01  4.7973e-01 -0.7368  0.46565  
Gini          2.8551e-01  2.3560e-01  1.2118  0.23287  
Schooling    -9.0607e-02  5.7300e-01 -0.1581  0.87517  
Corruption   -2.3402e-01  1.5362e-01 -1.5233  0.13575  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

❖ **Annex 11.** HC3 (No_Schooling2)

```
> coeftest(No_Schooling2, vcov = vcovHC(No_Schooling2, type = "HC3"))

t test of coefficients:

               Estimate  Std. Error t value Pr(>|t|)
(Intercept)  22.54012046 14.09579073  1.5991   0.1179
Tax           0.48239410  0.29130777  1.6560   0.1058
PIB          -0.00008542  0.00022672 -0.3768   0.7084
Unemployment -0.30308061  0.54148276 -0.5597   0.5789
Gini          0.28396104  0.23808779  1.1927   0.2402
HDI           0.01419501  0.06938710  0.2046   0.8390
Corruption   -0.22383139  0.16862611 -1.3274   0.1921
```

❖ **Annex 12.** RLM (Full model1)

```
================================================
                       Dependent variable:
                   -----------------------------
                           Informality
------------------------------------------------
Tax                          -0.619**
                             (0.295)

PIB                          -0.0001
                             (0.0001)

Unemployment                 -1.851**
                             (0.779)

Gini                         0.921***
                             (0.279)

Schooling                    -0.074
                             (1.565)

HDI                          0.467***
                             (0.128)

Corruption                   -0.035
                             (0.190)

Constant                     -1.428
                             (34.150)

------------------------------------------------
Observations                    48
Residual Std. Error        9.784 (df = 40)
================================================
Note:              *p<0.1; **p<0.05; ***p<0.01
```

❖ **Annex 13.** RLM (Full model2)

```
================================================
                       Dependent variable:
                   -----------------------------
                    `Informality by output`
------------------------------------------------
Tax                          0.596***
                             (0.192)

PIB                          -0.0001
                             (0.0001)

Unemployment                 -0.375
                             (0.499)

Gini                         0.210
                             (0.179)

Schooling                    0.196
                             (1.054)

HDI                          0.026
                             (0.086)

Corruption                   -0.230*
                             (0.130)

Constant                     21.657
                             (22.708)

------------------------------------------------
Observations                    46
Residual Std. Error        5.980 (df = 38)
================================================
Note:              *p<0.1; **p<0.05; ***p<0.01
```

❖ **Annex 14.** No Outliers (Full model1)

	Dependent variable: Informality
Tax	-1.028***
	(0.346)
PIB	-0.001**
	(0.0003)
Unemployment	-1.953**
	(0.762)
Gini	0.570**
	(0.280)
Schooling	-2.331
	(1.644)
HDI	0.180
	(0.167)
Corruption	0.191
	(0.196)
Constant	69.816*
	(40.498)
Observations	46
R2	0.922
Adjusted R2	0.908
Residual Std. Error	10.927 (df = 38)
F Statistic	64.354*** (df = 7; 38)
Note:	*p<0.1; **p<0.05; ***p<0.01

❖ **Annex 15.** No Outliers (Full model2)

	Dependent variable: \`Informality by output\`
Tax	0.405
	(0.267)
PIB	0.0001
	(0.0002)
Unemployment	-0.104
	(0.554)
Gini	0.159
	(0.217)
Schooling	0.600
	(1.245)
HDI	0.118
	(0.125)
Corruption	-0.186
	(0.154)
Constant	6.265
	(30.284)
Observations	43
R2	0.409
Adjusted R2	0.291
Residual Std. Error	7.866 (df = 35)
F Statistic	3.465*** (df = 7; 35)
Note:	*p<0.1; **p<0.05; ***p<0.01