

Measuring Informal GDP Using Survey Data

Nicola Corbellini*

August 18, 2025

Abstract

This paper estimates the share of informal GDP in 26 developing countries over a period that spans up to 2004–2023. The focus is on the concept of legal informality, defined as the activities of enterprises that are not formally registered. I propose a simple framework that formally characterizes informal GDP. I use data from various sources, including the World Bank Enterprise Survey and the International Labour Organization (ILO), to match the conditions derived from the framework. The resulting informality series conform to patterns documented in previous research, such as a negative correlation with GDP per capita and mild anticyclical behavior. Because I adopt a narrower definition of informality, the estimated levels tend to be lower than those reported in previous cross-country studies.

1 Introduction

A large share of economic activity in developing countries occurs within the informal sector. However, consistently measuring informality across countries presents several challenges. First, surveys often adopt varying definitions of informality. For example, some focus on different units, such as firms or employees, or emphasize different dimensions, such as the activities of unregistered firms rather than unreported activities of otherwise

*Department of Economics, Pennsylvania State University. Email: nzc5436@psu.edu.

I thank the Enterprise Analysis Unit of the Development Economics Global Indicators Group of the World Bank for the data, and Josh Wimpey for useful clarifications regarding the World Bank Enterprise Survey microdata. All errors are my own.

formal firms. Second, existing measures of informality rely on two main approaches, each with its own strengths and limitations. Direct methods, such as employment and enterprise surveys, ensure representativeness within a country and do not rely on strong assumptions, but are available for only a limited number of countries and years and are less suitable for cross-country comparisons. In contrast, indirect methods facilitate comparability across countries but rely on statistical or economic models characterized by strong assumptions about the relationship between informality and other economic variables.

This paper adopts a combination of direct and indirect methods to estimate the share of informal GDP in 26 developing countries over a period that spans up to 2004–2023, depending on the country. I focus on the concept of *legal informality*, defined as the economic activities of enterprises that are not formally registered. Compared to broader definitions, this approach excludes informal activities conducted by otherwise formal enterprises. From a methodological perspective, I introduce a simple framework that formally characterizes the concept of informal GDP. A key contribution of this paper is the use of statistics from the World Bank Enterprise Survey (WBES) to match some of the conditions derived from the proposed framework. Specifically, I use firm-level data to compute average sales and value added per worker in both the formal and informal sectors. This information is then combined with data on formal and informal employment from the International Labour Organization (ILO), along with other macroeconomic variables. The constructed series are designed so that fluctuations in informal GDP are primarily driven by changes in informal employment, as reported by the ILO. For each country, I derive two series of informal GDP: one based on average sales per worker, the other on average value added per worker. Since the former likely underestimates informal GDP, while the latter likely overestimates it, these two series can be interpreted as lower and upper bounds, respectively.

The estimated measures of informal GDP are negatively correlated with GDP per capita. However, some low-income countries exhibit relatively low levels of informality. This finding is not in contradiction with previous research, which has highlighted that factors beyond the stage of economic development, such as institutions or tax enforcement,

also influence informality. Regarding the evolution of informality over time, two patterns stand out. First, some countries maintain a stable level of informality throughout the period considered, while others experience a gradual decline. Second, the series exhibit a mild anticyclical behavior, aligning with previous studies documenting increases in informality during recessions. Finally, the estimated levels of informal GDP are somewhat lower than those reported in previous cross-country studies. This discrepancy is likely due to the narrower concept of informality adopted here, which excludes, for example, unreported activities conducted by formal enterprises.

The remainder of the paper is organized as follows. Section 2 provides an overview of the definitions and measurement approaches related to informality. Section 3 introduces the framework used to estimate informal GDP. Section 4 describes the data used. Section 5 discusses the main features of the estimated informality series. Section 6 concludes.

Related literature. This paper contributes to the literature estimating the size of the informal economy. Much of the existing research focuses on specific dimensions of informality or on individual countries. For example, [Orsi, Raggi, and Turino \(2014\)](#) develop a *dynamic general equilibrium* (DGE) model with an informal sector to estimate the size and evolution of the underground economy in Italy, while [Di Nola et al. \(2021\)](#) analyze the size and effects of tax evasion in the US.

Some studies, however, provide broader cross-country estimates of informality. [Schneider, Buehn, and Montenegro \(2010\)](#) adopt a *multiple indicators multiple causes* (MIMIC) approach to measure informality across 162 countries from 1999 to 2007. The main idea behind the MIMIC method is to retrieve an unobserved variable (share of informal GDP) using structural equations and the sample covariances between observed variables (e.g., self-employment and other macroeconomic variables). [Elgin and Öztunali \(2012\)](#) develop a DGE model in which a representative household allocates labor between the formal and informal sectors. The model is then calibrated using national account statistics to estimate

the size of the shadow economy in 161 countries from 1950 to 2009.¹ While this paper derives measures of informality for a narrower set of countries and years, it offers three key advantages over previous works. First, the base-year *levels* are derived from firm-level data rather than alternative methods such as the currency demand approach. Second, *changes* over time are mainly driven by changes in labor force survey statistics, rather than shifts in some of the national account statistics used to calibrate the model. Third, I introduce a quantitative estimate of a narrower concept of informality, namely legal informality, which is often conflated with other measures in earlier studies. Another relevant work is [Pappadà and Rogoff \(2025\)](#), who propose a measure of the informal economy based on the discrepancy between national account data and VAT tax records. The authors derive measures of informality for countries across the European Union during the period 1999–2020. Their approach, while more transparent and less dependent on model assumptions, focuses on a smaller group of developed economies and emphasizes a particular dimension of informality, namely tax evasion.

This paper also relates to previous work describing the main characteristics of informality. For example, [La Porta and Shleifer \(2014\)](#) document its prevalence in low- and middle-income countries and its negative correlation with GDP per capita. Along with other country-specific studies such as [Ulyssea \(2018\)](#), they also highlight the smaller size and lower productivity of informal businesses. This aligns with the paper’s finding that the share of employment outside the formal sector exceeds the share of GDP produced outside the formal sector. Finally, the anticyclical behavior of informality observed in the estimated series has been previously documented, for example, by [Leyva and Urrutia \(2020\)](#) and [Pappadà and Rogoff \(2025\)](#).

1. For both [Schneider, Buehn, and Montenegro \(2010\)](#) and [Elgin and Öztunali \(2012\)](#), updated series up to 2018 are available in [Elgin et al. \(2021\)](#).

2 Informality: Definitions and Measurements

Definitions. I follow the categorization proposed by the [World Bank \(2020\)](#), which distinguishes between three degrees of informality: (i) *legal informality*, referring to the business registration status, (ii) *fiscal informality*, concerning compliance with taxes and regulations, and (iii) *labor informality*, reflecting the type of contracts and benefits available to employees. The measure of informality used in this paper focuses on legal informality. Specifically, I employ data from the ILO and the WBES to collect employment and other balance sheet information of unregistered businesses operating in the informal sector. According to the [ILO \(n.d.\)](#), an informal sector business or enterprise (i) is unregistered and unincorporated, (i.e., it is not constituted as a legal entity separate from its owners) and (ii) it sells at least some of the goods or services it produces. This definition is consistent with the one adopted by the World Bank in its Informal Sector Enterprise Survey (ISES). In fact, the ISES focuses on businesses that lack a business license, do not exist on a business registry, and/or are not registered with the relevant tax authority ([Aga et al. \(2022\)](#)).

Measurements. There exist two main approaches to measuring informality: *direct methods* and *indirect methods*. Direct methods are collected from labor force surveys, household surveys, enterprise surveys, or a combination thereof. These approaches typically capture several dimensions of informality, such as business and employment informality. Their main advantage is that they require minimal assumptions. However, their infrequent implementation makes it difficult to obtain estimates that cover a broad range of countries and years. Moreover, cross-country comparisons are not always possible given the differences in the implementation of the surveys and/or the definitions of informality across countries.

Indirect methods, by contrast, involve statistical or economic models that link measures of informality with other observed economic variables. These methods can produce time series that are more comparable across countries and years, but they rely heavily on specific assumptions and are therefore subject to model misspecification.

As detailed in the next section, I combine direct methods, such as statistics derived from

ILO and World Bank data, with indirect methods, including the specification of an aggregate production function, to estimate informal GDP shares for a group of developing countries.

3 Informality Accounting

I consider a model economy in which aggregate output Y is the sum of output produced in the formal sector, Y_F , and output produced outside the formal sector, Y_O . In both sectors, there is a representative firm that transforms production inputs into output. Specifically, the formal sector's representative firm employs capital and labor to produce output, while the informal sector's representative firm employs labor only:

$$\begin{aligned} Y_F &= \theta_F K_F^\alpha N_F^{\gamma_F} \\ Y_O &= \theta_O N_O^{\gamma_O} \end{aligned}$$

Where K_F denotes formal sector's capital, N_F and N_O denote total employment in the formal and informal sector, α is the capital share in the formal sector, γ_F and γ_O are the labor share in the formal and informal sectors, and θ_F and θ_O denote TFP terms in the two sectors.

The objective is to derive the share of informal GDP S_O :

$$S_O = \frac{Y_O}{Y} = \frac{\theta_O N_O^{\gamma_O}}{\theta_F K_F^\alpha N_F^{\gamma_F} + \theta_O N_O^{\gamma_O}} \quad (1)$$

Note that θ_O and θ_F are the only unknowns in equation (1). I use employment series from ILO and capital series from the Penn World Table (PWT), while the production factor shares are estimated using data from the WBES. From this survey, I also compute country-level averages of sales per worker in the formal and informal sectors. Let the sales per worker averages be μ_F and μ_O . Therefore, for the years in which the WBES is conducted,

I use the following two equations to derive θ_F and θ_O :

$$\frac{Y_F}{N_F} = \theta_F K_F^\alpha N_F^{\gamma_F-1} = \mu_F \quad (2)$$

$$\frac{Y_O}{N_O} = \theta_O N_O^{\gamma_O-1} = \mu_O \quad (3)$$

This procedure allows me to derive estimates of informal GDP only for the years in which the WBES is conducted. I make the following assumptions to extend these measures backward and forward over time: (i) the estimated factor shares α , γ_F , and γ_O are constant over time, (ii) the growth rates of TFP in the two sectors $\delta\theta_{F,t} = \frac{\theta_{F,t}-\theta_{F,t-1}}{\theta_{F,t-1}}$ and $\delta\theta_O = \frac{\theta_{O,t}-\theta_{O,t-1}}{\theta_{O,t-1}}$ are such that $\delta\theta_F = \delta\theta_O$, and (iii) $\theta_{F,t}$ and $\theta_{O,t}$ satisfy the following equation for each year t in which employment data from the ILO are available:

$$\delta(\theta_{F,t} K_{F,t}^\alpha N_{F,t}^{\gamma_F} + \theta_{O,t} N_{O,t}^{\gamma_O}) = \delta(GDP_t) \quad (4)$$

Where $\delta(GDP_t)$ denotes the GDP growth rate $\frac{GDP_t - GDP_{t-1}}{GDP_{t-1}}$. GDP_t growth rates are taken from PWT and, for the most recent years, the World Bank. Equation (4) states that the growth rate of the TFP measures $\theta_{F,t}$ and $\theta_{O,t}$ is such that the model-implied growth rate of output equals the measured growth rate of GDP.

I then repeat the same steps using averages of value added per worker, in place of sales per worker, in the formal and informal sectors in equations (2) and (3). In Section 5, I present two distinct time series of informality for each country, one based on sales per worker and the other on value added per worker. While value added is a more appropriate measure than gross sales for assessing the relative contribution of the two sectors, firm-level data on intermediate purchases are sometimes missing. This limitation reduces both the size and representativeness of the samples in the value added-based series.

4 Data

WBES. I use data from the WBES to estimate the production function's factor shares and to compute average sales and value added per worker in the formal and informal sectors.² Specifically, I combine data from the standard WBES, which samples registered firms with at least five employees, and the ISES, which targets informal businesses.³ Both surveys focus exclusively on the non-agricultural sector. In addition, firm-level information is available on sector (manufacturing or services), number of employees, and annual sales. Sample weights are available for all observations in the formal sample, while they are available only for informal surveys conducted after 2016. Consequently, I assign equal weights to all observations within a country from informal surveys prior to 2016. Table 1 lists the number of firm-level observations, and the (weighted) average number of employees and yearly sales for each country, separately for the formal and informal sectors. Based on the available balance sheet data from the surveys, I compute value added differently across the two sectors. In the formal sector, value added is calculated by subtracting the value of intermediate goods purchases from annual sales. In the informal sector, where data on intermediate purchases are rarely available, I approximate value added by subtracting total electricity costs from annual sales. Table 2 shows the number of available observations along with (weighted) averages of the number of employees and value added for both sectors in each country. Note that the employment figures differ from those displayed in Table 1 because some observations are dropped when computing value added.

To estimate the formal sector's production function shares α and γ_F , I run the following country-specific regressions:

$$\ln y_i = \beta_0 + \alpha \ln k_i + \gamma_F \ln l_{F,i} + \rho_t + \rho_s + \varepsilon \quad (5)$$

2. The raw data used for this project are publicly available at <https://www.enterprisesurveys.org/en/enterprisesurveys>

3. The Appendix provides more details concerning WBES data.

Table 1: Summary Statistics (Sales).

Country	Year	Formal Sector			Informal Sector		
		Obs.	Average Employment	Average Sales (Th. 2017 USD)	Obs.	Average Employment	Average Sales (Th. 2017 USD)
Burkina Faso	2009	372	28.66	2247.4	102	3.6	20
Cameroon	2009	347	52.42	5509.7	121	3.04	14.9
Cote d'Ivoire	2009	508	16.11	1458.1	97	2.78	7.8
Madagascar	2009	355	59.56	1476.5	123	2.07	4.5
Mauritius	2009	383	44.93	5891.5	122	1.53	18.7
Nepal	2009	363	13.93	323.7	120	4.09	16.8
Niger	2009	137	16.93	3115.9	108	5.25	10.3
Angola	2010	323	34.06	308203.3	78	9.23	625
Argentina	2010	971	72.58	10866.4	308	1.59	6
Botswana	2010	232	61.61	8748.3	95	2.69	13.1
Guatemala	2010	435	96.1	5211.2	278	1.78	5.9
Mali	2010	248	19.92	3350.6	85	4.53	16.3
Rwanda	2011	190	36.14	1514.5	190	2.38	4.6
Egypt	2013	2443	39.18	1140.6	74	3.96	14
Myanmar	2014	540	31.6	303.7	276	5.41	39.5
Zimbabwe	2016	600	31.88	1325.7	499	1.96	8.7
Lao, People's Democratic Republic	2018	300	19.11	695.5	351	2	7.9
Zambia	2019	563	67.01	2112.7	659	1.87	1.6
Bangladesh	2022	966	114.84	1556.6	2851	6.99	14.9
Ghana	2022	664	27.83	904.9	2652	1.91	4.2
India	2022	9319	20.56	756.3	8471	1.64	5
Peru	2022	939	38.82	3368.5	1294	1.39	3.9
Sudan	2022	248	28.53	644764.4	1345	1.82	8.3
Cambodia	2023	519	94.89	1073.3	1277	1.57	3.6
Indonesia	2023	2237	23.14	1035.1	4896	1.52	5.7
Tanzania, United Republic of	2023	513	22.5	176	1377	1.54	1.4

Source: Own calculation based on WBES.

Where y_i , k_i , and $l_{F,i}$ denote value added, capital, and number of employees of formal firm i , respectively, while ρ_t and ρ_s are year and sector fixed effects.⁴ Value added and capital are expressed in 2017 USD.

For the informal sector, I run the following specification:

$$\ln y_i = \beta_0 + \gamma_O \ln l_{O,i} + \rho_t + \rho_s + \varepsilon \quad (6)$$

Where $l_{O,i}$ denotes the number of employees of informal firm i . I can then obtain country-specific estimates of α , γ_F , and γ_O that are used in equations (2) - (4)

ILO. I use data from the International ILO to obtain measures of formal and informal employment. For informal employment, I rely on the *number of employees outside the formal*

4. Year fixed effects are included in the regressions since the survey year does not always coincide with the year of operations.

Table 2: Summary Statistics (Value Added).

Country	Year	Formal Sector			Informal Sector		
		Obs.	Average Employment	Average Value Added (Th. 2017 USD)	Obs.	Average Employment	Average Value Added (Th. 2017 USD)
Burkina Faso	2009	52	38.24	1806.4	97	3.71	19.7
Cameroon	2009	90	91.94	10664.7	113	3	14.6
Cote d'Ivoire	2009	158	25.63	2056.5	89	2.73	7.8
Madagascar	2009	149	107.9	535.6	120	2.08	4.4
Mauritius	2009	138	67.61	2174.6	120	1.54	15.8
Nepal	2009	240	13.71	68.2	120	4.09	16.5
Niger	2009	35	18.44	272.1	93	5.27	7.6
Angola	2010	129	44.62	171004.8	51	7.82	761
Argentina	2010	658	81.86	7975.9	246	1.51	5.7
Botswana	2010	71	50.9	1567.6	85	2.74	13.5
Guatemala	2010	241	77.57	2276.3	254	1.78	5.9
Mali	2010	40	49.97	399.1	65	4.77	13.9
Rwanda	2011	38	83.03	1440.1	151	2.47	5
Egypt	2013	1632	48.54	818.8	74	3.96	14
Myanmar	2014	266	68.71	328.6	270	5.4	39.6
Zimbabwe	2016	280	42.26	1885.5	307	2.01	6.6
Lao, People's Democratic Republic	2018	121	30.2	253.1	283	1.79	8
Zambia	2019	155	108.17	1256.9	162	1.83	2.3
Bangladesh	2022	506	167.26	681	1941	9.88	19
Ghana	2022	162	44.11	646.7	1237	2.25	4.1
India	2022	5209	29.11	652.6	3872	1.77	4.5
Peru	2022	354	58.63	2188.6	415	1.52	4
Sudan	2022	18	52.42	2833.3	756	1.84	9.4
Cambodia	2023	260	346.14	2333	926	1.61	2.8
Indonesia	2023	865	39.26	970.3	3377	1.53	5.5
Tanzania, United Republic of	2023	150	18.65	118.6	710	1.69	1.6

Source: Own calculation based on WBES.

sector in non-agriculture for individuals aged 15 or more. This measure includes workers employed in at least one informal sector enterprise, defined as an unregistered and unincorporated enterprise (there is no legal entity separated from its owners) that sells at least some of the goods or services it produces. Note that this definition is consistent with the one used in the WBES.

The ILO harmonizes informality series by processing microdata files from national household survey using a consistent navigational path, ensuring that the resulting statistics are comparable across countries. However, I implement some adjustments to the ILO time series to address a few gaps and inconsistencies. First, I discard series whose figures are markedly inconsistent with preceding or subsequent surveys. In nearly all such cases, the survey source differs from the others available for the same country (e.g., a household survey instead of labor force surveys). Second, informal employment data are typically available only for a subset of years. When this occurs, I linearly interpolate employment

values for the missing years using data from the available years. The graphs presented in Section 5 clearly distinguish between years with observed employment data and those with interpolated values.

To compute formal sector employment, I subtract employment outside the formal sector from *total employment* in non-agriculture sector for individuals aged 15 or more.⁵

PWT and World Bank. I use data from the PWT for the time series on capital and GDP.⁶ Since the most recent version of the PWT includes data up to 2019, I extend the series using data from the World Bank. Specifically, for the capital series, I update the PWT data by adding investment, measured as gross capital formation, and subtracting capital depreciation:

$$K_t = (1 - \delta)K_{t-1} + I_t, \quad \text{for } t \geq 2020$$

Where K_t denotes capital, I_t denotes investment, while δ is the country-specific average depreciation rate between 2002 and 2019 taken from PWT.

However, capital and GDP series from PWT and the World Bank include the agricultural sector, which may lead to inconsistencies with the WBES data that focus exclusively on the non-agricultural sector. For this reason, I scale the GDP series by multiplying it by $(1-S_A)$, where S_A is the share of agricultural GDP from the WB, and I apply the same adjustment to the capital series. This approach assumes that the share of aggregate capital in the non-agricultural sector is equal to the non-agricultural share of GDP.

5 Results

Figures 1 to 4 display the informality series for 26 countries. The solid black lines represent the share of employees operating outside the formal sector, which is taken from ILO as the ratio between employment outside the formal sector and total employment in

5. The raw total and informal employment data are available at <https://ilostat.ilo.org/data/>.

6. PWT data are taken from [Feenstra, Inklaar, and Timmer 2015](#), available for download at www.ggdnet.net/pwt.

non-agriculture for individuals aged 15 or more. The dashed blue lines show the share of fiscal informality calculated using average sales per worker in the two sectors on the right-hand side of equations (2) and (3). The dotted red lines, instead, reflect calculations based on average value added per worker. Finally, diamonds indicate years for which informal employment data are available from ILO. Interpolated data are used when direct observations are unavailable. The following terminology is adopted throughout this section: the solid black lines are referred to as *informal employment*, the dashed blue lines as *sales-based informal GDP*, and the dotted red lines as *VA-based informal GDP*.

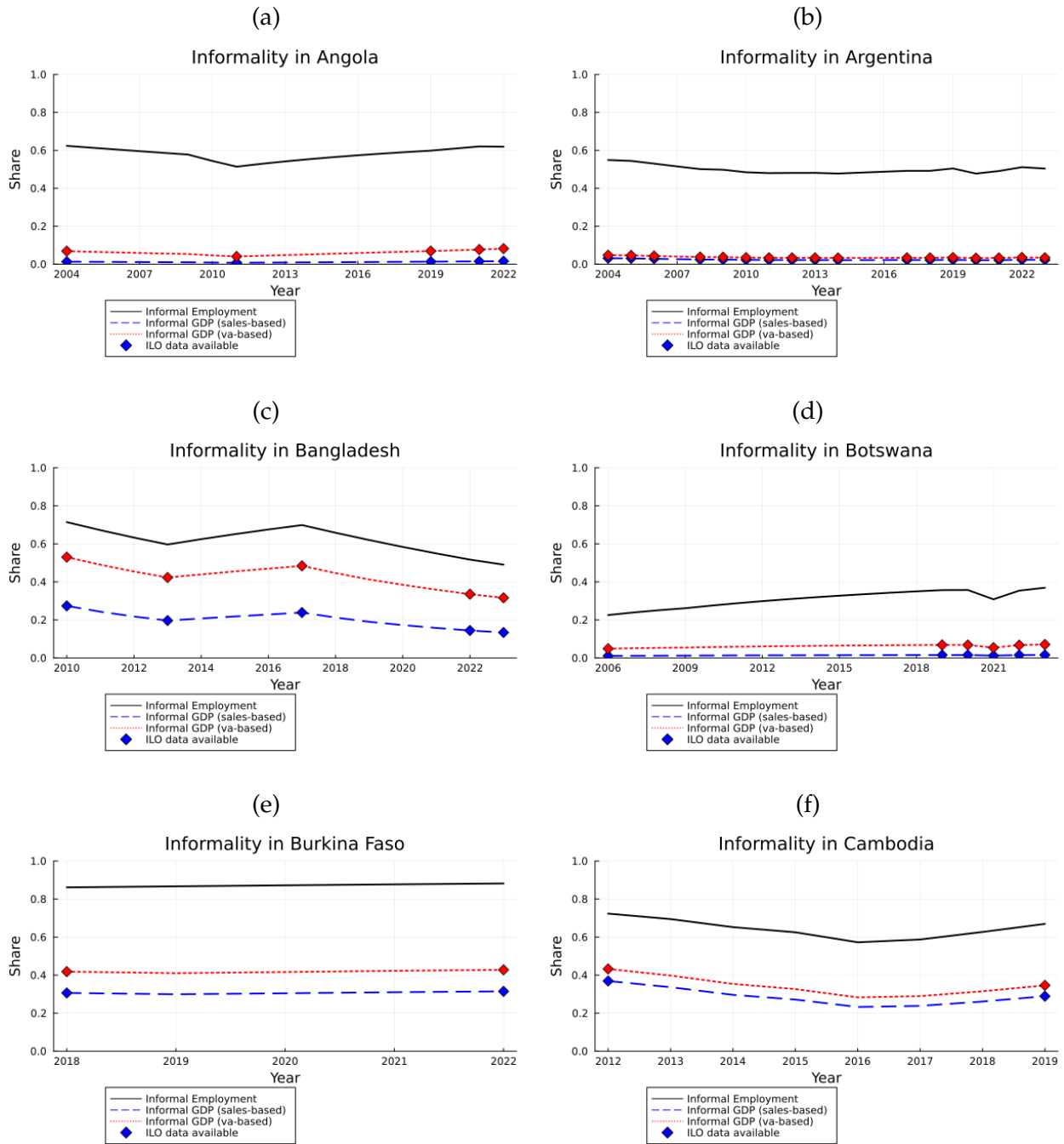
The following results stand out. First, in all countries except Myanmar, the share of informal employment is higher than the share of informal GDP (both sales- and VA-based), reflecting the higher average productivity of formal sector businesses. However, the magnitude of this discrepancy varies across countries. For instance, in Argentina and Peru, informal GDP accounts for less than 20%, while informal employment hovers around or above 50%. In contrast, several African and South/Southeast Asian countries exhibit narrower gaps between informal employment and informal GDP.

Second, sales-based informal GDP is generally lower than VA-based informal GDP. This pattern reflects a more intensive use of intermediate purchases by formal firms. Sales-based informal GDP tends to overstate the relative importance of the formal sector by not accounting for the relative higher share of intermediate purchases. Conversely, since data on intermediate purchases for informal businesses are rarely available (with electricity expenses used as a proxy when available), value added in the informal sector is likely to be overestimated. The two measures can be thought of as lower (sales-based) and upper (VA-based) bounds of the share of informal GDP.

Third, variations in informal GDP over time closely track variations in informal employment. While this feature partly occurs by construction, it represents an improvement over traditional measures of informal GDP, whose variations are generally driven by other macroeconomic statistics.

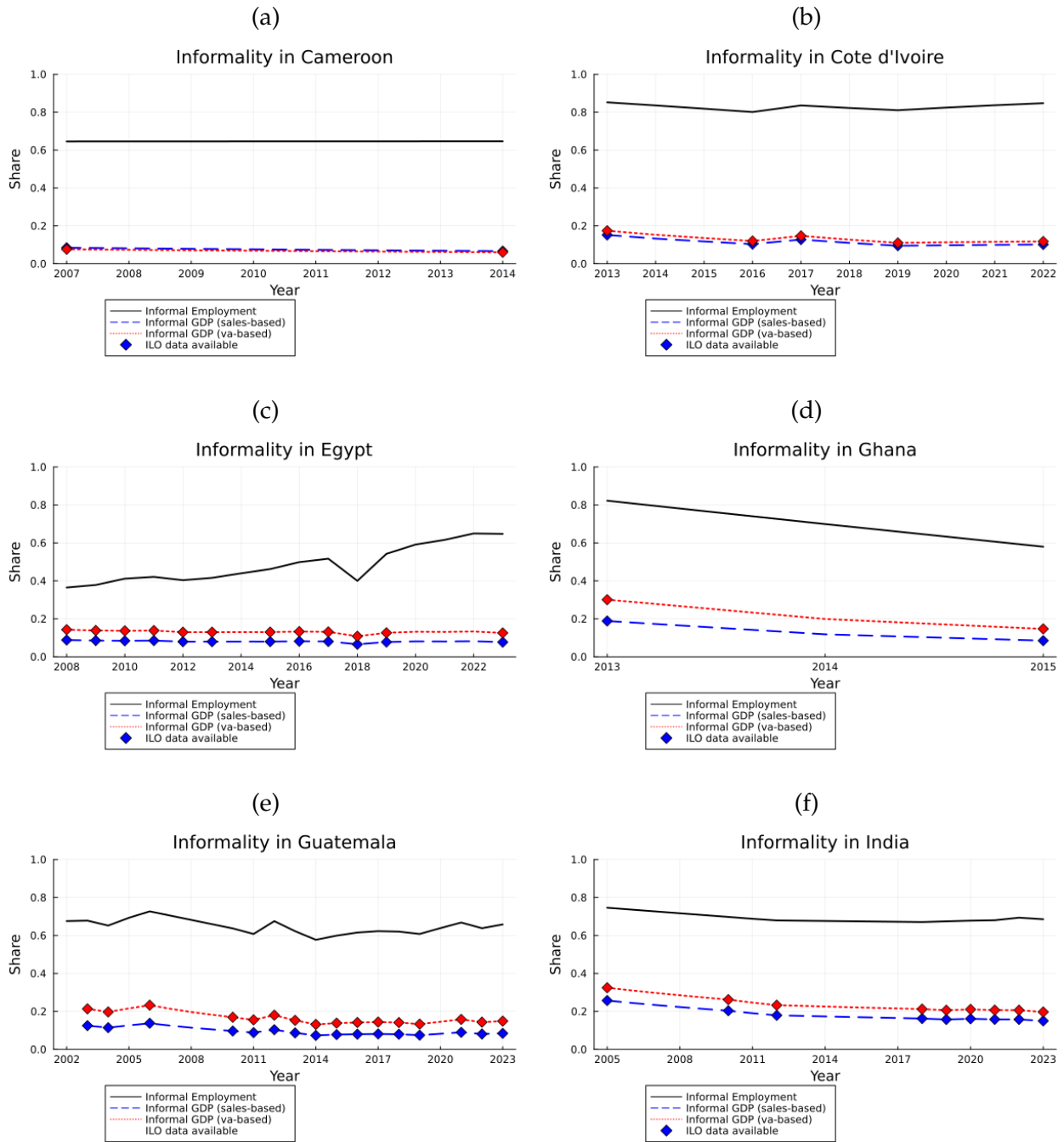
Finally, excluding short-term fluctuations, most countries display relatively steady informal GDP shares over the period considered, although a gradual decline is observed in a subset of countries.

Figure 1: Measures of Informality for Selected Countries



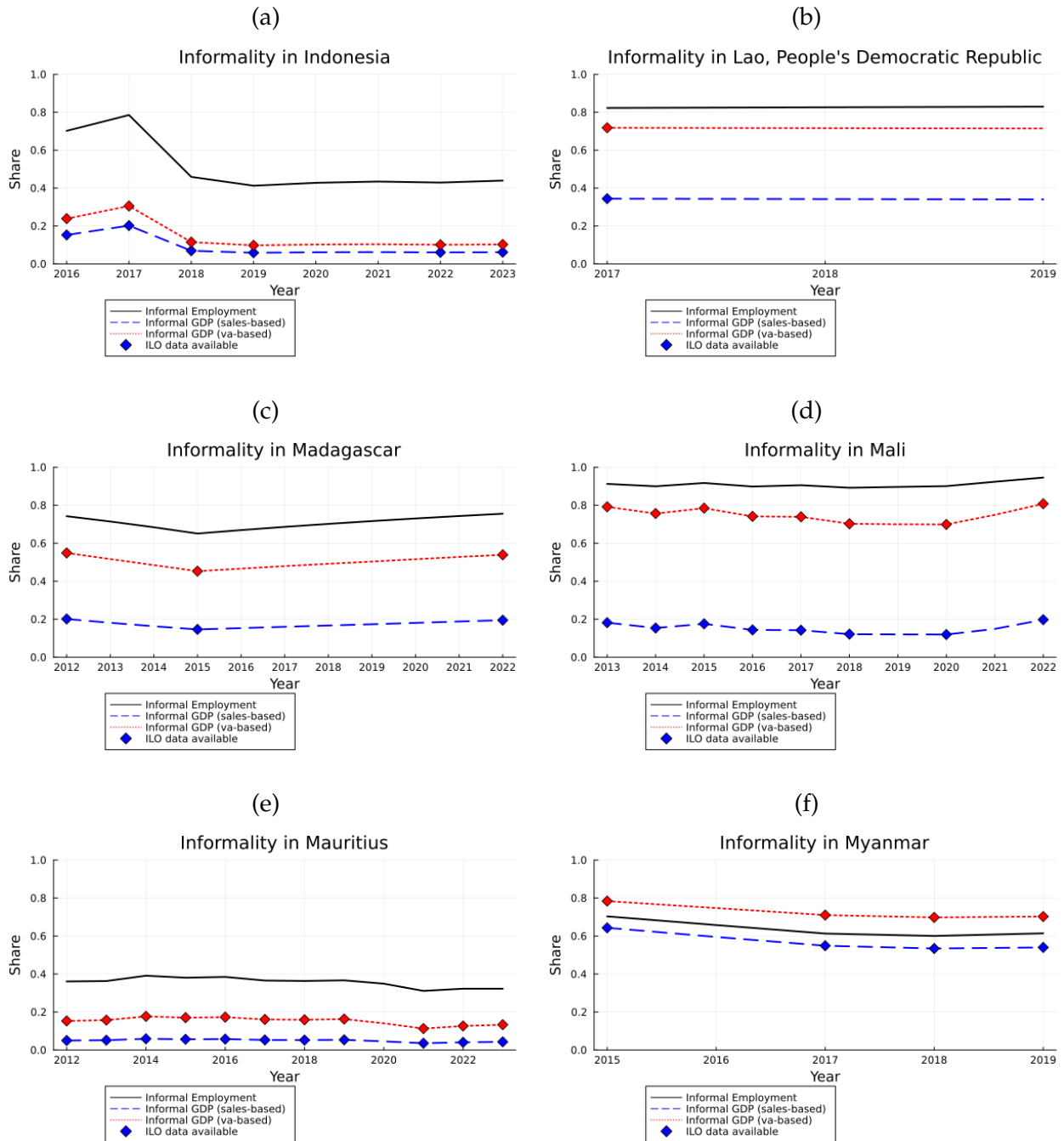
The solid black line represents the share of employment operating outside the formal sector. The dashed blue line is the share of fiscal informality based on sales per worker calculations, while the dotted red line is based on value added per worker calculations. Diamonds indicate years in which informal employment data are available from ILO. When this information is not available, interpolated data are used.

Figure 2: Measures of Informality for Selected Countries



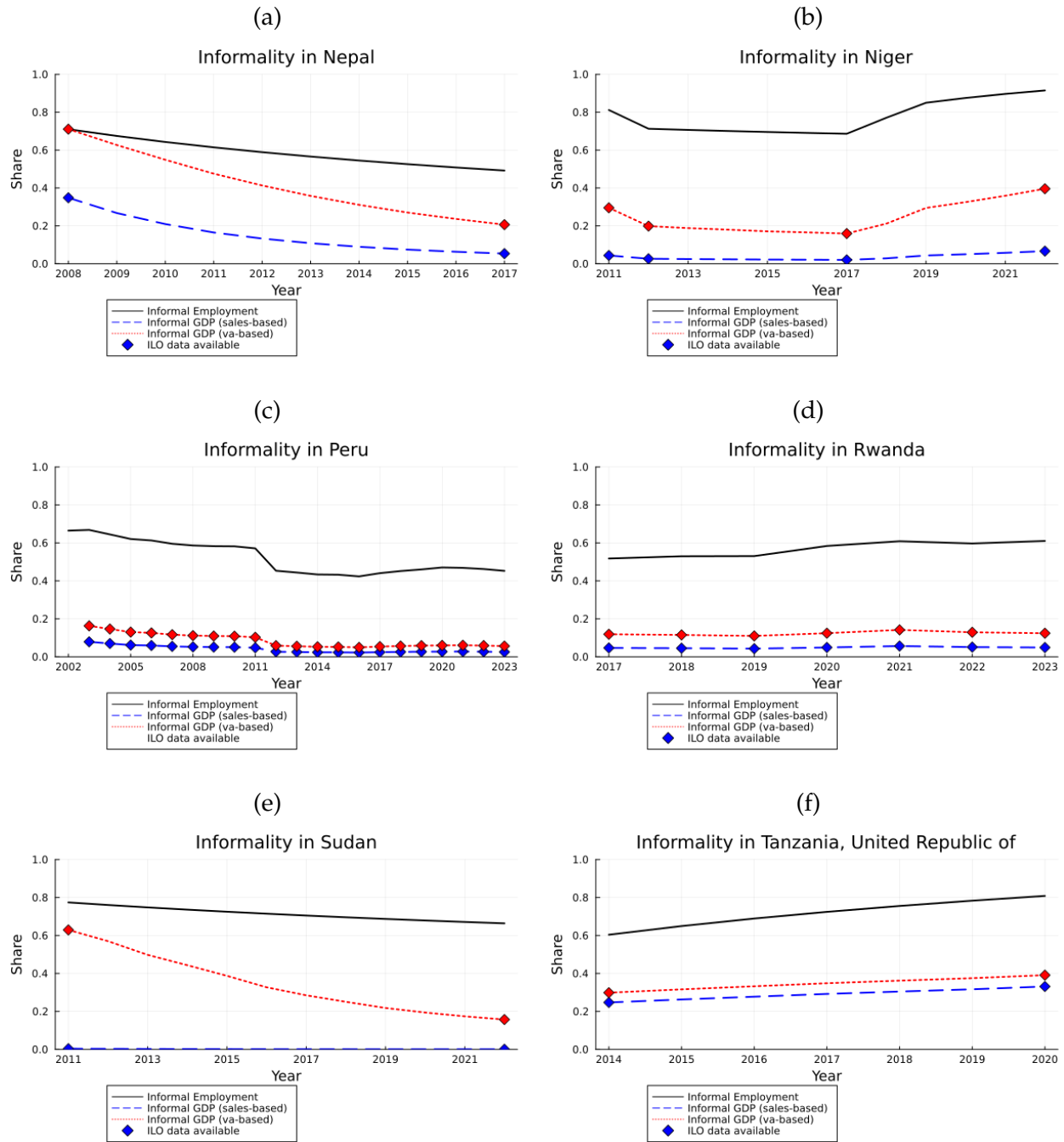
The solid black line represents the share of employment operating outside the formal sector. The dashed blue line is the share of fiscal informality based on sales per worker calculations, while the dotted red line is based on value added per worker calculations. Diamonds indicate years in which informal employment data are available from ILO. When this information is not available, interpolated data are used.

Figure 3: Measures of Informality for Selected Countries



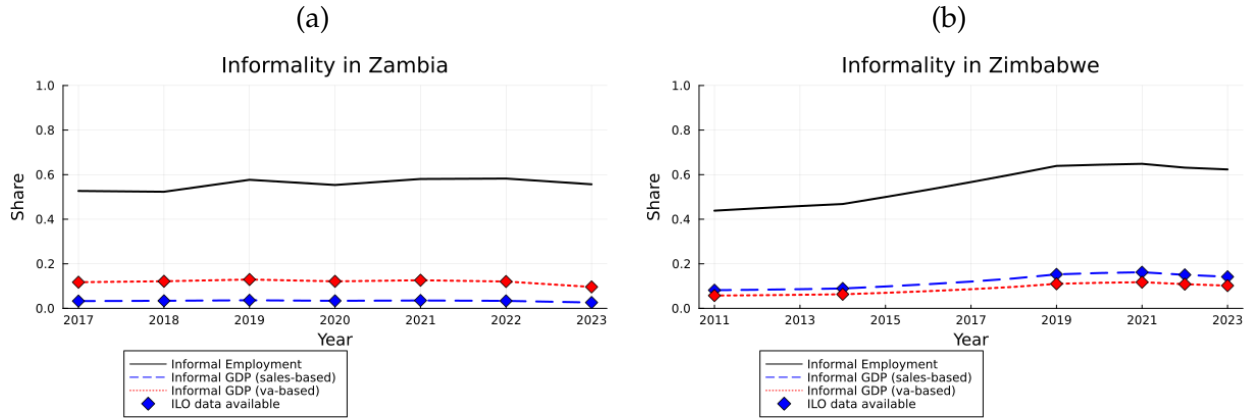
The solid black line represents the share of employment operating outside the formal sector. The dashed blue line is the share of fiscal informality based on sales per worker calculations, while the dotted red line is based on value added per worker calculations. Diamonds indicate years in which informal employment data are available from ILO. When this information is not available, interpolated data are used.

Figure 4: Measures of Informality for Selected Countries



The solid black line represents the share of employment operating outside the formal sector. The dashed blue line is the share of fiscal informality based on sales per worker calculations, while the dotted red line is based on value added per worker calculations. Diamonds indicate years in which informal employment data are available from ILO. When this information is not available, interpolated data are used.

Figure 5: Measures of Informality for Selected Countries



The solid black line represents the share of employment operating outside the formal sector. The dashed blue line is the share of fiscal informality based on sales per worker calculations, while the dotted red line is based on value added per worker calculations. Diamonds indicate years in which informal employment data are available from ILO. When this information is not available, interpolated data are used.

5.1 Relationship Between Informal GDP and GDP per Capita

Figure 6 displays the cross-country correlation between GDP per capita and sales-based (left panel) and VA-based (right panel) informal GDP. Each point represents a country-level average over the years for which the informality series is constructed, excluding data from 2020 on. Both panels show a negative correlation, consistent with previous studies documenting the prevalence of informality in low- and middle-income countries. However, The charts also reveal a group of countries, mainly in Africa, with relatively low informal GDP despite very low income per capita. Three considerations should be taken into account. First, this pattern has been previously documented and is not in conflict with previous studies on informality measures. Various factors beyond income per capita, such as institutional differences, can affect the size of the informal sector. Second, the measure of informality used in this paper, legal informality, does not capture all dimensions of informality, such as tax evasion or informal employment within formal enterprises. Third, for countries surveyed before 2016, firm-level data may not be representative of the business population, which could lead to flawed informal GDP measures.

(a)
(b)



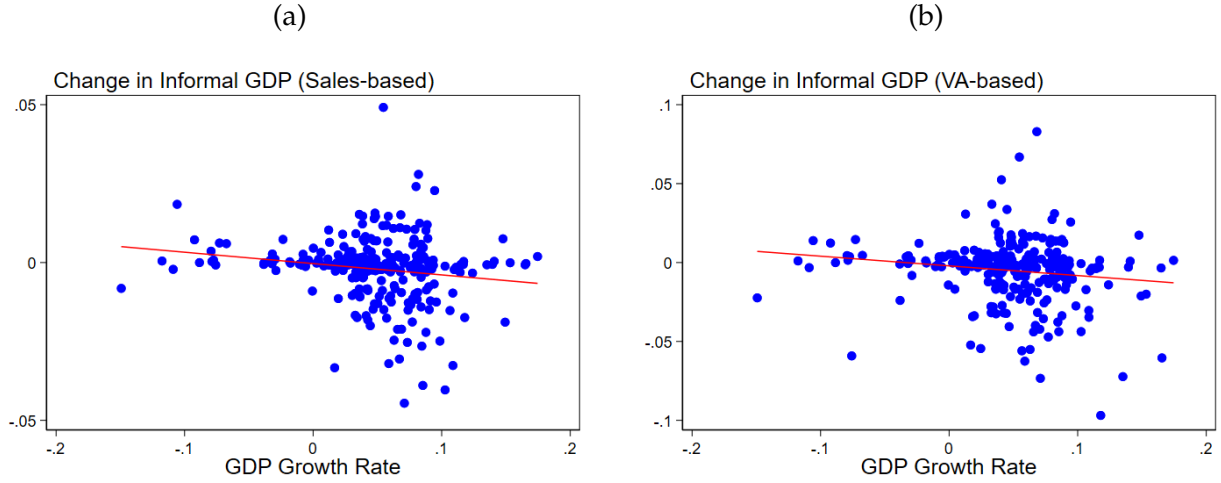
Several studies have documented the countercyclical nature of informality, Particularly its tendency to increase during recessions.⁷ These studies generally employ detailed country-specific information collected at the household or firm level, or both. Ideally, time series of informality should preserve this feature. Figure 7 illustrates the correlation between GDP growth and changes in sales-based (left panel) and VA-based (right panel) informal GDP. Each point represents a combination of GDP growth and change in informal GDP for a given country-year. Both panels display a negative correlation, suggesting that the measures constructed capture the countercyclical behavior of informality.

$$\Delta S_{O,t} = \alpha + \beta \Delta GDP_t + \varepsilon_t$$

Table 3 displays the estimates of the coefficient β , the sensitivity of informal GDP changes

18

Figure 7: Relationship Between Informal GDP Change and GDP Growth Rates



Source: Own calculation for informal GDP, PWT and World Bank for GDP growth. For visualization purposes, I exclude from the graph a couple of outliers (they are not excluded from the empirical analysis).

to GDP growth. The coefficients are negative, as predicted, and statistically significant at the 10% level.

Table 3: Relationship Between Informal GDP Change and GDP Growth Rates

	Informal GDP Change (sales-based)	Informal GDP Change (VA-based)
GDP Growth	-0.032* (0.019)	-0.057* (0.031)
Observations	144	144

Standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

5.3 Comparison With Existing Measures of Informal GDP

Table 4 compares country averages across different estimates of informal GDP. Columns (1) and (2) report the sales-based and VA-based informal GDP derived in this paper, while columns (3) and (4) show the DGE and MIMIC estimates from [Elgin et al. \(2021\)](#). For each country, averages are computed only for the years in which all four series are available.

The measures estimated in this paper tend to be lower than those reported previously. As shown in the bottom row of table 4, the cross-country averages for sales-based and

VA-based informal GDP are about 13% and 23%, compared to 34% and 39% for the DGE and MIMIC estimates. Several factors may explain this discrepancy. First, this paper focuses on a narrower dimension of informality, legal informality, which captures only the production of unregistered businesses, excluding unreported production by otherwise formal enterprises. Second, the analysis is restricted to the non-agricultural sector, which typically exhibits lower levels of informality than the agricultural sector. Third, the exclusion of formal firms with less than five employees in the WBES may lead to an overestimation of the formal sector average productivity, thereby underestimating the share of informal GDP. Conversely, the WBES may also fail to detect very large and highly productive formal firms. As a result, it is not ex-ante clear which effect prevails.

Table 4: Country Averages by Measures of Informal GDP

Country	Years	(1) Sales-Based	(2) VA-Based	(3) DGE	(4) MIMIC
Angola	2004-2018	1.1%	5.6%	40.8%	44.4%
Argentina	2004-2018	2.4%	3.7%	22.3%	24.4%
Bangladesh	2010-2018	22.6%	46.6%	30.2%	35.1%
Botswana	2006-2018	1.3%	6.0%	29.7%	30.9%
Burkina Faso	2018-2018	30.6%	41.8%	31.6%	38.7%
Cambodia	2012-2018	28.7%	34.3%	40.0%	46.0%
Cameroon	2007-2014	7.5%	6.8%	30.9%	32.2%
Cote d'Ivoire	2012-2018	12.4%	14.3%	44.9%	41.9%
Egypt	2008-2018	8.1%	13.2%	31.2%	34.2%
Ghana	2013-2015	13.1%	21.6%	37.9%	39.2%
Guatemala	2002-2018	10.1%	17.5%	47.6%	51.9%
India	2005-2018	19.7%	25.4%	19.3%	21.1%
Indonesia	2016-2018	14.1%	22.0%	15.7%	18.4%
Lao, People's Democratic Republic	2017-2018	14.6%	45.4%	21.8%	28.2%
Madagascar	2012-2018	16.7%	49.2%	36.9%	43.4%
Mauritius	2012-2018	5.4%	16.5%	20.5%	21.6%
Mozambique	2015-2015	6.2%	9.6%	28.2%	39.0%
Myanmar	2015-2018	54.6%	70.8%	26.6%	49.1%
Nepal	2008-2017	15.1%	41.6%	34.6%	36.4%
Niger	2011-2018	2.6%	19.6%	37.0%	40.2%
Peru	2002-2018	4.4%	9.3%	50.9%	57.2%
Rwanda	2017-2018	4.6%	11.7%	29.2%	35.4%
Sudan	2011-2018	0.2%	42.4%	20.8%	-
Tanzania, United Republic of	2014-2018	27.7%	33.1%	45.9%	54.2%
Zambia	2017-2018	3.3%	12.0%	40.6%	47.2%
Zimbabwe	2011-2018	10.0%	7.1%	64.0%	60.7%
Cross-Country Average		12.6%	23.3%	34.2%	39.1%

Source: Own calculation for columns (1) and (2), [Elgin et al. \(2021\)](#) for columns (3) and (4).

6 Conclusion

This paper provides a method to estimate the share of informal GDP using a simple framework and data from various surveys and sources. The estimated informality series exhibit three key features: (i) a negative correlation with GDP per capita, (ii) mild anti-cyclical behavior, (iii) lower levels than those reported in previous cross-country studies. While the current approach allows estimation for only a subset of countries and years, it offers a foundation for future research. Subsequent work could expand the time span, the number of countries, or both. Although the number of countries is constrained by the availability of informal WBES, the time dimension could be extended by incorporating country-specific labor force surveys in place of ILO data.

References

- Aga, Gemechu, David Francis, Filip Jolevski, Jorge Rodriguez Meza, and Joshua Seth Wimpey. 2022. "Surveying Informal Businesses, Methodology and Applications." *World Bank Policy Research Working Paper*.
- Di Nola, Alessandro, Georgi Kocharkov, Almuth Scholl, and Anna-Mariia Tkhir. 2021. "The Aggregate Consequences of Tax Evasion." *Review of Economic Dynamics* 40:198–227.
- Elgin, Ceyhun, M. Ayhan Kose, Franziska Ohnsorge, and Shu Yu. 2021. "Understanding Informality." *Working paper*.
- Elgin, Ceyhun, and Oguz Öztunali. 2012. "Shadow economies around the world: Model based estimates." *Working Papers 2012/05. Bogazici University*.
- Feenstra, Robert C., Robert Inklaar, and Marcel P. Timmer. 2015. "The Next Generation of the Penn World Table." *American Economic Review* 105 (10): 3150–3182. www.ggdc.net/pwt.
- ILO. n.d., https://ilostat.ilo.org/methods/concepts-and-definitions/description-labour-force-statistics/#elementor-toc_heading-anchor-28.
- La Porta, Rafael, and Andrei Shleifer. 2014. "Informality and development." *Journal of Economic Perspectives* 28 (3): 109–26.
- Leyva, Gustavo, and Carlos Urrutia. 2020. "Informality, labor regulation and the business cycle." *Journal of International Economics* 126.
- Orsi, Renzo, Davide Raggi, and Francesco Turino. 2014. "Size, Trend, and Policy Implications of the Underground Economy." *Review of Economic Dynamics* 17:417–436.
- Pappadà, Francesco, and Kenneth Rogoff. 2025. "Rethinking the Informal Economy and the Hugo Effect." *Journal of the European Economic Association*, <https://doi.org/10.1093/jeea/jvaf006>.

Schneider, Friedrich, Andreas Buehn, and Claudio Montenegro. 2010. "New Estimates for the Shadow Economies all over the World." *International Economic Journal* 24 (4): 443–461.

Ulyssea, Gabriel. 2018. "Firms, Informality, and Development: Theory and Evidence from Brazil." *American Economic Review* 108 (8): 2015–2047.

World Bank. 2020. "Re-thinking the Approach to Informal Businesses." *Washington, D.C.*,

Appendix: WBES Data

Description of data. Firm-level data from the WBES are available at <https://www.enterprisesurveys.org/en/data>. The data portal provides a comprehensive database covering formal business with at least five employees across countries and survey years. For most observations, the following key variables are available: annual sales, annual cost of intermediate purchases, value of capital, number of employees, industry classification (manufacturing or services), and country-specific sample weights. I exclude from the sample any observations with missing values for these variables. These data allow me to compute the (weighted) average sales and value added per worker—used on the right hand side of equation (2)—and estimate the production factor shares of equation (5).

The data portal also includes a combined database of the most recent informal surveys, which contain the same variables as the standard WBES, except for the value of capital and expenditures on intermediate purchases. To expand country coverage, I harmonize and merge data from informal surveys conducted prior to 2016. Unfortunately, sample weights are not available for these earlier surveys. In such cases, I assign equal weights to all observations within a country. Given the absence of firm-level data on intermediate purchases in the informal sector, I use annual electricity expenditures as a proxy and subtract them from annual sales to compute value added. In Section 5, I discuss the implications of this choice.

Data cleaning. Monetary values for sales and capital are expressed in local currencies. To ensure consistency across countries and with other macroeconomic variables from the PWT, I convert all monetary variables to constant 2017 USD. This conversion involves two steps: first, I use average annual exchange rates from the World Bank to convert current local currency values into current USD; second, I apply deflators from the Federal Reserve Economic Data (FRED) to express these values in constant 2017 USD.

Additionally, I exclude a small number of observations where a single firm's value added exceeds the sum of the value added of all other firms in the same country. While these large firms may exist, their inclusion would distort the estimation of average sales and

value added per worker. Of the 26 countries considered, only seven observations are excluded, three from the formal sector and four from the informal sector.

Finally, I replace estimated production factor shares $\hat{\alpha}$, $\hat{\gamma}_F$, and $\hat{\gamma}_O$ with cross-country averages whenever their estimated values are negative. This adjustment is applied in three cases: twice for formal sector shares (Angola and Lao PDR), and once for informal sector labor share (Cambodia).

Representativeness of the data. How do informal sector surveys ensure representativeness of the population of informal businesses? Typical informal businesses do not appear in official registrar records or tax rolls. To address this, [Aga et al. \(2022\)](#) propose a methodology based on Adaptive Cluster Sampling (ACS), which generates a representative sample of informal businesses using a geographic sampling approach. This method consists of the following two stages: in the first stage, ACS provides a probabilistic sample of surveyed units which are fully enumerated to list all informal businesses. In the second stage, businesses are randomly selected from the list for an interview.