Can innovation reduce the size of the informal economy? Evidence from panel data

KOUAKOU, Dorgyles C.M. and YEO, Kolotioloma I.H.

ENSTA Paris, The World Bank, CAPEC, CIRES, Université Alassane Ouattara

28 October 2023

Can innovation reduce the size of the informal economy? Evidence from panel data*

Dorgyles C.M. Kouakou† Kolotioloma I.H. Yéo‡

January 17, 2024

Abstract

A substantial body of literature has explored the determinants of the informal economy. However, this literature has predominantly focused on proximate causes such as taxation and unemployment, largely overlooking the role of innovation. This paper aims to fill this gap by examining the effect of innovation production on the size of the informal economy, utilizing a sample of 138 countries spanning the period from 2007 to 2018. We employ a two-step Generalized Method of Moments approach for a dynamic panel data model, addressing both the phenomenon of hysteresis in the development of informality and the endogeneity of innovation, along with several control variables. Estimations reveal that innovation reduces the size of the informal economy, emphasizing the significance of public innovation policies in addressing informality, with expected benefits in terms of tax revenue mobilization. This result remains robust across various controls, alternative estimation techniques, restricted samples, and different measures of both the informal economy and innovation. The study identifies economic development, domestic credit mobilization, and e-government as channels through which innovation influences the informal economy. We conclude by exploring possible public policies.

Keywords: Informal economy; Tax revenue mobilization; Innovation; Economic development; Domestic credit mobilization; E-government

JEL Codes: H25; H26; O17; O31; O38

---

*The authors contributed equally to this work. A previous iteration of this paper circulated under the title "Can innovation reduce the size of the informal economy? Econometric evidence from 138 countries." We appreciate the valuable comments and suggestions from Loesse J. Esso, Sophie Larribeau, Pierre Mohnen, Kouamé R. Moussa, and Hermann Ndoya. Any remaining errors or omissions are our responsibility.

†Unité d'Economie Appliquée, ENSTA Paris, Institut Polytechnique de Paris, and The World Bank. E-mail: dorgyles.kouakou@ensta-paris.fr. (Corresponding author)

‡CAPEC, CIRES, and Université Alassane Ouattara. E-mail: innocent.yeok@uao.edu.ci.

1 Introduction

The proliferation of informal economic activities is a significant contemporary development issue. These activities encompass all legal economic endeavors that would have contributed to the Gross Domestic Product (GDP) if they were officially recorded.[1] The extent to which these activities pose problems depends on the level of development in countries. Indeed, informality, like unemployment, is prevalent in all countries worldwide, irrespective of their level of development, but it is more pronounced in developing countries compared to developed economies. Estimations by Elgin et al. (2022) show that, in 2018, the informal production represented around 40% of GDP in Sub-Saharan Africa and Latin America and the Caribbean, and approximately 25% of GDP in East Asia and the Pacific, Europe and Central Asia, and Middle East and North Africa. It represented around 31% of GDP in South Asia and roughly 12% of GDP in North America. In terms of workforce, the informal economy provides jobs for around 86% of the active population in Africa, roughly 68% in Asia and Pacific, about 40% in the Americas, and approximately 25% in Europe and Central Asia (International Labour Organization 2018). As highlighted by Ulyssea (2018), the prominent role of the informal economy carries significant negative economic consequences for countries. Indeed, informality can lead to wage inequality, the loss of tax revenue, reduced productivity, and diminished economic growth, among other adverse effects.

Considering the adverse effects of informality, a substantial body of literature has emerged with the objective of uncovering the determinants of the size of the informal economy. A comprehensive review of this literature underscores that the potential role of innovation production as a determinant of the informal economy has been largely overlooked, despite its implications for economic, financial, and social development.[2] In fact, innovation can reasonably be expected to have a significant and negative impact on the size of the informal economy for at least three reasons.

First, innovation can reduce the size of the informal economy by promoting economic development. Indeed, innovation is recognized for driving economic development (Cantner et al. 2019, Schumpeter 1912), particularly through enhanced productivity (Amable et al. 2016). Moreover, innovation contributes to economic growth (Aghion & Howitt 1996, Akcigit & Kerr 2018), which, when sustained over time, can foster economic development.

[1] The term for these informal activities is the "informal economy," and the entities operating within the informal economy are referred to as "informal firms."

[2] Innovation is defined as the implementation of a new or significantly improved product (good or service), process, marketing method, or organizational method in business practices, workplace organization, or external relations (OECD & Eurostat 2005). This definition aligns with the Schumpeterian conception of innovation (see Schumpeter 1912). For a detailed understanding of innovation typology, readers may refer to OECD & Eurostat (2005).

Then, as a country becomes more developed, the size of its informal economy tends to decrease. This can be attributed to rising operational costs associated with informal activities, greater demand for modern manufactured products (typically found in the formal sector), the transition from informal (less educated) entrepreneurs to educated entrepreneurs with superior managerial skills who tend to operate formally, and increased economic opportunities (Berdiev & Saunoris 2016, Elbahnasawy 2021, Elbahnasawy et al. 2016, Goel & Nelson 2016, La Porta & Shleifer 2014), among other factors.

Second, innovation can also reduce the size of the informal economy by enhancing domestic credit mobilization. This is significant because financing constraints often hinder the formal registration of businesses (La Porta & Shleifer 2014). Access to credit is contingent, at least in part, on the creditworthiness of the borrower. Innovation can facilitate firms' access to credit by improving their productivity (Amable et al. 2016) and financial performance (Dong et al. 2020, Lu & Chesbrough 2022). Moreover, recent literature emphasizes that the adoption of financial innovation by banks enhances risk management, lowers the cost of capital, and results in increased credit availability and improved financing conditions for borrowers (Brewer III et al. 2000, Hirtle 2009, Nadauld & Weisbach 2012). Additionally, patents have been noted in the literature as instruments that enhance firms' access to external financing because they signal a firm's technological competencies (Hottenrott et al. 2016), can be used as collateral to secure funds (Mann 2018), and correlate with a firm's credit rating (Frey et al. 2019). Other studies by Bellucci et al. (2014), Chava et al. (2017), Freel (2007), and Jacolin et al. (2021) also shed light on the role of innovation in facilitating domestic credit mobilization. Domestic credit mobilization, and financial development in general, have been shown in the literature to significantly reduce the size of the informal economy (Berdiev & Saunoris 2016, Capasso & Jappelli 2013, Elbahnasawy et al. 2016). This is primarily due to the increased opportunities to fund the growth of formal businesses, among other factors.

Third, innovation can reduce the size of the informal economy through e-government. E-government involves the use of information and communication technologies (ICTs) by public authorities to enhance the delivery of public services (Elbahnasawy 2021). Technological innovation, in fact, promotes the development of e-government by facilitating the creation of advanced digital solutions, including online platforms, mobile applications, artificial intelligence, and blockchain technologies. Implementing these solutions in the delivery of public services leads to the modernization of public administration and a substantial enhancement in the efficiency and accessibility of these services (Yang & Rho 2007). E-government, in turn, plays a significant role in reducing the size of the informal economy by eliminating certain barriers to the formalization of informal enterprises (Elbahnasawy

2021, Williams 2023). Indeed, the modernization of public services through e-government helps reduce bureaucratic complexity, a major driver of informal activities, as emphasized by Djankov et al. (2002) and Goel & Nelson (2016). Additionally, e-government, by minimizing human interaction, contributes to the reduction of corruption (Elbahnasawy 2014), which encourages economic activities to shift toward the formal sector (Choi & Thum 2005, Dreher & Schneider 2010, Schneider 2010).

In summary, innovation has the potential to enhance economic development, domestic credit mobilization, and e-government, and these factors, in turn, should reduce the size of the informal economy. Given these considerations, it is crucial to conduct empirical analyses to formally explore the relationship between innovation and the size of the informal economy within a country. A negative and significant impact of innovation would underscore the importance of implementing public policies to enhance a country's innovation capabilities. The specific goal would be to limit the extent of the informal economy, consequently promoting greater economic progress through improved tax revenue mobilization, among other factors.

This paper addresses the gap in the literature by examining the effect of innovation production on the size of the informal economy, utilizing a comprehensive panel dataset covering 138 countries observed between 2007 and 2018. Econometric estimations reveal a consistently negative and significant effect of innovation on the size of the informal economy. This finding remains robust after conducting a wide range of sensitivity tests. The results also emphasize that economic development, domestic credit mobilization, and e-government are channels through which innovation influences the size of the informal economy. These findings underscore the potential of innovation policies to play a pivotal role in reducing informality on a global scale, with expected significant ramifications for economic performance and social development.

Ultimately, this paper contributes to the literature on two distinct levels. *First*, to the best of our knowledge, this study is the first to demonstrate that innovation production diminishes the size of the informal economy. This extends the existing literature on informal economy determinants by highlighting a factor that is not a proximate cause of informality. It also suggests that an approach centered on technology and creativity for reducing informality could be pertinent in confining the informal economy to a minimal scope. *Second*, to the best of our knowledge, this is also the first study to identify economic development, domestic credit mobilization, and e-government as transmission channels for the influence of innovation on the size of the informal economy. This contributes to understanding the mechanisms that underlie the macroeconomic-level relationship between innovation and informality on a global scale.

The remainder of the paper is organized as follows. Section 2 deals with the public policy and development issues relating to the reduction of informality. Section 3 reviews the recent literature on the determinants of the size of the informal economy. Section 4 presents the methodology. Section 5 describes the data and variables used to implement this methodology, and presents some descriptive evidence. Section 6 presents and discusses the estimation results. Section 7 presents a wide array of robustness checks. Section 8 investigates the transmission channels, and section 9 concludes. The appendices contain some materials discussed throughout the paper.

2 Reducing the size of the informal economy: What are the public policy and development issues?

The reduction of informality has gathered significant interest from both policymakers and academia given the substantial size of the informal economy in many regions around the world and its implications for public policy and development. Indeed, although the informal economy can act as a safety net (Loayza & Rigolini 2011), it should be noted that it creates a number of challenges that affect the productive system, public finance and monetary policy, as well as the social fabric.

At the level of the productive system, the informal economy is a source of low productivity as informal firms are usually less productive than their formal counterparts (La Porta & Shleifer 2014). As stressed by La Porta & Shleifer (2008), low productivity hampers the growth of informal firms. It also gives rise to the "working poor" phenomenon as highlighted by the International Labour Organization (2019).[3] Moreover, the informal sector produces goods that are similar to those of the formal economy but of lower quality (Banerji & Jain 2007). It has also been shown to limit economic growth (Loayza 2016).

At the level of public finance and monetary policy, note that the expansion of the informal economy limits tax revenue mobilization (Besley & Persson 2014). Indeed, in countries with a substantial informal sector, tax authorities, ceteris paribus, collect less tax revenue. For instance, in 2020, tax revenues represented on average 16% of GDP in Africa and 20% of GDP in Latin America and the Caribbean, compared with 34% of GDP in the Organisation for Economic Co-operation and Development (OECD) area.[4] The low level of tax revenue mobilization reduces the government's ability to finance its spending and promotes the growth of public debt (Cooray et al. 2017). In addition to restricting public spending, it has been shown in the literature that the informal economy makes monetary policy less

[3]The "working poor" phenomenon is when individuals work long hours but cannot provide proper sustenance for their families.

[4]See the Global Revenue Statistics Database.

effective, although it contributes to mitigating inflation volatility for most types of macroeconomic shocks (see Alberola & Urrutia 2020). This has significant importance from a Central Bank perspective.

On the social front, the informal economy exacerbates the vulnerability of part of the population it employs. In fact, informal workers have no social protection. According to the International Labour Organization (2017), more than half of the world's population (55%) had no social coverage. In regions where informal employment is widespread, the statistics are even more alarming. For instance, in Africa, at least 4 out of 5 people have no social protection (International Labour Organization 2017). The informal economy also accentuates poverty and inequality (Ohnsorge et al. 2022). The lack of social protection combined with low wages makes informal workers particularly vulnerable. Note that the recent COVID-19 pandemic has somehow also exacerbated the fragility of informal employment. Indeed, lockdown measures affected around 75% of informal workers worldwide (International Labour Organization 2020).

In summary, the high prevalence of the informal economy in many regions worldwide represents a significant public policy and development challenge. It is associated with several critical issues, including exacerbating economic and social vulnerabilities, increasing poverty and inequality, and reducing the effectiveness of economic policies. Additionally, it hampers economic performance, limits governments' capacity for effective engagement in self-financed development processes due to low tax revenue mobilization, and contributes to higher levels of public debt. As a result, the expansion of the informal economy may hinder progress toward achieving the Sustainable Development Goals (Ohnsorge et al. 2022). Therefore, it is imperative for governments and development agencies to prioritize addressing the informal economy in their development strategies.

Given the discussion above, it is essential to explore the policies that can effectively reduce informality on a global scale. This paper modestly contributes to this inquiry by conducting an econometric analysis of the role played by innovation.

3 What do we know about the determinants of the size of the informal economy?

The extant literature explains the size of the informal economy by focusing on a number of economic, institutional, political, and social factors. In this section, we give an overview of this literature by focusing on the recent studies.

At the economic level, there is evidence that when a country exhibits significant economic growth or development, the size of its informal economy tends to decrease (Berdiev & Saunoris 2016, Elbahnasawy 2021, Elbahnasawy et al. 2016, Goel & Nelson 2016, La Porta & Shleifer 2014). Inflation has been found to enlarge the size of the informal economy by increasing the demand for informal sector goods (Alm & Embaye 2013, Goel & Nelson 2016). Similarly, it has been argued that increases in the level of unemployment play as an incentive to work in the informal sector (Buehn & Schneider 2012, Dell'Anno & Solomon 2008), at least as a result of lack of opportunity, which increases the scope of the informal economy. Greater economic openness has been highlighted as a factor that reduces significantly the size of the informal economy (Blanton et al. 2018). In the same vein, Berdiev & Saunoris (2018) point out that economic globalization, a concept that is broader than economic openness, decreases the scope of the informal economy. In contrast to Berdiev & Saunoris (2018) and Blanton et al. (2018), Pham (2017) managed to demonstrate that trade and financial openness increase the size of the informal economy, as measured by informal employment. Trade restrictions also appeared to have a role in explaining the scope of the informal economy (Elbahnasawy et al. 2016). Chatterjee & Turnovsky (2018) finds that larger remittances are associated with a larger size of the informal economy.

Financial development has been highlighted in the literature as significantly and negatively impacting informality. It has been argued that financial development encourages firms to operate formally as external financing becomes available at a lower cost (Berdiev & Saunoris 2016, Capasso & Jappelli 2013). Taxation also plays a role in explaining the proliferation of informal activities. A high tax burden may stimulate informal activities by increasing production costs (Dabla-Norris et al. 2008, Djankov et al. 2010, Friedman et al. 2000, La Porta & Shleifer 2008, Schneider 2010), suggesting a positive effect on the size of the informal economy. However, the effect of taxation might also be negative. It has been argued that high taxation can reduce the informal economy when law enforcement institutions are strong and the credit market is developed (Mitra 2017). In addition to monetary costs such as taxes, Djankov et al. (2002) and Goel & Nelson (2016) have found that non-monetary costs, particularly lengthy procedures for starting a business and paying taxes, reduce incentives to operate in the formal sector. Excessive regulation, especially in the labor market, has also been found to stimulate the growth of informal activities (Schneider 2010, Schneider et al. 2010).

At the institutional level, the literature highlights the importance of good quality institutions in significantly limiting the proliferation of informal activities. For instance, Dabla-Norris et al. (2008) and Dreher et al. (2009) show that government efficiency helps reduce the size of the informal economy. Greater control of corruption has been found to significantly reduce informality (Choi & Thum 2005, Dreher & Schneider 2010, Schneider 2010). The enforcement of laws has also emerged in recent literature as an important determinant of the informal economy (Elbahnasawy 2021, Liu-Evans & Mitra 2019). The idea is that because informal activities are illegal, the more rule of law is respected and enforced in a country, the smaller the size of the informal economy.

At the political and social levels, the literature identifies several factors that play a crucial role in explaining the expansion of informal activities. Political stability is a key factor in this regard, as an unstable political environment limits the government's ability to effectively detect informal production (Elbahnasawy et al. 2016). Additionally, the type of political regime influences individuals' decisions to operate in the formal or informal sector. For example, in a democratic regime, tax policies often align more closely with citizens' preferences, reducing the likelihood of operating in the informal sector (Teobaldelli & Schneider 2013). Elbahnasawy (2021) demonstrates that internal conflict can increase the size of the informal economy by reducing tax compliance. Population size and urbanization are other social factors that have been found in the literature to contribute to limiting the proliferation of informal activities (Elbahnasawy 2021, Elgin & Oyvat 2013, Ndoya & Djeufack 2021).

Beyond these purely economic, institutional, political, and social factors, there have been very few recent studies that examine the role of ICTs (mobile phones and the internet), e-government, and financial mobile services in explaining the evolution of the informal economy. Indeed, research has shown that ICTs can significantly impact the size of the informal economy by improving human capital, financial development, and control of corruption (Ndoya et al. 2023). Financial mobile services can help reduce informal activities by facilitating access to financial services (Jacolin et al. 2021). Elbahnasawy (2021) shows that e-government may reduce informality by increasing efficiency in tax collection, among other factors.

From the foregoing, it is striking that the role of the production of innovation in explaining the size of the informal economy has been overlooked despite its implications for economic, financial, and social development. Additionally, the effect of innovation on the size of the informal economy is worth investigating as the reduction of informality is associated with a number of topical public policy and development issues as outlined previously. The next section presents the methodology we have adopted in this paper to make this investigation.

4 Methodology

This paper aims to analyze the effect of innovation on the size of the informal economy. To identify this effect, we must address two main econometric issues. First, as emphasized by Eilat & Zinnes (2002), the informal or shadow economy is established through a phenomenon of hysteresis, which makes it difficult to eliminate once it takes root. To be precise, the current level of the informal economy is, at least partly, a result of the previous level of informality. The informal economy exhibits a kind of persistence over time, as highlighted in previous empirical studies.[5] Second, innovation is endogenous. This endogeneity may arise from reverse causality, omitted variable bias, or measurement errors. Moreover, several control variables, including taxation, economic growth, political stability, and unemployment, may also exhibit endogeneity. Failing to correct for endogeneity would cast doubt on the reliability of the estimates.

In line with Elbahnasawy (2021), Elbahnasawy et al. (2016), Ndoya & Djeufack (2021), and Ndoya et al. (2023), we address these two econometric issues by estimating a dynamic panel data model using the system Generalized Method of Moment (GMM) estimator. The system GMM estimator is an improved version of the difference GMM estimator proposed by Arellano & Bond (1991). The methodology was first outlined by Arellano & Bover (1995) and then fully developed by Blundell & Bond (1998). The model to be estimated is specified as follows:

$$Inf_{i,t} = \alpha + \delta Inf_{i,t-1} + \lambda Innov_{i,t} + X'_{i,t}\beta + \mu_i + \upsilon_{i,t} \tag{1}$$

where $Inf_{i,t}$ denotes the output of the informal economy as a percentage of the official GDP for country i in year t, with $i = 1, ...,$N and $t = 1, ...,$T; N and T are the total numbers of countries and years, respectively. δ is a parameter to estimate, and $Inf_{i,t-1}$ is the lagged informal economy of country i. $Innov_{i,t}$ represents the level of innovation of country i in year t, and $X_{i,t}$ is a vector of control variables for country i in year t. α is a constant term, and λ is the main parameter of interest. It allows us to capture the effect of innovation on the size of the informal economy. β is a vector of parameters to estimate. μ_i denotes the unobserved time-invariant country-specific effects, and υ_{it} represents the usual error term which varies across countries and years. It has the usual properties, namely, a mean of 0, constant variance, and for all countries and years, the υ_{it} are independent of the regressors and μ_i, and uncorrelated with themselves.

As emphasized by Elbahnasawy (2021), the dynamic nature of the model allows us to account for the persistence of the informal economy (phenomenon of hysteresis) while the Instrumental Variables (IVs) help address the suspected endogeneity of all the regressors, not limited to innovation. This methodology enables the capture of country heterogeneity by including country effects. Country heterogeneity can stem from cross-country variations

[5]See Elbahnasawy (2021), Elbahnasawy et al. (2016), Ndoya & Djeufack (2021), and Ndoya et al. (2023), among others.

in economic, political, institutional, and social environments. The methodology also enables the handling of unobserved country-specific factors that could lead to omitted variable bias and efficient management of measurement errors (Baltagi et al. 2009).

We use the two-step variant of the system GMM because it is more asymptotically efficient than the one-step variant. However, it is worth noting that the two-step standard errors often exhibit significant downward bias. To address this bias, we employ the variance correction method proposed by Windmeijer (2005), which yields bias-corrected robust standard errors.

In the implementation of the system GMM methodology, we begin by first-differencing equation (1). This allows us to deal with fixed effects. The transformed model is described in equation (2) as follows:

$$\begin{aligned} Inf_{i,t} - Inf_{i,t-1} &= \delta(Inf_{i,t-1} - Inf_{i,t-2}) + \lambda(Innov_{i,t} - Innov_{i,t-1}) \\ &+ (X'_{i,t} - X'_{i,t-1})\beta + (\upsilon_{i,t} - \upsilon_{i,t-1}) \end{aligned} \tag{2}$$

Then, in a second step, the equations in levels and differences are simultaneously estimated. This is done using different sets of IVs. The IVs are lagged versions of the right-hand-side variables. To avoid the issue of IV proliferation, which occurs when there are too many IVs, the IV matrix is collapsed (Roodman 2009*a*,0).

As previously explained, the system GMM methodology allows us to deal efficiently with suspected endogeneity of all the regressors. This concerns all three theoretical sources of endogeneity: simultaneity bias/reverse causality (where the relationship between variables is bidirectional and simultaneous), omitted variables, and measurement errors, and is operationalized using the IV technique.[6] The system GMM methodology also helps avoid dynamic panel data bias (Nickell 1981). Note that in the estimations, we include time-specific dummy variables. This allows us to account for possible time-specific factors, such as the global financial crisis, that might have affected both the informal economy and innovation. By doing so, we eliminate any possible omitted variable bias that may be due to time-specific variables.

5 Data and preliminary analysis

In this section, we start by introducing the variables used in this study and the data sources. Following that, we present and discuss the key descriptive evidence.

[6]For a comprehensive discussion on the use of IVs to address endogeneity issues stemming from simultaneity bias (reverse causality), omitted variables bias, or measurement errors, refer to Angrist & Krueger (2001), Baltagi et al. (2009), and Becker (2016).

5.1 Data and variables

To assess the effect of innovation on the size of the informal economy, we utilize an unbalanced large panel dataset comprising 138 countries observed from 2007 to 2018.[7] The selection of countries in the dataset and the chosen time span are based on data availability for both the size of the informal economy and innovation. Notably, the unavailability of data on the composite and comprehensive measure of innovation production at the global level prior to 2007 dictated the time frame.

We utilize Elgin et al.'s (2022) estimates of the size of the informal economy, which reflect all legal economic activities that would have contributed to the GDP if they were recorded. This excludes all economic activities that are classified as criminal by public authorities. The estimates of the size of the informal economy are obtained using the MIMIC method. MIMIC-based estimates are widely used in the literature (see Dell'Anno (2016), Dreher & Schneider (2010), Elbahnasawy (2021), Elbahnasawy et al. (2016), Goel & Nelson (2016), Pham (2017), and Schneider (2010), among others). MIMIC is a type of structural equations model that combines multiple causes and outcome indicators of informal activities to estimate their relative size. It is based on the statistical theory of unobserved variables that draws upon the multiple causes and indicators of a given phenomenon to measure it. Combining the causes and outcome indicators of the informal economy allows us to better capture informal activities and, therefore, to have more reliable estimations of their scope.

An important feature of the MIMIC method is that it can readily be used to obtain the size of the informal economy of a large set of countries worldwide over time (more than 150 countries worldwide). Very importantly, this method takes into account both the levels of employment and productivity of the informal sector when measuring the size of the informal economy. This allows us to have a more comprehensive measure of the informal economy compared to estimations that focus on informal employment and which merely reflect the level of employment in the informal sector. Note that we use Elgin et al.'s (2022) MIMIC-based estimates of the size of the informal economy because they are more recent and/or available over a longer time span compared to other existing estimations (see Medina & Schneider (2018) and Medina & Schneider (2019), among others). They actually cover 160 economies observed over the period 1993-2018.

Two alternative measures of the informal economy are considered for robustness checks. The first one is Medina & Schneider's (2019) MIMIC-based estimates that are available for 157 countries from 1991 to 2017. This measure is used mainly because it is obtained using a set of variables that is not identical to the one used by Elgin et al. (2022). This allows us

[7] Table A1 in Appendix A lists these countries.

to check the robustness of our results to the choice of the variables included in the MIMIC. In the same line as Elgin & Oyvat (2013) and Ndoya et al. (2023), the second alternative measure is non-agricultural informal employment. Non-agricultural informal employment is considered, among others, because in many regions worldwide, informal employment is usually more prevalent outside of the agricultural sector. See, for instance, Figure 1 in Ndoya et al. (2023) for an illustration of this fact in 2018. The estimates of the size of the informal economy are obtained using survey data, that is, through a direct approach, which is not the case for the MIMIC-based measures.

To measure innovation, in line with Kouakou (2022), we use the innovation output index. This index is extracted from the report titled "Global Innovation Index," published yearly by the World Intellectual Property Organization (WIPO), Cornell University, and INSEAD, since 2007. It is computed based on two pillars, namely "knowledge and technology outputs" and "creative outputs." Knowledge and technology outputs encompass knowledge creation (e.g., patents), knowledge impact (e.g., high- and medium-high-tech manufacturing), and knowledge diffusion (e.g., high-tech net exports). Creative outputs refer to intangible assets (e.g., industrial designs), creative goods and services (e.g., creative goods exports), and online creativity (e.g., mobile app creation). The innovation output index is a composite index, providing a comprehensive measure of the production of innovation, as highlighted by Kouakou (2022). It measures a country's level of innovation production on a scale from 0 to 100, where a higher index indicates a higher level of innovation production.

Two alternative measures of innovation are used for robustness checks. The first measure is patent applications by residents, which is a traditional indicator of a country's innovation level. A higher number of patents indicates a higher level of innovation production in a country. The second alternative measure of innovation is Research and Development (R&D) intensity, expressed as R&D expenditure as a percentage of GDP. R&D plays a significant role in driving innovation, making it a suitable proxy. In the empirical literature on the determinants and effects of innovation (see, for instance, Gong & Hanley (2021)), R&D intensity is commonly used as an alternative measure of innovation. A higher R&D intensity in a country is associated with a greater likelihood of producing innovations.

Regarding the control variables, note that it is essential to include variables that exhibit correlations with both the size of the informal economy and innovation. This helps mitigate potential omitted variable bias and enhances estimation efficiency. To select these variables, we draw from the literature on the determinants of innovation and the size of the informal economy. The chosen variables include political stability, GDP growth, urban population, unemployment, trade openness, financial openness, and taxation. These variables constitute the set of baseline control variables.

Political stability plays a crucial role in our analysis as it allows us to consider the political environment's impact on both innovation and the extent of the informal economy. Political instability can hinder a government's ability to effectively detect informal production, thereby fostering the proliferation of informal activities (Elbahnasawy et al. 2016). Moreover, it has been noted in the literature that political instability tends to reduce the introduction of product innovation and weaken the national system of innovation (Allard et al. 2012, Krammer & Kafouros 2022). This is often attributed to declines in long-term investments, such as R&D, and frequent instability in the regulatory framework. It is worth noting that the informal sector often serves as a safety net (Loayza & Rigolini 2011). Informal employment tends to expand as the unemployment rate rises. To be precise, the informal sector absorbs part of the workforce unable to secure employment in the formal sector (Dell'Anno & Solomon 2008, Fields 1975). Furthermore, informal activities typically flourish in unfavorable economic contexts and contract as the economic situation improves. Therefore, an increase in economic growth is expected to reduce the size of the informal economy, while an increase in the unemployment rate is likely to lead to an expansion of the informal economy.

While innovation drives economic growth, economic growth can also fuel further innovation. This occurs through economic revitalization, increased business opportunities, and improved economic conditions, often resulting from expanded infrastructure development. Countries with high unemployment rates often experience increased international emigration (White & Buehler 2018). This encourages the "brain drain" phenomenon, where skilled workers emigrate in pursuit of better job opportunities and an improved quality of life abroad. However, the emigration of skilled workers may reduce innovation by depleting the country's human capital. Skilled employees play a vital role in innovation as they enhance firms' absorptive capacity (Leiponen 2005). This emigration can potentially hinder a country's innovation potential.

Taxation is another crucial economic factor considered in this research. Formal firms face various taxes that can be burdensome. High taxes can incentivize participation in the informal economy to reduce this burden, suggesting a positive correlation between taxation and the size of the informal economy (Dabla-Norris et al. 2008, La Porta & Shleifer 2008, Schipper 2020, Ulyssea 2018). However, the literature also notes that higher tax rates may have a negative impact on informality when institutions are of better quality and the credit market is developed (Goel & Nelson 2016, Mitra 2017). Therefore, the effect of taxation on the size of the informal economy is not straightforward. Taxation may also reduce the quantity of innovation produced in a country (Akcigit et al. 2022). High taxation increases production costs, which can impede innovation by reducing R&D investment. Trade openness is another economic factor relevant to understanding the scope of the informal economy (Elbahnasawy 2021, Elgin & Oyvat 2013). Indeed, trade openness can create economic opportunities, generate jobs, and promote economic progress, conditions that deter the proliferation of informal activities. It can also stimulate R&D investment (Teteryatnikova 2018), potentially increasing a country's likelihood to produce innovations, possibly due to greater exposure to international market competition.

Financial openness, by expanding financing opportunities through increased cross-border financial transactions, can potentially reduce the size of the informal economy. It may also stimulate innovation by providing greater support for innovative activities. Urbanization is another factor considered in this research to account for the social environment of countries. As cities modernize, they tend to attract rural workers seeking better-paying jobs (Harris & Todaro 1970, Todaro 1969), which are often found in the formal economy. Urbanization can thus play a significant role in reducing the informal economy's size (Elgin & Oyvat 2013, Ndoya & Djeufack 2021). Additionally, urbanization can have a notable impact on innovation, as shown by Chen et al. (2020). It has the potential to increase innovation at the country level by enhancing regional innovation capabilities.

For robustness checks, we consider a set of additional control variables, following the literature on the determinants of innovation and the size of the informal economy. These variables are government effectiveness, control of corruption, rule of law, inflation, trade restrictions, financial development, political system, and regime durability.

Government effectiveness and control of corruption control for the economic and institutional dimensions of governance in explaining informality, as outlined by Elbahnasawy (2021). These variables also contribute to long-term macroeconomic efficiency in the production of innovation (Kouakou 2022). The rule of law further addresses the institutional environment. Weaker law enforcement is detrimental to innovation, as it may fail to guarantee the property rights of innovators. Strong institutional quality makes it easier to detect informal economic activity, reinforces people's trust in institutions, and increases their likelihood to engage in the formal economy (Elbahnasawy 2021). Since informal activities are illegal (as they are not registered), stricter enforcement of the rule of law should reduce their prevalence in a country (Kouakou 2023). Political system and regime durability allow us to consider additional aspects of the political environment in countries, beyond what is covered by the political (in)stability variable. They are expected to have a negative impact on both innovation and the informal economy, as a deteriorating political environment is unfavorable for innovation and the development of the formal sector, as explained previously.[8]

[8] Political system and regime durability have recently been found by Elbahnasawy et al. (2016) to reduce the size of the informal economy.

Inflation is expected to increase the size of the informal economy. In a high inflation scenario, there is an increased demand for informal goods because they are relatively cheaper than those produced in the formal sector (Buehn & Schneider 2012, Goel & Nelson 2016). This stimulates the proliferation of informal activities. Inflation may also reduce innovation by weakening firms' ability to invest in R&D activities (Chu et al. 2019, Costamagna 2015). Financial development is expected to reduce the scope of the informal economy because a developed financial system provides greater access to low-cost financing, which encourages firms to operate in the formal sector (Berdiev & Saunoris 2016, Capasso & Jappelli 2013).[9] Financial development may also foster innovation (Hsu et al. 2014) due to increased financing for innovation activities. Restrictions on international trade may increase the size of the informal economy. One rationale is that these restrictions are correlated with rural areas and shift demand toward domestic output, making it more difficult to detect informal activities and stimulating their proliferation (Elbahnasawy et al. 2016). Restrictions on international trade may also reduce innovation, primarily by limiting exposure to competition on international markets.

Table A2 in Appendix A gives information on how each variable is measured. The data sources are also indicated. As explained in the Introduction section, we test three transmission channels in this study, that is, economic development, domestic credit mobilization, and e-government. The measures and data sources of these channels are also presented in Table A2.

5.2 Descriptive evidence

Descriptive statistics are presented in Table A3 in Appendix A. It emerges that over the study period, on average, the informal economy represents approximately 32% of GDP. It ranges from around 8% to 69% of GDP. Non-agricultural informal employment represents between approximately 1% and 96% of total employment, with an average level of around 33%. The average level of innovation is around 32 on a scale of 0 to 100. This means that, on average, countries have had a poor level of innovation over the study period. Indeed, 50 can be seen as an intermediate value that separates poor performers (countries that have an innovation index less than 50) from good performers (those countries having an innovation index higher than 50). The average level of R&D intensity is about 1% of GDP, with a minimum level of 0.01%. This is relatively low and corroborates the statistics on the innovation production index.

Furthermore, it is worth noting that, on average, approximately 60% of the population

[9] Financial development is a concept that is different and broader than financial openness. These two concepts should not be considered equivalent.

resides in urban areas, indicating a significant level of urbanization. The average unemployment rate is below 8% of the total labor force, but in several countries, it exceeds 25%, which is quite high. As seen in Table A3, the average total tax and contribution rate amounts to around 42% of profits, reflecting a high tax burden. Importantly, the average levels of government effectiveness, control of corruption, and the rule of law are positive. This indicates that, on average, the countries in the sample have demonstrated strong performances in terms of government effectiveness, control of corruption, and the enforcement of laws during the study period. However, it is worth noting that the variable measuring political stability exhibits a negative mean, suggesting a significant level of political instability throughout the study period.

Table A3 also reveals that the countries in the sample are, on the whole, quite open to international markets. The average trade openness rate stands at approximately 88% of GDP. In contrast, the average inflation rate hovers around 6%, indicating that some countries may need to make additional efforts to target inflation levels around the typical 2% threshold often sought in monetary policy. Moreover, the sample countries, on average, experience relatively low restrictions on international trade. The mean of the variable measuring these restrictions exceeds 50. The average level of domestic credit to the private sector is roughly 59% of GDP, which is significant. Nonetheless, the relatively high standard deviation of approximately 46% provides valuable context for understanding the substantial average domestic credit to the private sector. Taking a broader perspective on financial development, the average level appears to be low, with the mean of the financial development index being less than 0.5. Additionally, the e-government index, with an average of 0.53, suggests a moderate level of e-government adoption in the sample during the study period.

Correlation coefficients are reported in Table A4 in Appendix A. The correlation between the MIMIC-based estimates is equal to 97%, which is very high. This is an interesting result because the sets of variables used by Elgin et al. (2022) and Medina & Schneider (2019) are not identical, showing that Elgin et al.'s (2022) estimates of the size of the informal economy are not biased by the choice of the variables included in the MIMIC. The correlation between the Elgin et al. (2022) MIMIC-based estimates and non-agricultural employment is equal to around 60%, which is high. As expected, the innovation production index is highly correlated and negatively associated with all three measures of the size of the informal economy. It is also highly correlated and positively associated with R&D intensity and patent applications. These alternative measures of innovation are negatively correlated with all three measures of the size of the informal economy, with the correlation being more significant for R&D intensity. The transmission channels also exhibit high negative correlations with the measures of informality, but high positive correlations with innovation, which is expected. It also emerges many of the control variables are highly correlated with both innovation and the informal economy. This shows that the choice of the control variables is appropriate overall. Very importantly, this suggests that many of the control variables will effectively help to mitigate a possible omitted variable bias, thereby helping to better identify the effect of innovation on the size of the informal economy.

The variance inflation factors (VIFs) are reported in Table A5 in Appendix A. Regardless of the specification, the VIFs are all less than 5. This means that the econometric estimations do not suffer from collinearity issues.

Figure 1 presents two scatterplots illustrating the relationship between innovation and the size of the informal economy worldwide. The scatterplot displayed in the upper part of this figure is standard, while the one at the bottom considers the average levels of innovation and the informal economy for each country over the study period. Countries are represented by their standard country codes.

[Insert Figure 1 here]

The downward trend in the scatterplots indicates a negative correlation between innovation and the size of the informal economy. The higher the level of innovation, the smaller the size of the informal economy. Indeed, we can see that the most innovative countries (e.g. Switzerland (CHE), Sweden (SWE), Netherlands (NLD), UK (GBR), and Luxembourg (LUX)) have small informal economies. Conversely, in countries with a low level of innovation (e.g. Togo (TGO), Niger (NER), Zimbabwe (ZWE), Bolivia (BOL), and Georgia (GEO)), the size of the informal economy is large. While purely descriptive, this result suggests that fostering innovation may help public authorities curb the expansion of the informal economy.

Figures B1, B2, and B3 in Appendix B present world maps depicting the evolution of the size of the informal economy and innovation worldwide over the past years. These maps highlight regional patterns in the evolution of informality and innovation. Overall, regions with high levels of innovation appear to record low sizes of the informal economy. This reaffirms the potential negative association between these two variables, as illustrated in Figure 1.

6 Results

Table 1 presents the two-step system GMM estimates. The robust standard errors, as suggested by Windmeijer (2005), are reported in parentheses.

[Insert Table 1 here]

We observe from Table 1 that all diagnostics meet the required standards. The Arellano-Bond tests indicate rejection of the null hypothesis of no first-order autocorrelation but do not reject the null hypothesis of no second-order autocorrelation. The Hansen test for overidentifying restrictions does not reject the null hypothesis that all instruments are valid. The number of instruments is significantly fewer than the number of countries.

The two-step system GMM estimates reveal that innovation has a negative and significant effect on the size of the informal economy. In other words, a higher level of innovation in a country is associated with a lower size of its informal economy. As discussed in the Introduction section, a plausible explanation for this result can be derived from the economic development, domestic credit mobilization, and e-government channels. We formally test these three transmission channels in Section 8.

As for the controls, Table 1 indicates that economic growth plays a significant role in reducing the size of the informal economy. Specifically, a higher economic growth rate in a country is associated with a smaller informal economy. Indeed, the informal economy tends to diminish in countries experiencing substantial economic progress (Chong & Gradstein 2007, La Porta & Shleifer 2014). A key rationale behind this observation is that economic progress often brings about greater opportunities, and individuals with more opportunities are less inclined to operate outside the purview of public authorities. Similarly, financial openness is found to be associated with a reduction in informality. The expansion of financing opportunities resulting from increased cross-border financial transactions may indeed support the development of formal businesses.

Unemployment is found to have a positive and significant effect on the size of the informal economy. This implies that the higher a country's unemployment level, the larger its informal economy. This result aligns with the perspective that informality is, to some extent, a consequence of poor economic system management, leading to a scarcity of economic and employment opportunities. In countries with a deficit of such opportunities, individuals are more inclined to engage in informal activities to meet their needs. Nowadays, as previously emphasized by La Porta & Shleifer (2014), the informal economy supports the livelihoods of billions of people worldwide.

We see from Table 1 that an increased level of urbanization correlates with a reduced size of the informal economy. Consequently, significant urbanization of countries plays a role in constraining informality (Elgin & Oyvat 2013, Ndoya & Djeufack 2021). As noted earlier, this can be attributed, at least in part, to the fact that as cities undergo modernization, they attract rural workers in pursuit of more lucrative employment opportunities (Harris & Todaro 1970, Todaro 1969), typically available within the formal economy. This phenomenon hinders the growth of informal activities.

7 Robustness checks

Our estimations indicate that innovation significantly diminishes the size of the informal economy. In this section, we aim to assess the robustness of this main result through six different analyses. *First*, we assess the sensitivity of our primary finding to potential omitted variables by incorporating eight additional control variables. *Second*, we estimate the effect of innovation using the entropy balancing method, a novel impact evaluation technique, as an alternative approach to addressing the endogeneity of innovation. *Third*, we examine restricted samples. *Fourth*, we explore alternative measures of the size of the informal economy. *Fifth*, we also consider alternative measures of innovation. *Sixth*, we estimate the effect of innovation using pooled cross-section, fixed-effects, and random-effects regressions. As demonstrated by all these analyses, innovation consistently and significantly reduces the size of the informal economy. These results strongly support the robustness of our main finding.

7.1 Additional controls

We extend our baseline specification to a set of eight additional control variables. These variables are government effectiveness, control of corruption, rule of law, inflation, trade restrictions, financial development, political system, and regime durability.[10] These variables capture institutional, political, and economic aspects of the determinants of the size of the informal economy and innovation. As mentioned earlier, this analysis enables us to assess the robustness of our main finding against potential omitted variable bias.

To mitigate potential multicollinearity issues, the additional control variables are entered one at a time in the regression. The results are presented in Table 2.

[Insert Table 2 here]

All diagnostics are satisfactory, irrespective of the regression. We see from Table 2 that the effect of innovation remains strong and negative even with the inclusion of additional controls. This suggests that our finding is robust to a potential omitted variable bias. The observed stability of the coefficient of innovation across all eight specifications further supports the robustness of our results.

[10] For a discussion on the choice of these variables, refer to subsection 5.1. Their definitions are summarized in Table A2 in Appendix A.

7.2 Entropy balancing

Another approach to addressing the endogeneity of innovation involves employing the entropy balancing method. The methodology and advantages of this technique are detailed in Appendix C. This subsection begins by analyzing the performance of the entropy balancing method. To do so, we present, in Table 3, the summary statistics on balancing quality obtained from a (weighted) regression of the treatment variable ("Innovation 1") on the covariates (Tübbicke 2022). The balancing statistics include the R-squared and F-statistic from this regression (Tübbicke 2022), with the associated p-value for the F-statistic also reported. Table 3 presents the statistics both before and after applying the entropy balancing weighting used to estimate the treatment effect of innovation, allowing us to assess the performance of the entropy balancing method.

[Insert Table 3 here]

Prior to weighting, the R-squared stands at approximately 0.49, suggesting that the covariates account for roughly 49% of the variance in the treatment variable. Regarding the F-test, the p-value is 0.000, suggesting rejection of the null hypothesis that the covariates do not significantly influence the treatment variable overall.

After applying the entropy balancing weighting, the R-squared becomes 0, indicating that the covariates no longer induce differences in the treatment variable, as expected. Concerning the F-test, the F-statistic is 0, and the p-value is 1, indicating a failure to reject the null hypothesis that the covariates do not have a significant effect on the treatment variable overall. This aligns with the results from the R-squared. The balancing property is satisfied.

We can now proceed to interpret the treatment effect obtained through entropy balancing. The weights acquired in the first step of entropy balancing are then used in a second step to estimate the effect of innovation on the size of the informal economy, employing the weighted least squares method. The results are presented in Table 4.

[Insert Table 4 here]

Columns (1) to (4) present the effects without the matching covariates used in the first step of entropy balancing. In line with Balima et al. (2021), we include year and regional fixed effects in the second step of the entropy balancing methodology. Columns (5) to (8) repeat the exercise by incorporating the matching covariates used in the first step of entropy balancing. These covariates are lagged by one year to mitigate potential issues of reverse causality.

Year fixed effects control for macroeconomic shocks or time-related factors impacting the informal economy and innovation, while regional fixed effects allow us to account for region-specific factors affecting the informal economy, as highlighted by Elbahnasawy (2021), and innovation, as illustrated in Figure B2.[11] Incorporating regional fixed effects in a robustness check is important, as they provide a broader perspective on control variables. Furthermore, the inclusion of matching covariates in the second step of the entropy balancing methodology, akin to integrating control variables in a randomized experiment, enhances estimation efficiency.

As depicted in Table 4, innovation consistently and significantly diminishes the size of the informal economy, reaffirming previous results. However, it is noteworthy that the coefficients of innovation appear notably larger than the two-step system GMM estimate. This disparity may be attributed, at least in part, to the entropy balancing method, which does not account for the phenomenon of hysteresis in the establishment of the informal economy. Consequently, this method might have somewhat overestimated the effect of innovation on the informal economy. The difference in the magnitude of the effect of innovation may also be attributed to the inclusion of regional fixed effects in the second step of the entropy balancing methodology, consistent with the approach of Balima et al. (2021).

Figure 2 presents the Dose-Response Function (DRF) along with the corresponding 95% confidence interval. Standard errors are obtained using the bootstrap method (see Efron & Tibshirani 1986, MacKinnon 2006). The effectiveness of the bootstrap method in the context of entropy balancing for continuous treatments has been demonstrated by Vegetabile et al. (2021). This method was subsequently employed by Tübbicke (2022) to derive standard errors when estimating the DRF. In our analysis, we implement 500 bootstrap replications.

[Insert Figure 2 here]

The DRF illustrates how the size of the informal economy responds to various intensities (levels) of innovation. In Figure 2, a noticeable decrease in the size of the informal economy is evident as treatment intensity increases. In simpler terms, the higher the level of innovation, the smaller the size of the informal economy. This finding strongly reaffirms previous results regarding the negative effect of innovation on the scope of the informal economy.

[11] Consistent with Elbahnasawy (2021), we consider the regions identified by The World Bank, including Latin America and the Caribbean (LAC), North America (NA), South Asia (SA), Sub-Saharan Africa (SSA), Middle East and North Africa (MENA), East Asia and Pacific (EAP), and Europe and Central Asia (ECA).

7.3 Restricted samples

In this subsection, we narrow down the full sample either by focusing on a specific group of countries or by excluding particular regions or sets of countries. We then examine whether our main result remains consistent. This approach allows us to assess the robustness of our main finding with respect to restricted samples. Specifically, we consider three cases.

First, we focus on the sub-sample of developing countries as informality is an issue that is more prevalent in developing countries. Second, we remove countries from the G7. Indeed, G7 contains seven of the most advanced countries in the world. They exhibit significant levels of innovation and are among countries where informally is the less prevalent in the world. So, by removing G7 countries, we somehow try to focus on countries for which informally is more or less an important issue, while going beyond developing countries. Third, we remove from the sample countries from Sub-Saharan Africa and Latin America and the Caribbean. In fact, these regions have the highest levels of informality in the world. They also exhibit low levels of innovation. Hence, the negative effect of innovation on the size of the informal economy may have been influenced to some extent by the presence of countries from these regions in the sample. Thus, it is worth investigating whether our main result holds or not when countries from Sub-Saharan Africa and Latin America and the Caribbean are removed from the sample.

The two-step system GMM estimates are presented in Table 5. In each case, all the diagnostics are satisfactory.

[Insert Table 5 here]

In all cases, the effect of innovation is consistently negative and significant, indicating that innovation reduces the size of the informal economy. This reaffirms our previous findings.

7.4 Alternative measures of the size of the informal economy

In this subsection, we assess the robustness of our finding by examining alternative measures of the informal economy. We consider two specific measures. The first one is the MIMIC-based estimates of the informal economy, as proposed by Medina & Schneider (2019). The variables used in the MIMIC differ from those considered by Elgin et al. (2022). Therefore, utilizing Medina & Schneider's (2019) estimates enables us to evaluate the robustness of our main result concerning the choice of variables in the MIMIC when gauging the size of the informal economy. The second alternative measure is non-agricultural informal employment, aligning with the approach of Elgin & Oyvat (2013) and Ndoya et al. (2023). Unlike the MIMIC-based estimates, this measure is derived from survey data, representing a direct measurement approach. Importantly, exploring alternative measures of the informal economy allows us to test the resilience of our results to potential measurement errors in estimating the size of the informal economy.

We refer to the Medina & Schneider (2019) MIMIC-based estimates of the size of the informal economy as "Informal economy 2" and the non-agricultural informal employment measure as "Informal economy 3."[12] The two-step system GMM estimates obtained using these two variables as outcome variables are presented in Table 6. All diagnostic tests meet the required standards.

[Insert Table 6 here]

We see from Table 6 that the effect of innovation is consistently negative and significant for both alternative measures of the informal economy. This demonstrates the robustness of our main result, indicating that it is not likely influenced by potential errors in measuring the size of the informal economy.

7.5 Alternative measures of innovation

We further assess the robustness of our results by examining alternative measures of innovation. Specifically, we consider two alternative measures: patent applications by residents, denoted as "Innovation 2," and R&D intensity, denoted as "Innovation 3." For more comprehensive definitions and data sources of these variables, refer to Table A2 in Appendix A. Patent applications provide protection for inventions, encompassing products or processes that introduce new methods or novel technical solutions to problems. They also serve as records for inventions and the innovations associated with them. Therefore, the number of patents serves as a valuable indicator for measuring a country's level of production of innovation. R&D intensity serves as a reliable proxy for a country's level of innovation. Generally, the most innovative countries in the world are those with the highest levels of R&D intensity.

Table 7 displays the results obtained when using "Innovation 2" and "Innovation 3" as alternative measures of innovation. For each regression, the diagnostics are adequate. We observe that, irrespective of the specification, innovation consistently shows a negative and significant effect on the informal economy, underscoring the robustness of our findings against potential errors in measuring innovation.

[Insert Table 7 here]

[12] For detailed definitions and data sources of these variables, refer to Table A2 in Appendix A.

7.6 Additional robustness

In our ultimate robustness check, we estimate pooled cross-sectional, fixed-effects, and random-effects models. The estimates are available in Table A6 in Appendix A for reference. These findings further reaffirm our earlier results, emphasizing that innovation consistently diminishes the size of the informal economy.

8 Channels

Our results indicate that innovation reduces the size of the informal economy. The objective of this section is to elucidate the underlying mechanisms behind this finding, as introduced in the Introduction section. We investigate three potential mechanisms: economic development, domestic credit mobilization, and e-government. Our analysis encompasses two approaches. *First*, we conduct a descriptive analysis that involves examining the co-evolution of innovation and each potential channel, as well as the relationship between each potential channel and the informal economy. Scatterplots are employed for this analysis. *Second*, we formally test these channels using the entropy balancing method. Specifically, we use the entropy balancing method to assess both the impact of innovation on potential channels and the impact of each channel on the informal economy. This method enables us to identify causal effects while effectively addressing the endogeneity of innovation and the associated transmission channels.

Figure 3 presents scatterplots for the descriptive analysis, sequentially addressing economic development (part (a)), domestic credit mobilization (part (b)), and e-government (part (c)) as potential transmission channels.

[Insert Figure 3 here]

In part (a) of Figure 3, we observe that, on average, higher levels of innovation in a country correspond to greater economic development. Simultaneously, increased economic development is associated with a reduction in the size of the informal economy. These findings suggest that economic development serves as a channel through which innovation influences the size of the informal economy. We observe a similar result with domestic credit mobilization. In part (b) of Figure 3, it is evident that, on average, a country with higher levels of innovation also exhibits increased domestic credit mobilization. Concomitantly, higher levels of domestic credit mobilization are associated with a reduction in the size of the informal economy, suggesting that domestic credit mobilization serves as a transmission channel for the effect of innovation on informality. The descriptive analysis also indicates that e-government is another possible transmission channel for the effect of innovation. In part (c) of Figure 3, we observe that, on average, a country with higher levels of innovation also has higher levels of e-government. Simultaneously, increases in e-government levels are linked to a decrease in the size of the informal economy.

Let us now go one step further in the analysis of the channels by testing them formally using the entropy balancing method. The results of the estimations are reported in Table 8, with a sequential focus on economic development (part (I)), domestic credit mobilization (part (II)), and e-government (part (III)). For the effects of innovation on the channels, the summary statistics on balancing quality are those presented previously in Table 3 and are all adequate. As to the effects of the channels on the size of the informal economy, the summary statistics on balancing quality are reported in Tables A7 to A9, which can be found in Appendix A. They all meet the required standards.

[Insert Table 8 here]

In part (I) of Table 8, we observe that innovation has a positive and significant effect on economic development, while economic development significantly reduces the size of the informal economy. Similar results are found for domestic credit mobilization and e-government. Indeed, in part (II) of Table 8, we see that innovation stimulates domestic credit mobilization, and in turn, domestic credit mobilization decreases the informal economy. Regarding e-government, it is significantly and positively affected by innovation and has a negative effect on the size of the informal economy, as illustrated in part (III) of Table 8.

In summary, innovation increases economic development, domestic credit mobilization, and e-government, and these factors, in turn, reduce the size of the informal economy. This demonstrates that they are channels through which innovation impacts the size of the informal economy. The consistency of the results across different types of analysis argues in favor of the robustness and relevance of the highlighted channels.

9 Conclusion

This paper analyzes the effect of innovation production on the size of the informal economy worldwide, utilizing a large panel dataset covering 138 countries observed from 2007 to 2018. Econometric estimations demonstrate that innovation significantly reduces the size of the informal economy, and this result remains robust after conducting a series of checks. Economic development, domestic credit mobilization, and e-government play important roles as channels through which innovation influences the informal economy. These findings support the notion that innovation is a powerful driver of development. Countries that enhance their innovation capacity are more likely to achieve remarkable macroeconomic performance, including the reduction of their informal sectors, which is expected to improve tax revenue mobilization.

Effective public innovation policies are therefore warranted to limit the scope of the informal economy worldwide. Such policies are particularly needed in low-income countries as these countries generally have significantly lower levels of innovation than their developed counterparts due to weaker absorptive capacity and limited financial and knowledge resources, among others. A number of policies could be considered by countries to improve their levels of innovation.[13] Among others, countries could give firms tax incentives for R&D. R&D investment is, indeed, a key driver of innovation. In practice, firms that invest in R&D activities could be offered tax reduction. The extent of the reduction, often referred to as the "R&D tax credit," should be proportionate to the level of investment in R&D. While such a policy is implemented in developed countries like France (the famous "Crédit Impôt Recherche"), UK, and US, it does not exist in most developing countries. R&D tax credit is a lever that these countries can activate to provide firms with significant incentives to invest in R&D. As to developed countries, intensifying existing tax policies intended to give firms incentives for R&D may help increase the number of firms that invest in R&D, thereby improving their levels of innovation.

Besides, note that it can be difficult for small firms to innovate compared to their larger counterparts due to a weaker financing capacity, among others. Public policy may have a role in supporting these firms to improve their innovativeness. One option could involve providing direct financial support to small firms for their R&D activities, tailored to their financing capacity.[14] As we mentioned previously, lack of absorptive capacity and limited knowledge resources are factors that alter firms' ability to innovate, in particular, in developing economies. Public authorities could help to alleviate these issues by providing technical services and advice to firms that express a need. As Edler & Fagerberg (2017) rightly stressed, such a policy should help improve access to expertise, a milestone for successful innovation activities.

Beyond technical services and advice, however, policies for training are expected to play a key role in improving a country's level of innovation by enhancing both absorptive capacity and knowledge. The policies to be implemented could take the form of financial support for firms that make substantial investments in workers' training related to the technological knowledge relevant to their sector.

[13] For a taxonomy of innovation policy instruments, see Edler & Fagerberg (2017).

[14] For a recent discussion on the theoretical framework of the public financing of innovation, see Mazzucato & Semieniuk (2017).

The present research can be extended in several directions. Among others, future research can investigate the possible existence of a threshold effect in the impact of innovation on the size of the informal economy. The existence of such a threshold would indicate distinct regimes, where the impact of innovation on the informal economy varies from one regime to another. This analysis could be conducted using the panel threshold modeling approach. In this study, we focused on the effect of innovation output, that is, the production of innovation. The analysis could be extended to the innovation inputs, that is, to factors that drive a country's level of innovation (infrastructure, business sophistication, etc.). A comparative analysis of their effects on the informal economy compared to innovation output may lead to relevant conclusions regarding which element, innovation output or innovation inputs, public policies should prioritize to minimize the size of the informal economy. Further studies could also differentiate between female and male employment when measuring the informal economy through informal employment. This may help determine whether there is a gender gap in the impact of innovation on the informal economy.

Another avenue for future research consists in going beyond the simple production of innovation to deal with "innovation efficiency," that is, the efficiency in the production of innovation. Innovation efficiency refers to a country's ability to achieve the maximum possible level of innovation given its endowment in innovation inputs (R&D, infrastructure, business sophistication, etc.).[15] In fact, the more efficient a country is in producing innovations, the higher its level of innovation, which should significantly reduce the size of the informal economy. In the same line as Kouakou (2022), one could distinguish between short-run efficiency and long-run efficiency, and then analyze which between innovation short-run efficiency and innovation long-run efficiency reduces the most the size of the informal economy. Innovation efficiency scores can be obtained by making a stochastic frontier analysis.

Data availability statement

For access to the data, Stata Do-file, and R codes, please do not hesitate to contact the corresponding author.

Declaration of interest

There is no conflict of interest.

[15] While the concept of production efficiency, also known as "technical efficiency," has its foundations in microeconomics, there exists a substantial body of literature dedicated to the analysis of macroeconomic innovation efficiency. For a recent literature review, refer to Kouakou (2022).

References

Aghion, P. & Howitt, P. (1996), 'Research and development in the growth process', *Journal of Economic Growth* **1**(1), 49–73.

Akcigit, U., Grigsby, J., Nicholas, T. & Stantcheva, S. (2022), 'Taxation and innovation in the twentieth century', *The Quarterly Journal of Economics* **137**(1), 329–385.

Akcigit, U. & Kerr, W. (2018), 'Growth through heterogeneous innovations', *Journal of Political Economy* **126**(4), 1374–1443.

Alberola, E. & Urrutia, C. (2020), 'Does informality facilitate inflation stability?', *Journal of Development Economics* **146**, 102505.

Allard, G., Martinez, C. A. & Williams, C. (2012), 'Political instability, pro-business market reforms and their impacts on national systems of innovation', *Research Policy* **41**(3), 638–651.

Alm, J. & Embaye, A. (2013), 'Using dynamic panel methods to estimate shadow economies around the world, 1984–2006', *Public Finance Review* **41**(5), 510–543.

Amable, B., Ledezma, I. & Robin, S. (2016), 'Product market regulation, innovation, and productivity', *Research Policy* **45**(10), 2087–2104.

Angrist, J. D. & Krueger, A. B. (2001), 'Instrumental variables and the search for identification: From supply and demand to natural experiments', *Journal of Economic Perspectives* **15**(4), 69–85.

Arellano, M. & Bond, S. (1991), 'Some tests of specification for panel data: Monte carlo evidence and an application to employment equations', *The Review of Economic Studies* **58**(2), 277–297.

Arellano, M. & Bover, O. (1995), 'Another look at the instrumental variable estimation of error-components models', *Journal of Econometrics* **68**(1), 29–51.

Balima, H. W., Combes, J.-L. & Minea, A. (2021), 'The "dark side" of credit default swaps initiation: A close look at sovereign debt crises', *Macroeconomic Dynamics* **25**(1), 124–153.

Baltagi, B. H., Demetriades, P. O. & Law, S. H. (2009), 'Financial development and openness: Evidence from panel data', *Journal of Development Economics* **89**(2), 285–296.

Banerji, A. & Jain, S. (2007), 'Quality dualism', *Journal of Development Economics* **84**(1), 234–250.

Becker, S. O. (2016), 'Using instrumental variables to establish causality', IZA World of Labor, 250.

Bellucci, A., Favaretto, I. & Giombini, G. (2014), Does innovation affect credit access? New empirical evidence from Italian small business lending, IAW Discussion Papers 104, Tübingen.

Berdiev, A. N. & Saunoris, J. W. (2016), 'Financial development and the shadow economy: A panel VAR analysis', *Economic Modelling* **57**, 197–207.

Berdiev, A. N. & Saunoris, J. W. (2018), 'Does globalisation affect the shadow economy?', *The World Economy* **41**(1), 222–241.

Besley, T. & Persson, T. (2014), 'Why do developing countries tax so little?', *Journal of Economic Perspectives* **28**(4), 99–120.

Blanton, R. G., Early, B. & Peksen, D. (2018), 'Out of the shadows or into the dark? Economic openness, IMF programs, and the growth of shadow economies', *The Review of International Organizations* **13**, 309–333.

Blundell, R. & Bond, S. (1998), 'Initial conditions and moment restrictions in dynamic panel data models', *Journal of Econometrics* **87**(1), 115–143.

Brewer III, E., Minton, B. A. & Moser, J. T. (2000), 'Interest-rate derivatives and bank lending', *Journal of Banking & Finance* **24**(3), 353–379.

Buehn, A. & Schneider, F. (2012), 'Shadow economies around the world: novel insights, accepted knowledge, and new estimates', *International Tax and Public Finance* **19**, 139–171.

Cantner, U., Dettmann, E., Giebler, A., Guenther, J. & Kristalova, M. (2019), 'The impact of innovation and innovation subsidies on economic development in German regions', *Regional Studies* **53**(9), 1284–1295.

Capasso, S. & Jappelli, T. (2013), 'Financial development and the underground economy', *Journal of Development Economics* **101**, 167–178.

Chatterjee, S. & Turnovsky, S. J. (2018), 'Remittances and the informal economy', *Journal of Development Economics* **133**, 66–83.

Chava, S., Nanda, V. & Xiao, S. C. (2017), 'Lending to innovative firms', *The Review of Corporate Finance Studies* **6**(2), 234–289.

Chen, J., Wang, L. & Li, Y. (2020), 'Natural resources, urbanization and regional innovation capabilities', *Resources Policy* **66**, 101643.

Chinn, M. D. & Ito, H. (2006), 'What matters for financial development? Capital controls, institutions, and interactions', *Journal of Development Economics* **81**(1), 163–192.

Choi, J. P. & Thum, M. (2005), 'Corruption and the shadow economy', *International Economic Review* **46**(3), 817–836.

Chong, A. & Gradstein, M. (2007), 'Inequality and informality', *Journal of Public Economics* **91**(1), 159–179.

Chu, A. C., Cozzi, G., Furukawa, Y. & Liao, C.-H. (2019), 'Inflation and innovation in a schumpeterian economy with north–south technology transfer', *Journal of Money, Credit* *and Banking* **51**(2-3), 683–719.

Cooray, A., Dzhumashev, R. & Schneider, F. (2017), 'How does corruption affect public debt? An empirical analysis', *World Development* **90**, 115–127.

Costamagna, R. (2015), 'Inflation and R&D investment', *Journal of Innovation Economics & Management* **17**(2), 143–163.

Dabla-Norris, E., Gradstein, M. & Inchauste, G. (2008), 'What causes firms to hide output? The determinants of informality', *Journal of Development Economics* **85**(1), 1–27.

Dell'Anno, R. (2016), 'Analyzing the determinants of the shadow economy with a "separate approach". An application of the relationship between inequality and the shadow economy', *World Development* **84**, 342–356.

Dell'Anno, R. & Solomon, O. H. (2008), 'Shadow economy and unemployment rate in USA: is there a structural relationship? An empirical analysis', *Applied Economics* **40**(19), 2537–2555.

Djankov, S., Ganser, T., McLiesh, C., Ramalho, R. & Shleifer, A. (2010), 'The effect of corporate taxes on investment and entrepreneurship', *American Economic Journal: Macroeconomics* **2**(3), 31–64.

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.

Dong, Y., Skowronski, K., Song, S., Venkataraman, S. & Zou, F. (2020), 'Supply base innovation and firm financial performance', *Journal of Operations Management* **66**(7-8), 768–796.

Dreher, A. (2006), 'Does globalization affect growth? Evidence from a new index of globalization', *Applied Economics* **38**(10), 1091–1110.

Dreher, A., Kotsogiannis, C. & McCorriston, S. (2009), 'How do institutions affect corruption and the shadow economy?', *International Tax and Public Finance* **16**, 773–796.

Dreher, A. & Schneider, F. (2010), 'Corruption and the shadow economy: an empirical analysis', *Public Choice* **144**, 215–238.

Edler, J. & Fagerberg, J. (2017), 'Innovation policy: what, why, and how', *Oxford Review of Economic Policy* **33**(1), 2–23.

Efron, B. & Tibshirani, R. (1986), 'Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy', *Statistical Science* **1**(1), 54–75.

Eilat, Y. & Zinnes, C. (2002), 'The shadow economy in transition countries: Friend or foe? A policy perspective', *World Development* **30**(7), 1233–1254.

Elbahnasawy, N. G. (2014), 'E-government, internet adoption, and corruption: An empirical investigation', *World Development* **57**, 114–126.

Elbahnasawy, N. G. (2021), 'Can e-government limit the scope of the informal economy?', *World Development* **139**, 105341.

Elbahnasawy, N. G., Ellis, M. A. & Adom, A. D. (2016), 'Political instability and the informal economy', *World Development* **85**, 31–42.

Elgin, C., Kose, M. A., Ohnsorge, F. & Yu, S. (2022), Understanding the informal economy: Concepts and trends, *in* 'The Long Shadow of Informality: Challenges and Policies', World Bank Publications, chapter 2, pp. 33–92.

Elgin, C. & Oyvat, C. (2013), 'Lurking in the cities: Urbanization and the informal economy', *Structural Change and Economic Dynamics* **27**, 36–47.

Fields, G. S. (1975), 'Rural-urban migration, urban unemployment and underemployment, and job-search activity in LDCs', *Journal of Development Economics* **2**(2), 165–187.

Fong, C., Hazlett, C. & Imai, K. (2018), 'Covariate balancing propensity score for a continuous treatment: Application to the efficacy of political advertisements', *The Annals of Applied Statistics* **12**(1), 156–177.

Freel, M. S. (2007), 'Are small innovators credit rationed?', *Small Business Economics* **28**(1), 23–35.

Frey, C. B., Neuhäusler, P. & Blind, K. (2019), 'Patents and corporate credit risk', *Industrial and Corporate Change* **29**(2), 289–308.

Friedman, E., Johnson, S., Kaufmann, D. & Zoido-Lobaton, P. (2000), 'Dodging the grabbing hand: the determinants of unofficial activity in 69 countries', *Journal of Public Economics* **76**(3), 459–493.

Goel, R. K. & Nelson, M. A. (2016), 'Shining a light on the shadows: Identifying robust determinants of the shadow economy', *Economic Modelling* **58**, 351–364.

Gong, Y. & Hanley, A. (2021), 'Exports and new products in China – a generalised propensity score approach with firm-to-firm spillovers', *The Journal of Development Studies* **57**(12), 2136–2155.

Gygli, S., Haelg, F., Potrafke, N. & Sturm, J.-E. (2019), 'The KOF globalisation index–revisited', *The Review of International Organizations* **14**, 575–575.

Hainmueller, J. (2012), 'Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies', *Political Analysis* **20**(1), 25–46.

Harris, J. R. & Todaro, M. P. (1970), 'Migration, unemployment and development: A two-sector analysis', *The American Economic Review* **60**(1), 126–142.

Hirtle, B. (2009), 'Credit derivatives and bank credit supply', *Journal of Financial Intermediation* **18**(2), 125–150.

Hottenrott, H., Hall, B. H. & Czarnitzki, D. (2016), 'Patents as quality signals? The implications for financing constraints on R&D', *Economics of Innovation and New Technology* **25**(3), 197–217.

Hsu, P.-H., Tian, X. & Xu, Y. (2014), 'Financial development and innovation: Cross-country evidence', *Journal of Financial Economics* **112**(1), 116–135.

International Labour Organization (2017), World social protection report 2017–19: Universal social protection to achieve the sustainable development goals, Technical report, International Labour Organization.

International Labour Organization (2018), Women and men in the informal economy: A statistical picture (Third edition), Technical report, International Labour Organization.

International Labour Organization (2019), Advancing social justice: Shaping the future of work in Africa, Technical report, International Labour Organization.

International Labour Organization (2020), ILO Monitor: COVID-19 and the world of work. Third edition, Technical report, International Labour Organization.

Jacolin, L., Keneck Massil, J. & Noah, A. (2021), 'Informal sector and mobile financial services in emerging and developing countries: Does financial innovation matter?', *The World Economy* **44**(9), 2703–2737.

Kouakou, D. C. M. (2022), 'Separating innovation short-run and long-run technical efficiencies: Evidence from the Economic Community of West African States (ECOWAS)', *European Journal of Comparative Economics* **19**(1), 103–141.

Kouakou, D. C. M. (2023), 'Competing against 'invisibles': the effect of competition from informal firms on formal firms' R&D', *Eurasian Business Review* **13**(1), 87–117.

Krammer, S. M. S. & Kafouros, M. I. (2022), 'Facing the heat: Political instability and firm new product innovation in sub-Saharan Africa', *Journal of Product Innovation Management* **39**(5), 604–642.

Kullback, S. (1959), *Information Theory and Statistics*, Wiley, New York.

La Porta, R. & Shleifer, A. (2008), 'The unofficial economy and economic development', *Brookings Papers on Economic Activity* pp. 275–352.

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

Leiponen, A. (2005), 'Skills and innovation', *International Journal of Industrial Organization* **23**(5), 303–323.

Liu-Evans, G. & Mitra, S. (2019), 'Informality and bank stability', *Economics Letters* **182**, 122–125.

Loayza, N. V. (2016), 'Informality in the process of development and growth', *The World Economy* **39**(12), 1856–1916.

Loayza, N. V. & Rigolini, J. (2011), 'Informal employment: Safety net or growth engine?', *World Development* **39**(9), 1503–1515.

Lu, Q. & Chesbrough, H. (2022), 'Measuring open innovation practices through topic modelling: Revisiting their impact on firm financial performance', *Technovation* **114**, 102434.

MacKinnon, J. G. (2006), 'Bootstrap methods in econometrics', *Economic Record* **82**(s1), S2–S18.

Mann, W. (2018), 'Creditor rights and innovation: Evidence from patent collateral', *Journal of Financial Economics* **130**(1), 25–47.

Mazzucato, M. & Semieniuk, G. (2017), 'Public financing of innovation: new questions', *Oxford Review of Economic Policy* **33**(1), 24–48.

Medina, L. & Schneider, F. (2018), 'Shadow economies around the world: What did we learn over the last 20 years?', Working Paper No. 2018/017, International Monetary Fund, Washington, D.C.

Medina, L. & Schneider, F. (2019), 'Shedding light on the shadow economy: A global database and the interaction with the official one', CESifo Working Paper No. 7981.

Mitra, S. (2017), 'To tax or not to tax? When does it matter for informality?', *Economic Modelling* **64**, 117–127.

Nadauld, T. D. & Weisbach, M. S. (2012), 'Did securitization affect the cost of corporate debt?', *Journal of Financial Economics* **105**(2), 332–352.

Ndoya, H. & Djeufack, A. (2021), 'Urbanization, governance and informal economy: an African tale', *Economics Bulletin* **41**(3), 1525–1540.

Ndoya, H., Okere, D., Laure Belomo, M. & Atangana, M. (2023), 'Does ICTs decrease the spread of informal economy in Africa?', *Telecommunications Policy* **47**(2), 102485.

Nickell, S. (1981), 'Biases in dynamic models with fixed effects', *Econometrica* **49**(6), 1417–1426.

OECD & Eurostat (2005), *Oslo Manual: Guidelines for Collecting and Interpreting Innovation Data*, 3rd edn, OECD Publishing, Paris.

Ohnsorge, F., Okawa, Y. & Yu, S. (2022), Lagging behind: Informality and development, *in* 'The Long Shadow of Informality: Challenges and Policies', World Bank Publications, chapter 4, pp. 123–204.

Pham, T. H. H. (2017), 'Impacts of globalization on the informal sector: Empirical evidence from developing countries', *Economic Modelling* **62**, 207–218.

Robins, J. M., Hernán, M. A. & Brumback, B. (2000), 'Marginal structural models and causal inference in epidemiology.', *Epidemiology* **11**(5), 550–560.

Roodman, D. (2009*a*), 'How to do xtabond2: An introduction to difference and system GMM in Stata', *The Stata Journal* **9**(1), 86–136.

Roodman, D. (2009*b*), 'A note on the theme of too many instruments', *Oxford Bulletin of Economics and Statistics* **71**(1), 135–158.

Schipper, T. C. (2020), 'Informality, innovation, and aggregate productivity growth', *Review of Development Economics* **24**(1), 125–143.

Schneider, F. (2010), 'The influence of public institutions on the shadow economy: An empirical investigation for OECD countries', *Review of Law & Economics* **6**(3), 441–468.

Schneider, F., Buehn, A. & Montenegro, C. E. (2010), 'Shadow economies all over the world: New estimates for 162 countries from 1999 to 2007', Policy Research Working Paper No. WPS 5356, World Bank Group, Washington, D.C.

Schumpeter, J. A. (1912), *The Theory of Economic Development*, Harvard University Press, Cambridge.

Svirydzenka, K. (2016), 'Introducing a new broad-based index of financial development', Working Paper No. 2016/005, International Monetary Fund, Washington, D.C.

Teobaldelli, D. & Schneider, F. (2013), 'The influence of direct democracy on the shadow economy', *Public Choice* **157**, 543–567.

Teteryatnikova, M. (2018), 'R&D in trade networks: The role of asymmetry', *International Journal of Industrial Organization* **61**, 307–350.

Todaro, M. P. (1969), 'A model of labor migration and urban unemployment in less developed countries', *American Economic Review* **59**(1), 138–148.

Tübbicke, S. (2022), 'Entropy balancing for continuous treatments', *Journal of Econometric Methods* **11**(1), 71–89.

Ulyssea, G. (2018), 'Firms, informality, and development: Theory and evidence from Brazil', *American Economic Review* **108**(8), 2015–2047.

Vegetabile, B. G., Griffin, B. A., Coffman, D. L., Cefalu, M., Robbins, M. W. & McCaffrey, D. F. (2021), 'Nonparametric estimation of population average dose-response curves using entropy balancing weights for continuous exposures', *Health Services and Outcomes Research Methodology* **21**(1), 69–110.

White, R. & Buehler, D. (2018), 'A closer look at the determinants of international migration: decomposing cultural distance', *Applied Economics* **50**(33), 3575–3595.

Williams, C. C. (2023), Theorising an e-government approach towards formalising the informal economy, *in* 'Formalization of the Informal Economy: An E-government Approach', Springer, pp. 21–56.

Windmeijer, F. (2005), 'A finite sample correction for the variance of linear efficient two-step GMM estimators', *Journal of Econometrics* **126**(1), 25–51.

Yang, K. & Rho, S.-Y. (2007), ‘E-government for better performance: Promises, realities, and challenges’, *International Journal of Public Administration* **30**(11), 1197–1217.

Yiu, S. & Su, L. (2018), ‘Covariate association eliminating weights: a unified weighting framework for causal effect estimation’, *Biometrika* **105**(3), 709–722.

Zhu, Y., Coffman, D. L. & Ghosh, D. (2015), ‘A boosting algorithm for estimating generalized propensity scores with continuous treatments.’, *Journal of Causal Inference* **3**(1), 25–40.

Table 1. Two-step system GMM estimates

Informal economy 1 (lag)	0.819***
	(0.063)
Innovation 1	-0.049***
	(0.016)
Political stability	0.004
	(0.012)
GDP growth	-0.003***
	(0.001)
Urban population	-0.137**
	(0.058)
Unemployment	0.016*
	(0.009)
Trade openness	0.033
	(0.023)
Financial openness	-0.010*
	(0.006)
Taxation	-0.029
	(0.021)
Constant	1.271***
	(0.413)
Time dummies	Yes
Fisher test (p-value)	0.000
Arellano-Bond test for AR(1) (p-value)	0.000
Arellano-Bond test for AR(2) (p-value)	0.758
Hansen test of overidentifying restrictions (p-value)	0.882
Number of instruments	34
Number of countries	131
Average observations per country	8.76
Observations	1,148

Notes: As expected, the 95% confidence interval for the coefficient of the lagged dependent variable does not contain the value 1. "Informal economy 1" and "Innovation 1" are our main measures of the size of the informal economy and innovation, respectively. See Table A2 for a description of the variables. Windmeijer-corrected robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2. Robustness checks – Additional controls, two-step system GMM estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Informal economy 1 (lag)	0.819***	0.844***	0.827***	0.831***	0.810***	0.822***	0.829***	0.781***
	(0.070)	(0.075)	(0.066)	(0.069)	(0.074)	(0.086)	(0.059)	(0.101)
Innovation 1	-0.048***	-0.046***	-0.049***	-0.044***	-0.049***	-0.046***	-0.048***	-0.048**
	(0.017)	(0.015)	(0.017)	(0.015)	(0.016)	(0.016)	(0.017)	(0.020)
Main control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Government effectiveness	0.003							
	(0.020)							
Control of corruption		0.015						
		(0.017)						
Rule of law			0.010					
			(0.017)					
Inflation				-0.000				
				(0.001)				
Trade restrictions					-0.015			
					(0.073)			
Financial development						-0.013		
						(0.071)		
Political system							0.001	
							(0.002)	
Regime durability								-0.009
								(0.021)
Constant	1.246***	1.167***	1.242***	1.198***	1.342**	1.174**	1.137***	1.435**
	(0.424)	(0.437)	(0.415)	(0.414)	(0.522)	(0.510)	(0.368)	(0.568)
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fisher test (p-value)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Arellano-Bond test for AR(1) (p-value)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Arellano-Bond test for AR(2) (p-value)	0.785	0.749	0.815	0.947	0.737	0.549	0.996	0.662
Hansen test of overidentifying restrictions (p-value)	0.899	0.892	0.905	0.813	0.832	0.964	0.954	0.982
Number of instruments	34	34	34	34	34	37	34	34
Number of countries	131	131	131	131	131	130	126	127
Average observations per country	8.76	8.76	8.76	8.76	8.76	8.76	8.80	8.78
Observations	1,148	1,148	1,148	1,148	1,148	1,139	1,109	1,115

Notes: In all the regressions, as expected, the 95% confidence interval for the coefficient of the lagged dependent variable does not contain the value 1. In (1), (2), and (3), we exclude political stability to mitigate collinearity issues. Similarly, introducing additional controls one at a time into the baseline specification helps to address collinearity problems. Windmeijer-corrected robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3. Robustness checks – Entropy balancing: Summary statistics on balancing quality

	R-squared	F-statistic	p-value
Before balancing	0.485	151.75	0.000
After balancing	0	0	1

Notes: Results from a (weighted) regression of the treatment variable on the covariates. The treatment variable is "Innovation 1" (see Table A2). The set of covariates includes political stability, GDP growth, urban population, unemployment, trade openness, financial openness, and taxation. All the covariates are lagged by one year to prevent potential issues of reverse causality.

Table 4. Robustness checks – Entropy balancing method

Informal economy 1	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Innovation 1	-0.500***	-0.681***	-0.372***	-0.582***	-0.500***	-0.620***	-0.327***	-0.484***
	(0.071)	(0.063)	(0.050)	(0.054)	(0.062)	(0.068)	(0.048)	(0.053)
Covariates in the second step	No	No	No	No	Yes	Yes	Yes	Yes
Year fixed effects in the second step	No	Yes	No	Yes	No	Yes	No	Yes
Regional fixed effects in the second step	No	No	Yes	Yes	No	No	Yes	Yes
R-squared	0.198	0.340	0.410	0.498	0.378	0.426	0.595	0.626
Observations	1,137	1,137	1,137	1,137	1,137	1,137	1,137	1,137

Notes: This table presents the effect of innovation on the size of the informal economy, obtained using the entropy balancing method. The treatment variable is innovation. The outcome variable is the size of the informal economy. The set of covariates includes political stability, GDP growth, urban population, unemployment, trade openness, financial openness, and taxation. All the covariates are lagged by one year to prevent potential issues of reverse causality. In line with Balima et al. (2021), we include year and regional fixed effects in the second step of the entropy balancing methodology. Year fixed effects control for macroeconomic shocks or time-related factors impacting the informal economy and innovation. Regional fixed effects allow us to control for region-specific factors affecting the informal economy (see Elbahnasawy 2021) and innovation. In alignment with Elbahnasawy (2021), we consider the regions identified by The World Bank, which include Latin America and the Caribbean (LAC), North America (NA), South Asia (SA), Sub-Saharan Africa (SSA), Middle East and North Africa (MENA), East Asia and Pacific (EAP), and Europe and Central Asia (ECA). Unreported constant included. Robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5. Robustness checks – Restricted samples, two-step system GMM estimates

	(1) Developing countries	(2) Removing G7 countries	(3) Removing LAC and SSA countries
Informal economy 1 (lag)	0.895***	0.817***	0.631***
	(0.045)	(0.068)	(0.176)
Innovation 1	-0.050**	-0.039**	-0.088***
	(0.021)	(0.017)	(0.033)
Main control variables	Yes	Yes	Yes
Constant	0.171	1.329***	2.957*
	(0.290)	(0.480)	(1.758)
Time dummies	Yes	Yes	Yes
Fisher test (p-value)	0.000	0.000	0.000
Arellano-Bond test for AR(1) (p-value)	0.000	0.000	0.008
Arellano-Bond test for AR(2) (p-value)	0.700	0.903	0.698
Hansen test of overidentifying restrictions (p-value)	0.997	0.955	0.687
Number of instruments	32	34	28
Number of countries	84	124	80
Average observations per country	8.30	8.76	9.11
Observations	697	1,086	729

Notes: In all the regressions, as expected, the 95% confidence interval for the coefficient of the lagged dependent variable does not contain the value 1. The dependent variable is "Informal economy 1." See Table A2 for its description. G7 includes Canada, France, Germany, Italy, Japan, the United Kingdom, and the United States. LAC: Latin America and the Caribbean. SSA: Sub-Saharan Africa. Windmeijer-corrected robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6. Robustness checks – Alternative measures of the size of the informal economy, two-step system GMM estimates

	(1) "Informal economy 2"	(2) "Informal economy 3"
Informal economy 2 (lag)	0.562***	
	(0.171)	
Informal economy 3 (lag)		0.667***
		(0.110)
Innovation 1	-0.086**	-0.474**
	(0.039)	(0.220)
Main control variables	Yes	Yes
Constant	2.409***	2.301
	(0.824)	(1.569)
Time dummies	Yes	Yes
Fisher test (p-value)	0.000	0.000
Arellano-Bond test for AR(1) (p-value)	0.003	0.049
Arellano-Bond test for AR(2) (p-value)	0.210	0.329
Hansen test of overidentifying restrictions (p-value)	0.982	0.787
Number of instruments	31	53
Number of countries	131	65
Average observations per country	7.88	5.78
Observations	1,032	376

Notes: In both regressions, as expected, the 95% confidence interval for the coefficient of the lagged dependent variable does not contain the value 1. See Table A2 for a description of the variables. Windmeijer-corrected robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7. Robustness checks – Alternative measures of innovation, two-step system GMM estimates

	(1) "Innovation 2"	(2) "Innovation 3"
Informal economy 1 (lag)	0.915***	0.953***
	(0.029)	(0.021)
Innovation 2	-0.006**	
	(0.003)	
Innovation 3		-0.019**
		(0.009)
Main control variables	Yes	Yes
Constant	0.763**	0.293
	(0.320)	(0.225)
Time dummies	Yes	Yes
Fisher test (p-value)	0.000	0.000
Arellano-Bond test for AR(1) (p-value)	0.001	0.013
Arellano-Bond test for AR(2) (p-value)	0.284	0.782
Hansen test of overidentifying restrictions (p-value)	0.487	0.627
Number of instruments	37	36
Number of countries	114	114
Average observations per country	8.25	7.03
Observations	940	801

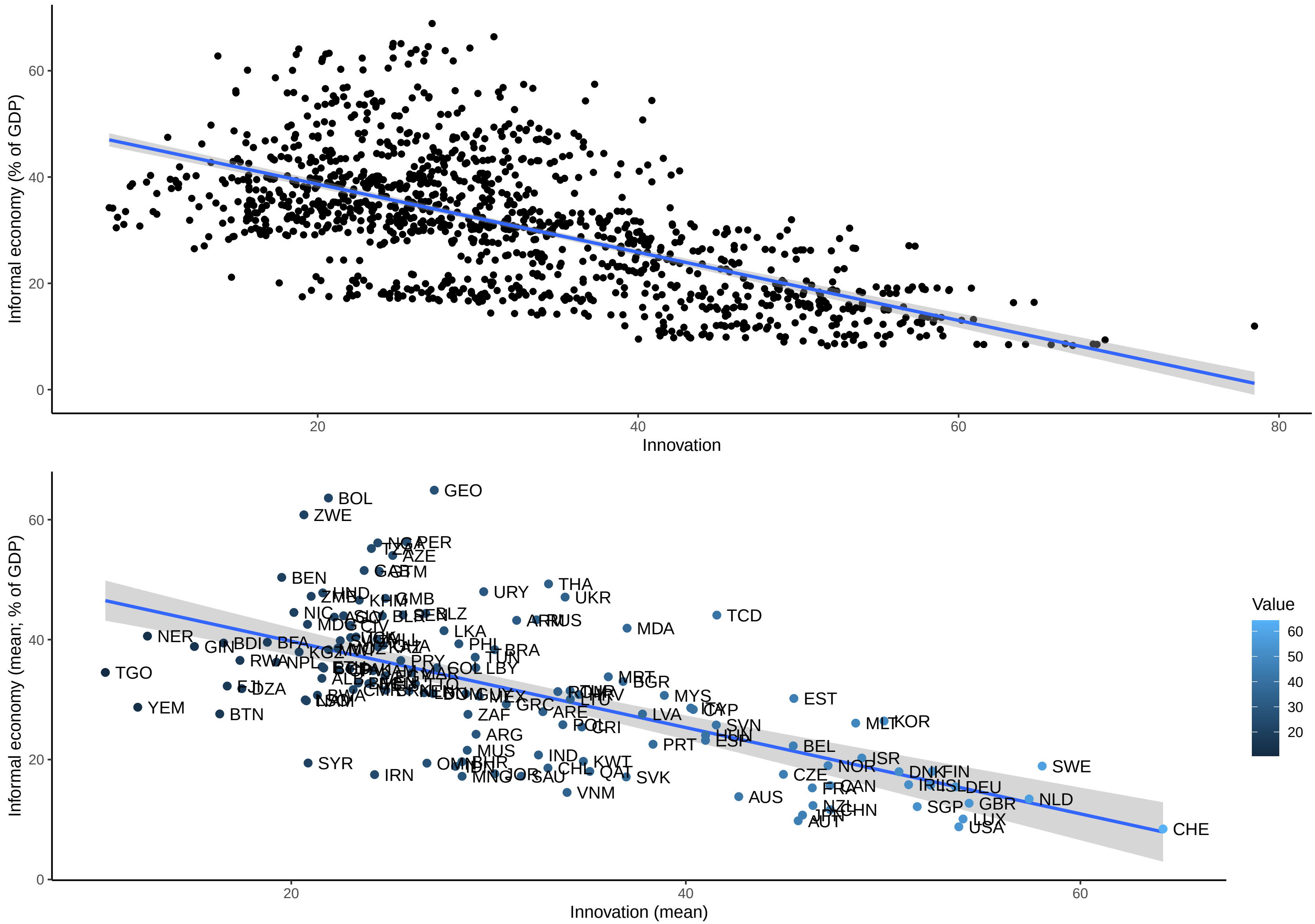
Notes: In both regressions, as expected, the 95% confidence interval for the coefficient of the lagged dependent variable does not contain the value 1. The dependent variable is "Informal economy 1." See Table A2 for a description of the variables. Windmeijer-corrected robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8. Analysis of transmission channels, using the entropy balancing method

(I) Economic development channel								
	(A) Impact of innovation on economic development							
Economic development	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Innovation 1	0.970***	1.346***	0.496***	0.893***	0.992***	1.144***	0.589***	0.763***
	(0.110)	(0.131)	(0.106)	(0.109)	(0.077)	(0.074)	(0.057)	(0.068)
Covariates in the second step	No	No	No	No	Yes	Yes	Yes	Yes
Year fixed effects in the second step	No	Yes	No	Yes	No	Yes	No	Yes
Regional fixed effects in the second step	No	No	Yes	Yes	No	No	Yes	Yes
R-squared	0.179	0.335	0.498	0.574	0.689	0.726	0.837	0.848
Observations	1,117	1,117	1,117	1,117	1,117	1,117	1,117	1,117
	(B) Impact of economic development on the size of the informal economy							
Informal economy 1	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Economic development	-0.193***	-0.197***	-0.261***	-0.268***	-0.190***	-0.192***	-0.278***	-0.277***
	(0.034)	(0.031)	(0.037)	(0.033)	(0.027)	(0.025)	(0.020)	(0.020)
Covariates in the second step	No	No	No	No	Yes	Yes	Yes	Yes
Year fixed effects in the second step	No	Yes	No	Yes	No	Yes	No	Yes
Regional fixed effects in the second step	No	No	Yes	Yes	No	No	Yes	Yes
R-squared	0.066	0.065	0.275	0.276	0.290	0.291	0.661	0.660
Observations	1,330	1,330	1,330	1,330	1,330	1,330	1,330	1,330
(II) Domestic credit mobilization channel								
	(A) Impact of innovation on domestic credit mobilization							
Domestic credit mobilization	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Innovation 1	0.696***	0.948***	0.299***	0.542***	0.729***	0.891***	0.366***	0.530***
	(0.113)	(0.096)	(0.094)	(0.090)	(0.103)	(0.113)	(0.091)	(0.097)
Covariates in the second step	No	No	No	No	Yes	Yes	Yes	Yes
Year fixed effects in the second step	No	Yes	No	Yes	No	Yes	No	Yes
Regional fixed effects in the second step	No	No	Yes	Yes	No	No	Yes	Yes
R-squared	0.151	0.279	0.349	0.396	0.338	0.391	0.524	0.543
Observations	1,058	1,058	1,058	1,058	1,058	1,058	1,058	1,058
	(B) Impact of domestic credit mobilization on the size of the informal economy							
Informal economy 1	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Domestic credit mobilization	-0.300***	-0.296***	-0.231***	-0.226***	-0.300***	-0.295***	-0.172***	-0.169***
	(0.030)	(0.029)	(0.020)	(0.020)	(0.030)	(0.029)	(0.023)	(0.023)
Covariates in the second step	No	No	No	No	Yes	Yes	Yes	Yes
Year fixed effects in the second step	No	Yes	No	Yes	No	Yes	No	Yes
Regional fixed effects in the second step	No	No	Yes	Yes	No	No	Yes	Yes
R-squared	0.325	0.325	0.397	0.399	0.372	0.375	0.523	0.523
Observations	1,253	1,253	1,253	1,253	1,253	1,253	1,253	1,253
(III) E-government channel								
	(A) Impact of innovation on e-government							
E-government	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Innovation 1	0.199***	0.302***	0.100***	0.228***	0.184***	0.267***	0.075***	0.179***
	(0.019)	(0.025)	(0.031)	(0.026)	(0.017)	(0.018)	(0.019)	(0.021)
Covariates in the second step	No	No	No	No	Yes	Yes	Yes	Yes
Year fixed effects in the second step	No	Yes	No	Yes	No	Yes	No	Yes
Regional fixed effects in the second step	No	No	Yes	Yes	No	No	Yes	Yes
R-squared	0.239	0.568	0.502	0.721	0.614	0.756	0.745	0.829
Observations	577	577	577	577	577	577	577	577
	(B) Impact of e-government on the size of the informal economy							
Informal economy 1	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
E-government	-0.798***	-0.933***	-1.086***	-1.148***	-0.798***	-0.790***	-1.358***	-1.535***
	(0.166)	(0.134)	(0.318)	(0.320)	(0.143)	(0.114)	(0.144)	(0.163)
Covariates in the second step	No	No	No	No	Yes	Yes	Yes	Yes
Year fixed effects in the second step	No	Yes	No	Yes	No	Yes	No	Yes
Regional fixed effects in the second step	No	No	Yes	Yes	No	No	Yes	Yes
R-squared	0.255	0.410	0.436	0.508	0.509	0.569	0.726	0.741
Observations	741	741	741	741	741	741	741	741

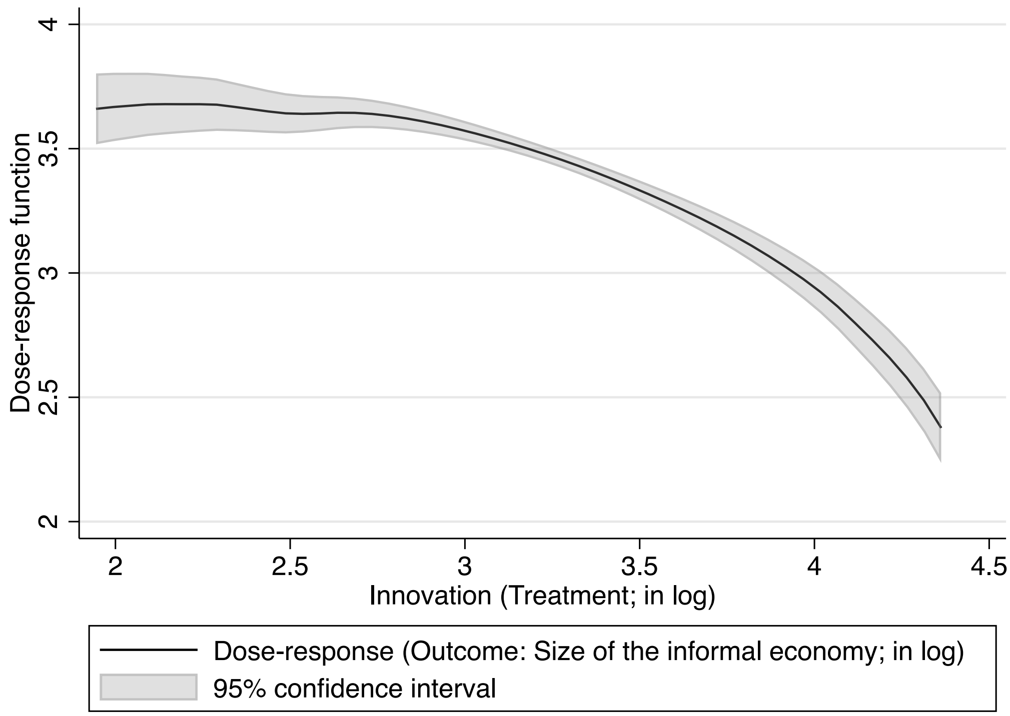
Notes: This table presents the effect of innovation on each channel on the one hand, and the influence of each channel on the size of the informal economy on the other hand, as obtained using the entropy balancing method. In line with Balima et al. (2021), we introduce year and regional fixed effects in the second step of the entropy balancing methodology. Year fixed effects control for macroeconomic shocks or time-related factors, while regional fixed effects allow us to control for region-specific factors. In alignment with Elbahnasawy (2021), we consider the regions identified by The World Bank, which include Latin America and the Caribbean (LAC), North America (NA), South Asia (SA), Sub-Saharan Africa (SSA), Middle East and North Africa (MENA), East Asia and Pacific (EAP), and Europe and Central Asia (ECA). Unreported constant included. Robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 1. Joint evolution of the informal economy and innovation (2007-2018)



Source: Authors.

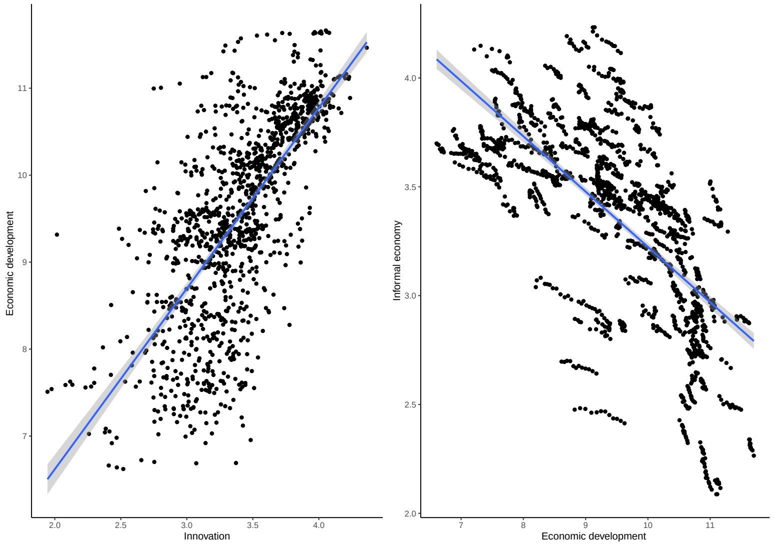
Figure 2. Dose-response function



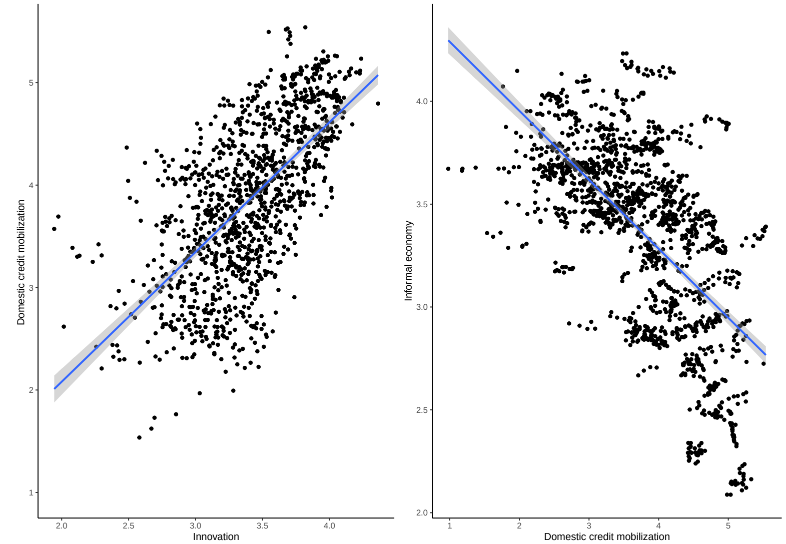
Source: Authors.

Notes: The treatment variable is "Innovation 1," and the outcome variable is "Informal economy 1."

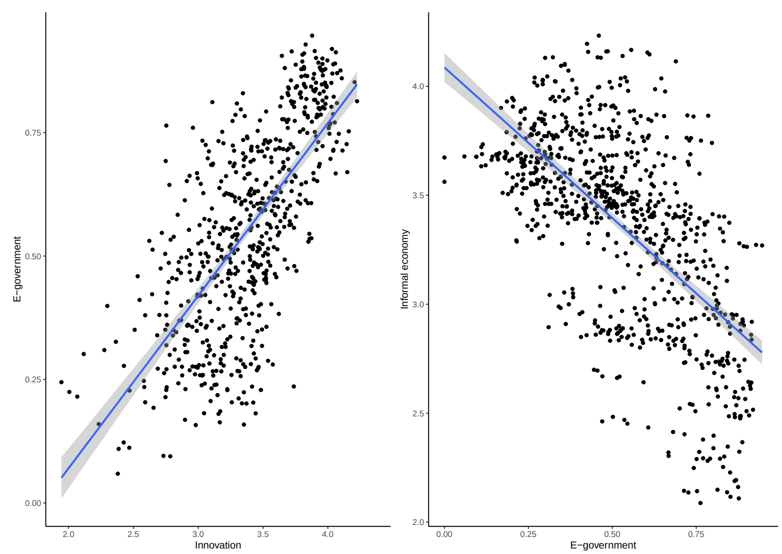
Figure 3. Visualizing the mediating role of the transmission channels through graphical representation



(a) Economic development channel



(b) Domestic credit mobilization channel



(c) E-government channel

Source: Authors.

Appendix A

Table A1. List of the countries

Albania	Cabo Verde	Gambia	Kuwait	New Zealand	Sri Lanka
Algeria	Cambodia	Georgia	Kyrgyz Republic	Nicaragua	Swaziland
Angola	Cameroon	Germany	Latvia	Niger	Sweden
Argentina	Canada	Ghana	Lebanon	Nigeria	Switzerland
Armenia	Chile	Greece	Lesotho	Norway	Syrian Arab Republic
Australia	China	Guatemala	Libya	Oman	Tajikistan
Austria	Colombia	Guinea	Lithuania	Pakistan	Tanzania
Azerbaijan	Costa Rica	Guyana	Luxembourg	Paraguay	Thailand
Bahrain	Côte d'Ivoire	Honduras	Madagascar	Peru	Togo
Bangladesh	Croatia	Hungary	Malawi	Philippines	Trinidad and Tobago
Belarus	Cyprus	Iceland	Malaysia	Poland	Tunisia
Belgium	Czech Republic	India	Mali	Portugal	Turkey
Belize	Denmark	Indonesia	Malta	Qatar	Uganda
Benin	Dominican Republic	Iran	Mauritania	Romania	Ukraine
Bhutan	Ecuador	Ireland	Mauritius	Russian Federation	United Arab Emirates
Bolivia	Egypt	Israel	Mexico	Rwanda	United Kingdom
Bosnia and Herzegovina	El Salvador	Italy	Moldova	Saudi Arabia	United States
Botswana	Estonia	Jamaica	Mongolia	Senegal	Uruguay
Brazil	Ethiopia	Japan	Morocco	Singapore	Venezuela
Brunei Darussalam	Fiji	Jordan	Mozambique	Slovak Republic	Vietnam
Bulgaria	Finland	Kazakhstan	Namibia	Slovenia	Yemen
Burkina Faso	France	Kenya	Nepal	South Africa	Zambia
Burundi	Gabon	Korea (Rep.)	Netherlands	Spain	Zimbabwe

Table A2. Definition of the variables and data sources

	Definition	Source
Informal economy 1	Output of the informal economy as a percentage of the official GDP. (in log)	Elgin et al. (2022)
Informal economy 2	Output of the informal economy as a percentage of the official GDP. (in log)	Medina & Schneider (2019)
Informal economy 3	Informal employment as a percentage of the total non-agricultural employment. (in log)	ILOSTAT
Innovation 1	Innovation output index. Ranges from 0 to 100. (in log)	WIPO, Cornell University, and INSEAD
Innovation 2	Patent applications by residents. (in log)	WDI
Innovation 3	R&D intensity, measured as R&D expenditure as a percentage of GDP. (in log)	WDI
Political stability	Political stability and absence of violence/terrorism index. Ranges from approximately -2.5 (weak performance) to approximately 2.5 (strong performance).	WGI
GDP growth	Annual percentage growth rate of the real GDP.	WDI
Urban population	Urban population as a percentage of the total population. (in log)	WDI
Unemployment	Unemployment as a percentage of the total labor force. (in log)	WDI
Trade openness	Sum of exports and imports of goods and services as a percentage of GDP. (in log)	WDI
Financial openness	De jure capital account openness index. The higher the index, the more financially open the country is.	Chinn & Ito (2006)
Taxation	Total tax and contribution as a percentage of profit. (in log)	WDI
Government effectiveness	Government effectiveness index. Ranges from approximately -2.5 (weak performance) to approximately 2.5 (strong performance).	WGI
Control of corruption	Control of corruption index. Ranges from approximately -2.5 (weak performance) to approximately 2.5 (strong performance).	WGI
Rule of law	Rule of law index. Ranges from approximately -2.5 (weak performance) to approximately 2.5 (strong performance).	WGI
Inflation	Inflation, measured by the annual growth rate of the GDP implicit deflator.	WDI
Trade restrictions	De jure trade globalization index . Ranges from 0 (high restrictions) to 100 (low restrictions). (in log)	Dreher (2006) and Gygli et al. (2019)
Financial development	Financial development index. Ranges from 0 to 1.	Svirydzenka (2016)
Political system	Polity index. Ranges from -10 (strongly autocratic) to 10 (strongly democratic).	Polity5 Database version 2018
Regime durability	Number of years since the most recent regime change or the end of transition period defined by the lack of stable political institutions. (in log)	Polity5 Database version 2018
Economic development	GDP per capita in constant 2017 international Dollar (PPP). (in log)	WDI
Domestic credit mobilization	Domestic credit to private sector as a percentage of GDP. (in log)	WDI
E-government	E-government development index. Ranges from 0 to 1.	United Nations E-Government Development Database

Notes: log: Natural logarithm. PPP: Purchasing Power Parity. WDI: World Development Indicators (The World Bank). WGI: Worldwide Governance Indicators. WIPO: World Intellectual Property Organization. ILOSTAT: Database on labor statistics by the International Labour Organization (ILO).

Table A3. Summary statistics

	Observation	Mean	Std. dev.	Min.	Max.
Informal economy 1	1,652	3.36	0.45	2.09	4.23
		[31.54]	[12.41]	[8.07]	[68.91]
Informal economy 2	1,518	3.19	0.50	1.63	4.14
		[27.15]	[11.71]	[5.10]	[62.80]
Informal economy 3	513	3.02	1.11	0.26	4.57
		[33.30]	[27.10]	[1.30]	[96.20]
Innovation 1	1,259	3.39	0.40	1.94	4.36
		[31.95]	[12.12]	[6.99]	[78.47]
Innovation 2	1,163	5.60	2.82	0	14.15
		[16,198.74]	[90,488.91]	[1]	[1,393,815]
Innovation 3	991	-0.53	1.22	-4.57	1.57
		[1.04]	[1.01]	[0.01]	[4.80]
Political stability	1,644	-0.08	0.92	-3.01	1.62
GDP growth	1,652	3.43	4.85	-50.34	86.83
Urban population	1,656	4.01	0.47	2.29	4.61
		[60.26]	[22.47]	[9.86]	[100]
Unemployment	1,656	1.75	0.80	-2.21	3.39
		[7.55]	[5.53]	[0.11]	[29.62]
Trade openness	1,602	4.34	0.50	3.03	6.08
		[87.60]	[51.21]	[20.72]	[437.33]
Financial openness	1,632	0.53	1.60	-1.93	2.31
Taxation	1,580	3.63	0.45	2.08	5.66
		[42.10]	[25.36]	[8.00]	[285.90]
Government effectiveness	1,644	0.11	0.94	-2.26	2.43
Control of corruption	1,644	0.03	1.01	-1.68	2.44
Rule of law	1,644	0.05	0.97	-2.26	2.12
Inflation	1,652	5.77	9.84	-27.63	200.77
Trade restrictions	1,644	4.00	0.44	2.48	4.55
		[59.48]	[22.46]	[11.89]	[94.35]
Financial development	1,644	0.36	0.24	0.06	1
Political system	1,584	4.65	5.95	-10	10
Regime durability	1,592	2.92	1.19	0.00	5.14
		[30.66]	[31.61]	[0.00]	[170.00]
Economic development	1,620	9.45	1.15	6.61	11.70
		[21,923.49]	[21,429.11]	[740.45]	[120,647.82]
Domestic credit mobilization	1,523	3.78	0.82	0.98	5.54
		[59.18]	[45.59]	[2.66]	[254.67]
E-government	822	0.53	0.20	0	0.95

Note: For the variables that are measured in natural logarithm, the values without logarithm are presented in brackets.

Table A4. Correlation coefficients

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
(1) Inf. eco. 1	1																							
(2) Inf. eco. 2	0.97	1																						
(3) Inf. eco. 3	0.59	0.67	1																					
(4) Innov. 1	-0.62	-0.64	-0.75	1																				
(5) Innov. 2	-0.48	-0.47	-0.28	0.55	1																			
(6) Innov. 3	-0.62	-0.67	-0.71	0.73	0.64	1																		
(7) Pol. stab.	-0.49	-0.55	-0.69	0.55	0.14	0.49	1																	
(8) GDP gr.	0.10	0.11	0.32	-0.15	-0.12	-0.18	-0.06	1																
(9) Urban pop.	-0.43	-0.45	-0.44	0.54	0.36	0.47	0.44	-0.19	1															
(10) Unemploy.	-0.01	-0.04	-0.29	0.12	0.10	0.11	0.09	-0.21	0.24	1														
(11) Tra. open.	-0.20	-0.23	-0.54	0.30	-0.22	0.13	0.46	0.02	0.25	0.07	1													
(12) Fin. open.	-0.44	-0.49	-0.56	0.50	0.06	0.33	0.49	-0.13	0.45	0.04	0.24	1												
(13) Taxation	0.19	0.19	0.15	-0.05	0.28	0.02	-0.22	-0.02	-0.09	0.11	-0.34	-0.23	1											
(14) Gov. eff.	-0.72	-0.77	-0.75	0.76	0.41	0.75	0.73	-0.14	0.55	0.05	0.34	0.62	-0.22	1										
(15) C. of corr.	-0.69	-0.75	-0.72	0.68	0.32	0.69	0.76	-0.12	0.50	0.06	0.32	0.56	-0.23	0.93	1									
(16) R. of law	-0.72	-0.79	-0.76	0.74	0.37	0.73	0.76	-0.13	0.49	0.08	0.33	0.61	-0.23	0.95	0.95	1								
(17) Inflation	0.17	0.19	0.29	-0.20	-0.06	-0.22	-0.24	0.04	-0.17	0.01	-0.13	-0.23	0.11	-0.30	-0.27	-0.29	1							
(18) Trade rest.	-0.51	-0.55	-0.73	0.65	0.31	0.54	0.52	-0.22	0.56	0.18	0.41	0.59	-0.27	0.71	0.62	0.68	-0.21	1						
(19) Fin. dev.	-0.73	-0.75	-0.60	0.73	0.66	0.74	0.52	-0.17	0.57	0.04	0.15	0.52	-0.16	0.84	0.76	0.80	-0.27	0.65	1					
(20) Pol. sys.	-0.13	-0.19	-0.48	0.31	0.11	0.31	0.36	-0.12	0.18	0.23	0.02	0.36	0.09	0.44	0.41	0.44	-0.17	0.34	0.35	1				
(21) R. durab.	-0.54	-0.55	-0.44	0.51	0.32	0.43	0.57	-0.09	0.42	0.10	0.13	0.44	-0.09	0.62	0.58	0.60	-0.18	0.46	0.56	0.24	1.00			
(22) Econ. dev.	-0.65	-0.68	-0.79	0.71	0.44	0.64	0.61	-0.22	0.80	0.16	0.37	0.61	-0.26	0.80	0.71	0.75	-0.23	0.76	0.77	0.24	0.54	1		
(23) D. cred. m.	-0.61	-0.62	-0.56	0.64	0.47	0.63	0.51	-0.20	0.52	0.06	0.29	0.46	-0.23	0.75	0.68	0.71	-0.29	0.67	0.81	0.29	0.48	0.71	1	
(24) E-gov.	-0.63	-0.68	-0.74	0.70	0.54	0.71	0.58	-0.27	0.69	0.09	0.27	0.60	-0.15	0.82	0.73	0.77	-0.21	0.76	0.81	0.36	0.57	0.87	0.75	1

Notes: Inf. eco. 1: Informal economy 1; Inf. eco. 2: Informal economy 2; Inf. eco. 3: Informal economy 3; Innov. 1: Innovation 1; Innov. 2: Innovation 2; Innov. 3: Innovation 3; Pol. stab.: Political stability; GDP gr.: GDP growth; Urban pop.: Urban population; Unemploy.: Unemployment; Tra. open.: Trade openness; Fin. open.: Financial openness; Gov. eff.: Government effectiveness; C. of corr.: Control of corruption; R. of law: Rule of law; Trade rest.: Trade restrictions; Fin. dev.: Financial development; Pol. sys.: Political system; R. durab.: Regime durability; Econ. dev.: Economic development; D. cred. m.: Domestic credit mobilization; E-gov.: E-government.

Table A5. Variance inflation factors (VIFs) in the different specifications

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Innovation 1	1.92	2.72	2.27	2.60	1.92	2.24	2.64	1.93	1.97
Political stability	1.90				1.91	1.91	2.00	2.00	2.27
GDP growth	1.16	1.15	1.15	1.15	1.16	1.20	1.17	1.16	1.17
Urban population	1.71	1.76	1.73	1.69	1.71	1.78	1.81	1.74	1.76
Unemployment	1.17	1.17	1.17	1.17	1.17	1.22	1.18	1.30	1.18
Trade openness	1.44	1.32	1.31	1.31	1.44	1.53	1.52	1.42	1.49
Financial openness	1.66	1.71	1.68	1.75	1.67	1.80	1.66	1.79	1.74
Taxation	1.23	1.24	1.25	1.25	1.23	1.28	1.26	1.29	1.23
Government effectiveness		3.22							
Control of corruption			2.31						
Rule of law				2.85					
Inflation					1.08				
Trade restrictions						2.86			
Financial development							2.65		
Political system								1.45	
Regime durability									1.91
Mean	1.52	1.79	1.61	1.72	1.48	1.76	1.77	1.57	1.63

Notes: We report the VIFs using "Innovation 1," our primary measure of innovation, in the regressions. Utilizing alternative measures, namely "Innovation 2" and "Innovation 3," produces the same conclusion. Results obtained with these alternative measures are available upon request. In models (2), (3), and (4), we exclude political stability to mitigate collinearity issues. Similarly, introducing additional variables one at a time into the baseline specification helps address collinearity problems.

Table A6. Additional robustness

	OLS estimates	Fixed-effects estimates	Random-effects estimates
Innovation 1	-0.015***	-0.015**	-0.018***
	(0.005)	(0.007)	(0.006)
Political stability	-0.015***	-0.015***	-0.016***
	(0.003)	(0.004)	(0.004)
GDP growth	-0.002***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)
Urban population	-0.063*	-0.063	-0.125**
	(0.037)	(0.068)	(0.056)
Unemployment	0.031***	0.031***	0.031***
	(0.004)	(0.007)	(0.007)
Trade openness	0.003	0.003	0.001
	(0.007)	(0.010)	(0.010)
Financial openness	-0.002	-0.002	-0.003
	(0.001)	(0.002)	(0.002)
Taxation	-0.013	-0.013	-0.013
	(0.009)	(0.014)	(0.014)
Constant	3.780***	3.652***	3.936***
	(0.153)	(0.273)	(0.220)
Time dummies	Yes	Yes	Yes
Country dummies	Yes	No	No
Fisher/Wald test (p-value)	0.000	0.000	0.000
Observations	1,148	1,148	1,148

Notes: The dependent variable is "Informal economy 1." OLS: Ordinary least squares. Robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A7. Entropy balancing – Summary statistics on balancing quality when the treatment variable is economic development

	R-squared	F-statistic	p-value
Before balancing	0.761	602.05	0.000
After balancing	0	0	1

Notes: Results from a (weighted) regression of the treatment variable on the covariates. The treatment variable is "Economic development." The set of covariates includes political stability, GDP growth, urban population, unemployment, trade openness, financial openness, and taxation. All the covariates are lagged by one year to prevent potential issues of reverse causality.

Table A8. Entropy balancing – Summary statistics on balancing quality when the treatment variable is domestic credit mobilization

	R-squared	F-statistic	p-value
Before balancing	0.431	135.06	0.000
After balancing	0	0	1

Notes: Results from a (weighted) regression of the treatment variable on the covariates. The treatment variable is "Domestic credit mobilization." The set of covariates includes political stability, GDP growth, urban population, unemployment, trade openness, financial openness, and taxation. All the covariates are lagged by one year to prevent potential issues of reverse causality.

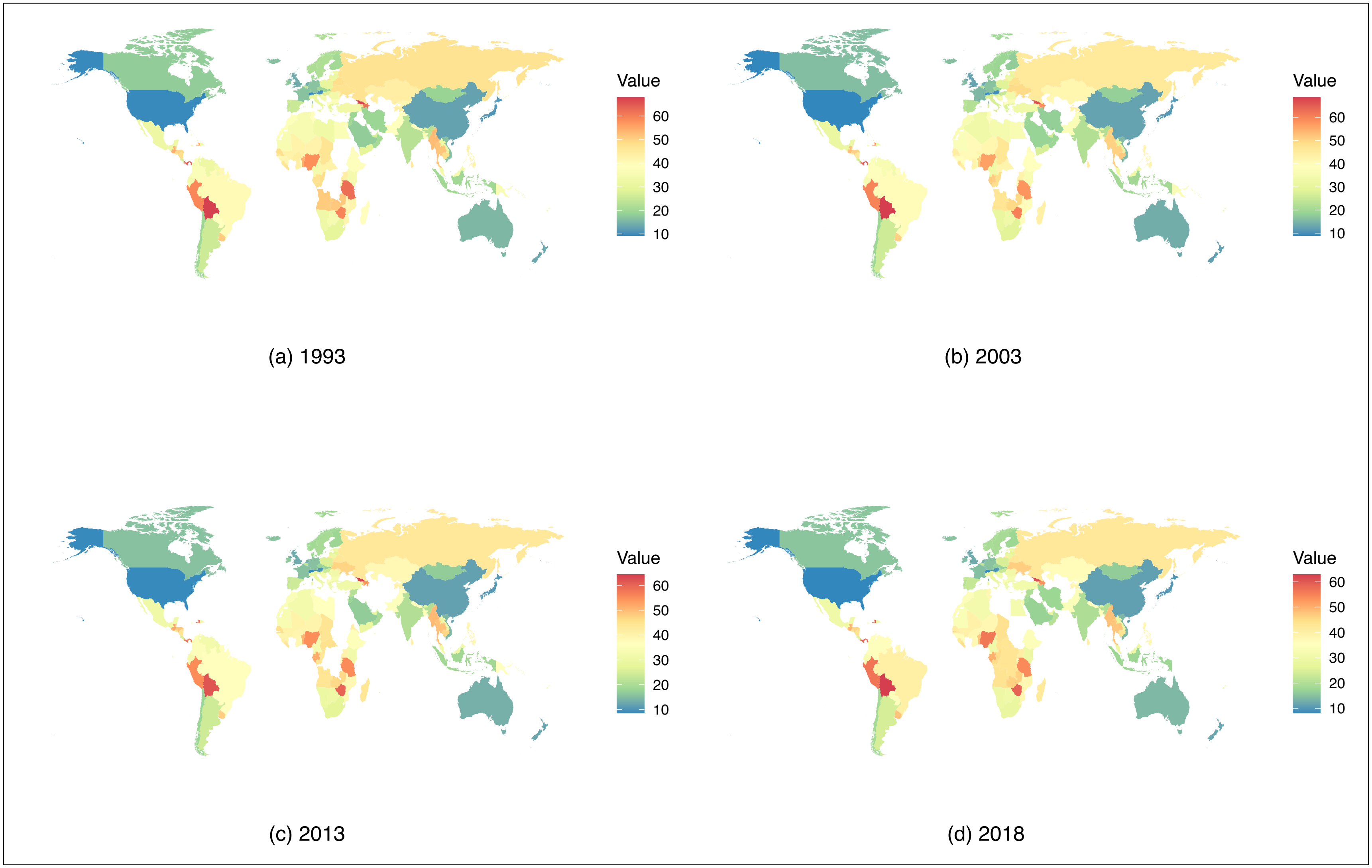
Table A9. Entropy balancing – Summary statistics on balancing quality when the treatment variable is e-government

	R-squared	F-statistic	p-value
Before balancing	0.622	172.82	0.000
After balancing	0	0	1

Notes: Results from a (weighted) regression of the treatment variable on the covariates. The treatment variable is "E-government." The set of covariates includes political stability, GDP growth, urban population, unemployment, trade openness, financial openness, and taxation. All the covariates are lagged by one year to prevent potential issues of reverse causality.

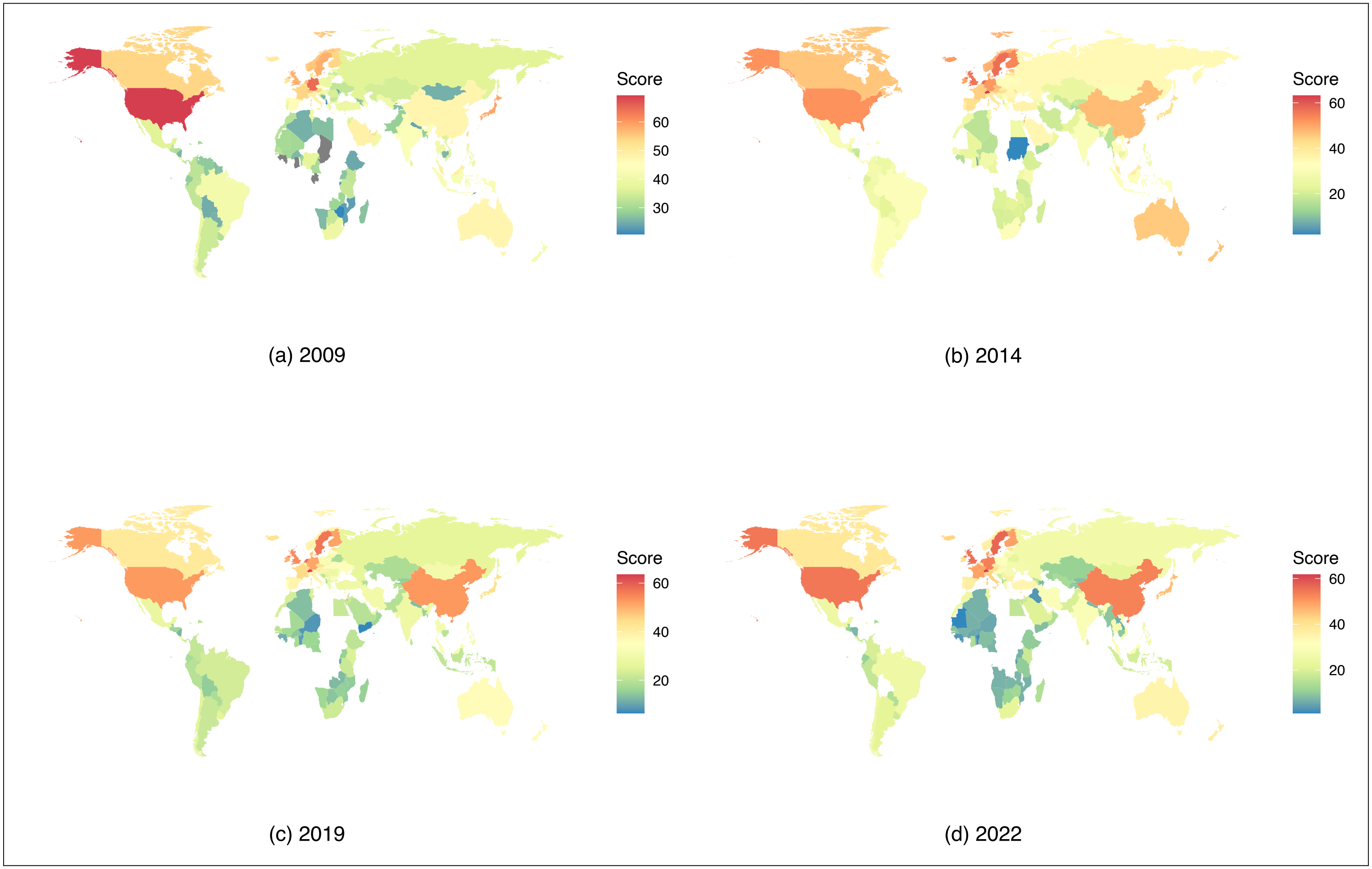
Appendix B

Figure B1. World maps depicting the informal economy (% of GDP)



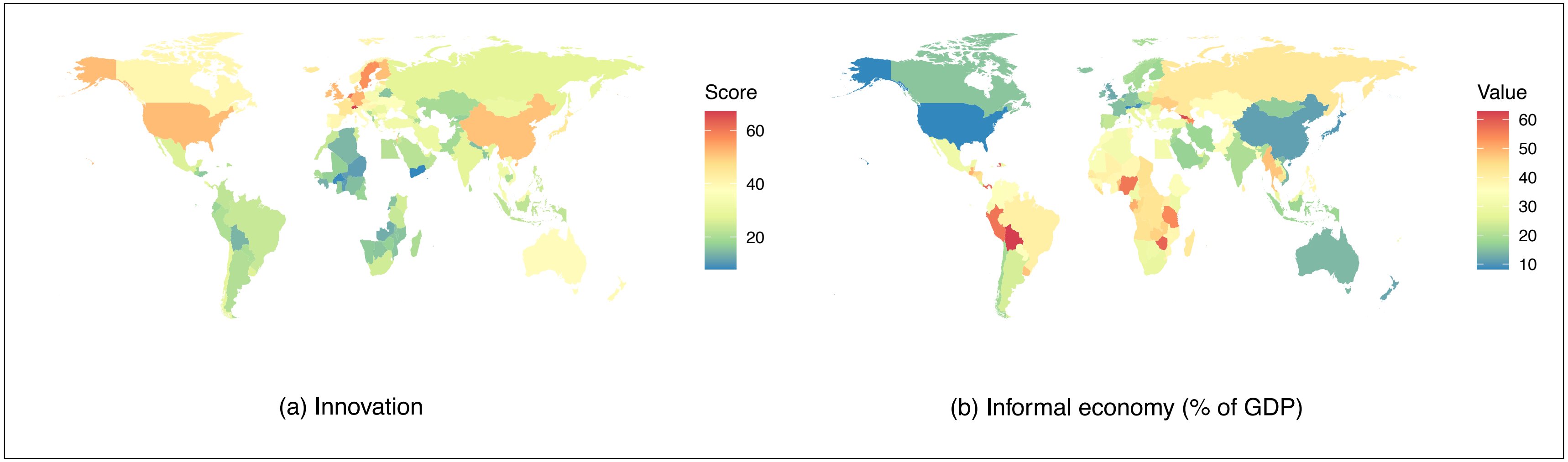
Source: Authors.

Figure B2. World maps depicting innovation (innovation output index)



Source: Authors.

Figure B3. World maps depicting innovation and the informal economy in 2018



Source: Authors.

Appendix C – Entropy balancing methodology

This paper aims to analyze the effect of innovation production on the size of the informal economy. Establishing a causal link between innovation and the informal economy is a more significant challenge than it may initially appear. Indeed, innovation may be endogenous for various reasons. One of the most critical factors is that achieving a certain level or intensity of innovation, whether low, medium, or high, is not a random process. The level or intensity of innovation can be influenced by a country's political stability, economic growth, urbanization, unemployment, trade openness, financial openness, and taxation, among other factors. In econometric terms, this introduces a selection bias. These factors may also affect the size of the informal economy, making innovation endogenous.

Entropy balancing is a novel impact evaluation technique that effectively corrects for the endogeneity of innovation. This method was initially developed by Hainmueller (2012) for binary treatments and has been extended more recently by Vegetabile et al. (2021) and Tübbicke (2022) to accommodate continuous treatments. In this paper, for robustness checks, we employ the extended entropy balancing method tailored for continuous treatments, given that our treatment variable, innovation, is continuous. The outcome variable is the size of the informal economy. It is worth noting that the existing empirical literature on entropy balancing has primarily focused on binary treatments. This paper is one of the first to employ the extended entropy balancing methodology for continuous treatments.

In our context, where the treatment is continuous, all units received some treatment with different intensity or dose.[16] Estimating the treatment effect of innovation using the entropy balancing method involves two consecutive steps.

The first step consists in computing weights so that in the re-weighted sample, the balancing property is respected. Entropy balancing for continuous treatments is essentially a weight-based covariate balancing scheme that addresses a globally convex optimization problem to derive balancing weights. This is achieved by minimizing the deviation from (uniform) base weights while adhering to zero correlation and normalization constraints.

Assume that T_i is a non-negative variable denoting the treatment variable (innovation) for unit i, where treated units have $T_i > 0$. Unlike the binary treatment case, balancing weights (w_i) are computed for the treated units. This is done to estimate the average outcomes of treated units under specific treatment doses or intensities (Tübbicke 2022). Define $X_i \in \mathbb{R}^K$ as a vector of pre-treatment covariates and Y_i as a post-treatment outcome (the size of the informal economy) for unit i, where K is the number of covariates. Let $\tilde{X}_i$ be the de-meaned

[16] The assessment of the impact of continuous treatments has attracted significant interest recently in the economics literature. Tübbicke (2022) briefly reviews some of the recent studies.

version of X_i. Similarly, define $\tilde{T}_i^r$ as the de-meaned rth order term of the treatment intensity. Consider the column vector $g(r, \tilde{X}_i, \tilde{T}_i) = (\tilde{X}_i', \tilde{T}_i, ..., \tilde{T}_i^r, \tilde{X}_i'\tilde{T}_i, ..., \tilde{X}_i'\tilde{T}_i^r)'$. Entropy balancing weights are obtained by solving the following constrained optimization problem:

$$\begin{cases} & \min_{w} H(w) = \sum_{i|T_i>0} h(w_i) \\ \text{s.t.:} & \sum_{i|T_i>0} w_i g(r, \tilde{X}_i, \tilde{T}_i) = 0, \\ & \sum_{i|T_i>0} w_i = 1, \\ & \text{and } w_i > 0 \ \forall i|T_i > 0 \end{cases} \tag{3}$$

where H is the loss function. It is minimized subject to both balancing constraints in terms of $g(r, \tilde{X}_i, \tilde{T}_i)$ and normalizing constraints that weights are strictly positive and sum up to one. The fundamental concept of entropy balancing is to calculate weights that render the treatment variable uncorrelated with the covariates. However, despite its intuitiveness, this approach may prove insufficient in eliminating bias arising from observed covariates when analyzing the causal effect of a continuous treatment (Tübbicke 2022, Yiu & Su 2018). Indeed, as emphasized by Tübbicke (2022), achieving uncorrelatedness between $\tilde{T}_i$ and $\tilde{X}_i$ does not guarantee independence; the distributions of covariates may still differ across the treatment intensity distribution, even with flexible $\tilde{X}_i$. To overcome this challenge, entropy balancing weights also ensure that higher orders of the treatment variable are uncorrelated with the covariates. This is operationalized by choosing r that shows the order to which the treatment variable (innovation) has been rendered uncorrelated with the covariates. In practice, weights are usually estimated for $r = 1, 2,$ and 3. Through Monte-Carlo simulations, Tübbicke (2022) demonstrated that setting $r = 2$ yields superior results in terms of both bias and Root Mean Squared Error (RMSE) compared to choices of $r = 1$ and $r = 3$. From this backdrop, we set $r = 2$ in our analysis. This means that both T_i and T_i^2 are rendered uncorrelated with the covariates. As it will be seen, this leads to achieve excellent covariate balance.

The optimization problem is solved using the Lagrange method. The entropy metric by Kullback (1959) is employed, defined as $h(w_i) = w_i \ln(w_i/q_i)$, where a uniform base weight scheme $q_i = 1/N_1$, with N_1 representing the size of the treatment group, is set as the default. The loss function is undefined for non-positive weights and reaches its minimum when $w_i = q_i$. Therefore, when using the Lagrange method, the normalizing constraint that weights must be positive can be omitted. This yields the following equation:

$$\min_{w,\lambda,\gamma} \mathcal{L}(w, \lambda, \gamma) = \sum_{i=1}^{N_1} w_i \ln(w_i/q_i) - \lambda \left(\sum_{i=1}^{N_1} w_i - 1 \right) - \gamma' \left(\sum_{i=1}^{N_1} w_i g(r, \tilde{X}_i, \tilde{T}_i) \right) \tag{4}$$

where $\mathcal{L}$ is the Lagrange function, and λ and γ are Lagrange multipliers. Solving equation (2) yields the weights (w_i) as a function of the base weights (q_i), γ, and the data $g(r, \tilde{X}_i, \tilde{T}_i)$. We obtain the following equation:

$$w_i = \frac{q_i \exp\left(\gamma' g(r, \tilde{X}_i, \tilde{T}_i)\right)}{\sum_{i=1}^{N_1} q_i \exp\left(\gamma' g(r, \tilde{X}_i, \tilde{T}_i)\right)} \tag{5}$$

where λ was cancelled out. To obtain the final expression for w_i, we need to estimate γ. To do so, we proceed in three steps. First, we derive the dual Lagrange function ($\mathcal{L}^d$) by substituting (3) into the Lagrange function $\mathcal{L}$. Then, we determine γ^*, the value of γ at the optimum, by differentiating $\mathcal{L}^d$ with respect to γ. Finally, we obtain the balancing weights by replacing γ with γ^* in equation (3).

In the second step, the balancing weights obtained in the first step are used in a regression analysis to determine the treatment effect and the Dose-Response Function (DRF). More specifically, the treatment effect of innovation is obtained using the weighted least squares method. We perform a nonparametric estimation of the DRF using local linear regression with an Epanechnikov kernel.

By combining weighting and regression approaches, the entropy balancing method has several advantages over conventional impact evaluation methods such as Generalized Propensity Score (GPS) or Difference-in-Differences (DID). First, the re-weighting scheme of entropy balancing allows us to achieve a high degree of covariate balance even in the case of small samples. Second, as stressed by Tübbicke (2022), entropy balancing helps obviate the estimation of the GPS, which is notoriously difficult to estimate. Many GPS-based methods may require implementing an iterative estimation procedure until satisfactory covariate balance is achieved. Third, the entropy balancing scheme is nonparametric. This means that there is no need to define an empirical model for either the treatment variable (innovation) or the outcome variable (size of the informal economy) to obtain balancing weights. Consequently, entropy balancing reduces model dependency by avoiding misspecification issues related, for instance, to the functional form of the empirical model.

Fourth, the treatment effects obtained using the entropy balancing method are not biased by multicollinearity issues. The covariates are actually orthogonalized with respect to the treatment variable. Fifth, entropy balancing allows us to avoid information loss and retain efficiency for the subsequent estimations by using a more flexible re-weighting scheme that keeps the weights as close as possible to the base weights. Sixth, as the entropy balancing method combines weighting and regression analysis, it allows us to account for unobservable factors related to the panel structure of the data in the estimation of the treatment effect by including fixed effects in the second step. Seventh, as shown by Tübbicke (2022) through Monte-Carlo simulations, the entropy balancing method for continuous treatment with $r = 2$ can outperform in terms of bias and RMSE other re-weighting approaches such as Generalized Boosted Modeling (see Zhu et al. 2015), Covariate Balancing GPS (see Fong et al. 2018), and Inverse Probability Weighting with continuous treatment (see Robins et al. 2000).

Eight, the entropy balancing methodology is very versatile. The balancing weights obtained in the first step can be used in any standard regression model of the outcome variable (size of the informal economy) on the treatment variable (innovation) to obtain the treatment effects, provided that this model is one that would have been estimated in the absence of any re-weighting scheme. Ninth, from a computational perspective, entropy balancing is appealing. The optimization problem used to compute the weights is globally convex and well-behaved. In general, it requires only a few seconds to attain the weighting solution, even in the case of moderately large datasets.