

Analysing the informal economy: Data challenges, research design, and research transparency

Rob Ackrill¹  | Eghosa Igudia² 

¹Department of Economics, Nottingham Business School, Nottingham Trent University, Nottingham, UK

²Department of Economics, Faculty of Business and Law, De Montfort University, Leicester, UK

Correspondence

Rob Ackrill, Department of Economics, Nottingham Business School, Nottingham Trent University, Nottingham NG1 4BU, UK.
Email: robert.ackrill@ntu.ac.uk

Abstract

Analyses of the informal economy are vital to understanding economic activity in developing countries, but challenges abound when analysing something that is so difficult to measure. In this article, we explore these challenges and what they mean for research and research design in studies of the informal economy. We review key economic theories developed to study the informal economy, before critically discussing the range of methods of data collection available for researchers. All approaches to data collection have limitations, yet many studies fail to reflect adequately on these in their design and presentation. In reviewing data sources for economic analyses of informality, we give particular attention to surveys designed and led by academic researchers, the consideration of which is omitted from leading official organisations' own reports on data collection in the informal economy. In the absence of a unifying theory and data sources that deliver the 'comprehensive and accurate' data called for by some authors we argue, first, that no one source of data is intrinsically superior; and second, that research on the informal economy must embrace transparency—around the data used, their alignment

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Review of Development Economics* published by John Wiley & Sons Ltd.

with the underlying theory adopted, and the detailed arrangements for data collection and analysis.

KEYWORDS

data challenges, data collection, developing countries, informal economy, research design

1 | INTRODUCTION

It is 50 years since the notion of the ‘informal economy’ entered economic, political, and academic discourse (Hart, 1973; ILO, 1972). Its importance is without question. It was estimated before COVID-19 to employ 2 billion people, or over 60% of the world’s employed (ILO, 2018), representing over 90% of global micro and small enterprises.¹ For Sub-Saharan Africa, wherein the concept was first developed, the IMF (2017) estimated that over the period 2010–2014, the informal economy ranged from just over 20% of the whole economy in Mauritius to 65% in Nigeria, Africa’s largest economy and most populous country. That said, the ‘informal economy is difficult to measure [...] because activities within it cannot be directly observed, and for the most part, participants in the informal economy do not want to be accounted for’.²

Our purpose in this paper is to provide an overview of the challenges researchers face when seeking to analyse something that is so important, yet which cannot be directly observed. We have a particular empirical focus on informality in developing countries, because not only do they have relatively large informal economies, but also they are more likely to suffer from limitations regarding sufficient and reliable official data, that would enable research on the informal economy using indirect methods (see below).

We present several directions for future research to help address some of these challenges. We first set out various economic theories that have sought to frame the informal economy. We then consider the practical implementation of such research, in terms of the core triad of data needs, methods of data collection, and methods of data analysis. Orthodox economic perspectives on informal economy research frequently follow standard methodological choices in economics, including the research methods adopted and techniques applied to the chosen data. The difficulty in measuring the informal economy, however, can result in significant challenges in applying these techniques. From the foregoing, we derive our research question:

RQ. What are the data challenges faced in economic analyses of the informal economy?

In what follows, we explore the conceptual boundaries that have shaped, and in some cases arguably limited, economic analyses and understandings of informality, especially in developing countries. We reflect on the elements of four key economic theories of the informal economy, discussed in the next section, each developed specifically for this purpose. From this, we explore their data needs, and associated methods of data collection and analysis. In this paper, we do not address in detail those theories that have been adopted from other relevant and cognate areas (e.g., theories of entrepreneurship; see, amongst many, Bruton et al., 2012; Galdino et al., 2018; Ketchen et al., 2014). An important distinction here is between the economic theories that have been developed specifically for analysis of the informal economy,

and those more general theories from other disciplines, such as around entrepreneurship, that have been adopted for use in analyses of the informal economy.

2 | ECONOMIC THEORIES OF THE INFORMAL ECONOMY: A REVIEW

In this section, we review key theories that dominate not only theoretical economic explanations for the existence of informality, but also empirical analyses of the nature and scale of its presence.³ We identify the key features of each of these theories. In so doing, and to aid analytical and narrative clarity, we merge some theories where appropriate, based on their key features. We then provide in Table 1 a concise summary of their key features, their strengths, and their weaknesses.

Before we introduce the different theories, we define the informal economy as embracing economic activities that take place outside of routinised government oversight (reflected in synonyms such as ‘untaxed economy’ or ‘hidden economy’). Its origins lie in efforts to understand the labour market characteristics of developing countries (e.g., Dell’Anno, 2008; Hart, 1973; ILO, 1972), where Hart observed workers in urban settings who appeared to be unemployed in a Western sense, yet were still engaged in economic activities that enabled them to survive.

Since these early studies on developing countries, much economic research has focused on developed countries and, from the early 1990s, the former centrally planned ‘transition’ economies of central, eastern, and south-eastern Europe, and former states of the Soviet Union. This shift has also influenced the nature of debates around the informal economy. Notably, a debate on the merits and feasibility of formalising informal economic activity in the countries of, say, south-eastern Europe (see, amongst many, Williams, 2023) will take a very different form to such a debate held regarding the countries of sub-Saharan Africa where, as noted, the informal economy may represent the majority of economic activity, in contexts of generally weaker institutions and governance.

In this context, debates will reflect much more fundamental issues, such as whether it is even appropriate to maintain a distinction between informal and formal economic activity (Sindzingre, 2006), when ‘as much as 90 percent of the money produced [...] comes through informal channels’ (Neuwirth, 2013, p. 66). Such distinctions can appear rather artificial when compared against citizens’ lived reality and ‘everyday’ experiences in the Global South (inter alia, Pratt, 2019). We start our review of theories of the informal economy with three theories—dualist, structuralist, and legalist—which have long-been prominent in research (see, among many, Chen et al., 2002; Roberts, 2014). We then explore jointly, for reasons set out below, voluntarist, neoliberal, and post-structuralist theories.

2.1 | Dualist theory

Dualist theory (also referred to as modernisation theory—see, inter alia, Williams & Nadin, 2010; Williams et al., 2012) views the economy as two discrete and essentially disconnected parts, the formal and the informal. Initially, this duality was between ‘the traditional/subsistence (i.e., informal) and modern/industrialized (i.e., formal) sectors’ (Dell’Anno, 2022, p. 1611). The implication was that industrialisation and urbanisation would

TABLE 1 Key characteristics of IE theories.

Theory	Notable characteristics	Strengths	Weaknesses
Dualist	<ul style="list-style-type: none"> • Unconnected, dichotomous economies, FE, and IE • Individuals engage in the IE for survival reasons due to a skills-gap, inadequate education; represent migrants to urban centres, experience poverty • IE arises because of systemic flaws in the FE, lack of economic growth, unemployment, corruption, and rapid population growth 	<ul style="list-style-type: none"> • Has helped to focus attention on IE • Highlights some of the key motivations for informality • Put spotlight in employment and unemployment • Logic of survival still relevant 	<ul style="list-style-type: none"> • Fails to address interconnected FE and IE • Withering of IE refuted empirically • Overlooks pull factors into the IE • Focuses on necessity over opportunity • Does not reflect informal institutions
Structuralist	<ul style="list-style-type: none"> • FE and IE are interconnected, interdependent • IE driven by firm-related factors including the nature of capitalism, subordinated and imbalanced firms in well-linked relationships, the need to be competitive, firms' responses to unions' power, and state regulations • Formal firms use IE to reduce bureaucracy and costs • Individuals engage in the IE because they need extra income, working at a second job/business. The IE represents a seed-bed for entrepreneurship 	<ul style="list-style-type: none"> • Advances dualist by recognising FE-IE links • Brings awareness to concerns over exploitation • Sees potential basis for developing entrepreneurial skills in the IE 	<ul style="list-style-type: none"> • Does not accommodate either the existence of the IE having positive aspects; nor the possibility of operating in the IE being a positive choice • Focuses on exploitation at the cost of potential earning opportunities in the IE • Focuses on opportunity in the FE • Does not reflect informal institutions
Legalist	<ul style="list-style-type: none"> • Firms pushed into the IE, to avoid bureaucracy and costs • Reduce or remove the bureaucracy and firms will formalise 	<ul style="list-style-type: none"> • Advances structuralist, by seeing IE as valid alternative to FE • Puts focus on legal-bureaucratic structures • Recognises opportunity in the IE 	<ul style="list-style-type: none"> • Motives for IE omit wider economic concerns and necessity to operate in the IE • Does not reflect informal institutions

TABLE 1 (Continued)

Theory	Notable characteristics	Strengths	Weaknesses
Voluntarist/ neoliberal/ post-structuralist	<ul style="list-style-type: none"> Firms choose to operate in the IE, to avoid bureaucracy and costs Individuals working in the IE are entrepreneurial, carrying out an implicit cost-benefit analysis before deciding to engage in the IE Signal for the government to reduce FE burdens 	<ul style="list-style-type: none"> Advances legalist, by seeing IE as valid alternative to FE Cost-benefit idea could aid policymakers set taxes and regulations, to give firms in FE the freedoms enjoyed by IE firms Celebrates opportunity in the IE 	<ul style="list-style-type: none"> Motives for IE omit wider economic concerns and necessity to operate in the IE Does not reflect informal institutions

Abbreviations: FE, formal economy; IE, informal economy.

see the informal wither away. As the ILO (1972) and Hart (1973), showed, however, the formal and informal coexisted *within* the urban labour market: the informal sector was not withering away. With hindsight, it is easy to see that early theorising on informality could not foresee the changing nature of informality that urbanisation would bring, in particular where industrialisation failed to create sufficient jobs to match the rate of rural-urban migration.

Dualist theory also leads to the widespread argument that informality should be discouraged and, ultimately, formalised. In the context of developing countries with significant informal economies and given inadequate formal urban job-creation, however, the informal economy provides essential goods and services either unavailable from the formal economy, or available from the informal economy at a lower price, even including goods and services from multinationals. The informal economy has also been particularly important at times of economic crisis. As such, dualism is empirically limited in its relevance, but it offers a historical reference point, from which other theories evolved.

2.2 | Structuralist theory

Structuralist theory was popularised by Moser (1978) and Castells and Portes (1989), and is also referred to as the political economy approach (Dell'Anno, 2022; see, inter alia, Williams & Kedir, 2018; Williams & Windebank, 2015). It argues that whilst the formal and informal economies in developing countries are distinct, they are interconnected. The informal economy exists as a result of structured development within the capitalist mode of production (Chen, 2012; Henry, 1978), with capitalist participants driven primarily by the desire to make a profit (Moser, 1978, p. 22). The informal economy serves the formal economy, as a result of formal firms' attempts to reduce labour costs, increase competitiveness, reduce the power of labour unions, avoid or limit state regulations, and respond to global competition and industrialisation.

This subordination of the informal to the formal leads, in particular, to the exploitation of 'women, youth, immigrants, and unskilled' (Dell'Anno, 2022, p. 1628), a view consistent with

another strand of the informal economy literature, where these groups are more likely to be found participating in the informal economy out of survival needs, rather than choice (Fairlie & Fossen, 2018; Williams et al., 2012). Structuralist perspectives on the informal economy see urbanisation and industrialisation as creating a new type of economy that sits within, and is enabled by, global deregulation, within which countries have ‘peripheral and semi-peripheral roles’ (Roberts, 2014, p. 422), ‘where economic value is re-appropriated from the periphery and semi-periphery to the core through processes of trade and investment’ (Roberts, 2014, p. 422). This theory therefore focuses on the countries of the Global South (*contra* Dell’Anno, 2022). Although structuralist theory does not foresee the natural withering of the informal economy, the exploitation it is subjected to presents a strong argument for policy efforts to formalise its activities (Williams et al., 2012, pp. 530–531), arising out of an insufficiently regulated economy (*inter alia*, Williams & Kedir, 2018).

2.3 | Legalist theory

Legalist theory contrasts with one key facet of structuralist theory: the existence of, and participation in, the informal economy does not arise out of a lack of official oversight of the formal economy, but out of too much. Firms wilfully choose to operate in the informal economy, rather than doing so out of necessity, in order to avoid the cost, time and processes associated with the formal registration of their businesses, and the costs of remaining formal: taxes, salaries, (over) regulation, and the high cost of public utilities (Becker, 2004; De Soto, 1989; Dell’Anno & Halicioglu, 2010). Thus, the root cause of informality is the ‘hostile legal system [which] leads the self-employed to operate informally with their own informal extra-legal norms’ (Chen, 2012).

Formalisation will enable informal firms to carry out asset conversion to real capital, and unleash their productive potential if the state simplifies bureaucratic process (Chen, 2012; De Soto, 1989). Legalist theory is thus distinct from the first two theories insofar as it starts from the existence of formal and informal economies, and then asks which one individuals looking to establish or operate a firm will choose; a choice based on the disincentives that bureaucracy creates for the formal option. Address that and firms will formalise.

2.4 | Voluntarist theory, neoliberal theory, and post-structural theory

The citations above confirm that those first three theories have been utilised in many papers over the years. We can also see a progression over time in thinking around formality and informality, captured in Table 1. Since then, theorising on the informal economy has developed in different ways. We identify here three theories—voluntarist, neoliberal, and post-structuralist—that, whilst having their differences, share a specific form of analytical progression beyond legalist theory.

Legalist theory implies that firms would operate in the formal economy, were it not for the rules and regulations that push individuals into the informal economy. In contrast, voluntarist theory, neoliberal theory, and post-structuralist theory see individuals as having agency over their decision-making, as a result of which they make a deliberate and rational choice as to which economy to operate in, on the basis of a(n implicit) cost–benefit analysis. Voluntarist

theory (see also ‘free-riding’: Dell’Anno, 2022), observes the deliberate choice made freely by individuals. Neoliberal theory also sees individuals acting rationally: indeed, they ‘are celebrated as heroes throwing off the shackles of a burdensome state’ (Williams et al., 2012, p. 531). The appropriate response to such informality is to deregulate the economy, to enable all firms to operate as these informal firms do (Williams et al., 2012).

Post-structuralist theory shares with voluntarist and neoliberal theories the feature of individual agency in decision-making. Where it differs is in the conditions under which individuals exercise that agency, which can be several and varied, ‘with complex and messy characters and logics’ (Williams et al., 2012). These motives contrast with voluntarist and neoliberal theories only insofar as they represent multiple potential agentic factors: developing personal relations, as a form of personal expression, for redistributive/social motivations, and as a deliberate act of resistance to exploitation in the neoliberal global economy (see, inter alia, Igudia et al., 2016; Williams & Nadin, 2010; Williams et al., 2012).

Post-structuralist theory also, therefore, has similarities with other strands of the informal economy literature: the gap, or incongruence, between formal institutions (the laws, policies, and so forth that either force or encourage participation in the informal economy) and the informal institutions of norms, beliefs, practices, kinship, and so forth (see, inter alia, Dell’Anno, 2022; Hyden & Williams, 1994; Igudia et al., 2022; Webb et al., 2009, 2013). Post-structuralist theory thus offers a bridge between theories of the informal economy and neo-institutionalist theory, notably through institutional incongruence.

3 | ANALYSES OF THE INFORMAL ECONOMY

The foregoing has set out the key focal points of analysis laid down in the principal theories of the informal economy. This raises important questions concerning the data required for such analyses. In this section, we reflect on the general approaches to data collection for informal economy analyses. This is necessary, given that the existing literature on the informal economy is noted for a lack of clarity over relevant theory, and methods of data collection and analysis (Dell’Anno & Halicioglu, 2010), whilst ‘the “appropriate” methodology to assess its scope has not yet been agreed upon’ (Feld & Schneider, 2010, p. 112). Similarly, Galdino et al. (2018, p. 241) found that ‘the majority of the studies, independently from the methodology used, fail to provide a thorough explanation of the data collection and data analysis processes as well as to provide justification for the methodology chosen’.

Transparency is important in any research, but its absence represents a particular oversight in analyses of the informal economy, given the inherent challenges in analysing the unobservable (see Feige, 2016a, 2016b; Schneider, 2016, whose debate in these articles we explore further below). Further, there is a ‘lack of a broadly accepted method of measurement [that] has hampered research efforts aimed at understanding shadow economies: their size, their determinants, their relations to the stages of economic development, and their responsiveness to various policy measures’ (Putnins & Sauka, 2015, p. 472).

Given these concerns, we reflect on the ways in which data on the informal economy can be collected. The literature identifies two broad approaches: *directly* from those engaged in informal activities; and *indirectly* via official statistics on variables that represent proxies for informal economic activity. We consider each in turn, before reflecting in more depth on the limitations of each.

3.1 | The informal economy and direct data collection

Surveys are a common way of collecting data for research on the informal economy. There is, however, a distinction to be drawn between surveys for the collection of ‘official statistics’ by ‘official’ bodies and their researchers, and surveys created and distributed directly by (academic) researchers in the context of individual research projects. Specifically, data collected in the field by academics will, or should, be designed in the context of a specific research project and research questions. This allows for close-to-optimum alignment of data collected and research needs. In particular, this will apply to in-depth small-n (often single country or even single-city) studies. We also suggest that, Open Data policies notwithstanding, most primary data collected in this manner will rarely be used again, other than indirectly via the subsequent citing of the resulting publications. This contrasts with the use of secondary data collected as official statistics. These will be utilised repeatedly, with different researchers using data in different ways for different studies. In this context, the research needs to adapt to the data collected, rather than the other way around. It facilitates, for example, large-n multi-country studies, where the depth of analysis is different—but again, this returns us to the question of alignment of data and analysis, rather than the intrinsic superiority, or not, of specific methods of data collection. We discuss below the challenges around survey-based data collection, but these must also be weighed against the data needs of specific research projects.

The OECD (2002) offers a useful starting point for official statistics. They identify several options available for data collection and analysis, including labour force surveys, household income and expenditure surveys, informal sector enterprise surveys, informal sector mixed household-enterprise surveys—to which questions or modules on the informal economy can be added (see also, *inter alia*, Benjamin et al., 2014; ILO, 2013; IMF, 2021). They also identify ‘independent surveys’, but they are unclear as to whether or not this describes surveys undertaken by organisations behind official statistics. The OECD and other sources cited above also provide caveats with each type of survey design, to which we return below.

An alternative to collecting official statistics via surveys is for researchers themselves to design and distribute surveys. This allows them potentially to collect exactly the data they wish to, consistent with other aspects of their research design such as their specific research questions and chosen theory. They also offer a potentially more precise and fine-grained targeting of the questions and participants than might be available via secondary data, depending on the detailed research design. On the other hand, this method can raise concerns over reliability, robustness, and representativeness, related to the survey’s design, distribution and respondent attitudes to completing the survey—although transparency can help to address these potential concerns with such research. With consideration of such surveys omitted from the official sources above, we reflect further on them below.

3.2 | The informal economy and indirect data collection

This second category of data derives from the idea that although the informal economy is difficult to measure, its activities leave traces behind that can be analysed through the use of proxy data. These traces can be found in the discrepancy between national income and expenditure, the discrepancy between the official and actual labour force, via transactions carried out in the economy, currency demand in the economy, or electricity consumption (Schneider & Enste, 2000).

Looking first at discrepancy-based methods, a key principle behind National Income Accounting (NIA) methods is that calculations based on income, output, and expenditure should all produce the same final estimate. In the informal economy literature, different national income estimates from the different methods are attributed to informal activities, 'if an independent estimate of the expenditure side of the accounts is available' (Schneider & Enste, 2000, p. 92). The use of labour force statistics, assumes that any decline in the measured formal labour force indicates a movement of labour into the informal labour force, 'if total labor force participation is assumed to be constant' (Schneider & Enste, 2000, p. 93). The transaction method (Feige, 1979), based on Fisher's equation for the quantity theory of money, $M \times V = P \times T$, assumes 'that the relation over time of the volume of transactions and official GNP is constant. [Feige] uses the value of total transactions (PT) as an estimate of nominal GNP; then, the informal economy is the difference between nominal GNP and the official GNP' (Restrepo-Echavarria, 2014, p. 468).

The currency (cash) demand⁴ approach dates back to Cagan (1958), with further developments and applications by, inter alia, Gutmann (1977), Tanzi (1980, 1983), and Dell'Anno and Halicioglu (2010). If the tax burden sends individuals into the informal economy, and assuming the informal economy is cash-based, then an increase in the tax burden will lead to an increase in the demand for cash. There is then an econometric estimation of the demand for cash, with models typically including data on tax rates, wages and salaries (as an indicator of money holding), and interest rates (as the opportunity cost of holding cash)—see, inter alia, Schneider and Enste (2000), OECD (2002), and Restrepo-Echavarria (2014).

The approach based on the physical inputs for economic activity—specifically electricity consumption (Kaufmann & Kaliberda, 1996)—assumes that with an electricity/gross domestic product (GDP) elasticity of close to unity (an assumption with empirical support; Restrepo-Echavarria, 2014, p. 469), informal economic activity is reflected in the difference between official GDP and total electricity consumed.

These methods are grouped together as 'indirect' data collection, to enable adjustments to be made to official data. That said, the IMF (2021) distinguishes the labour input method from the others—the former classified as an 'official statistics' estimation method, the others being 'non-official'. Similarly, the OECD (2002, chapter 5) classifies the labour input method as a national-account, production approach, but it classifies (in chapter 12) the monetary and electricity consumption approaches as macro-model methods. Thus even official statistics adopt indirect data collection methods to estimate what the OECD refers to as the non-observed economy, as a supplement to the directly collected data collected routinely for their databases (notably GDP). It is also worth noting at this point that every single one of the foregoing approaches in this section relies on major assumptions.

An alternative macro-model method is the multiple indicators multiple causes, or MIMIC, method, a popular example of the latent variable approach. Rather than focusing on a single key variable, such as money demand or electricity consumption, latent variable approaches start from the position that there is a non-observed dependent variable, the informal economy, but multiple observed independent or explanatory variables. The MIMIC model has been used extensively to estimate the size of countries' informal economies (see, perhaps most notably, Schneider et al., 2010). That said, MIMIC models have been criticised for significant potential weaknesses and limitations. This found perhaps its ultimate expression in the 2016 exchanges between Edgar Feige and Friedrich Schneider (in, chronologically, Feige, 2016a, 2016b; Schneider, 2016; plus summaries in IMF, 2015, 2021). This debate shows how controversial the estimates from MIMIC models are considered to be by some.

Feige makes a general critique about the lack of ‘transparency’ associated with most methods, but he is particularly vitriolic about Schneider’s MIMIC estimates, which ‘suffer from conceptual flaws, apparent manipulation of results and insufficient documentation for replication’, given which he questions ‘their place in the academic, policy and popular literature’. (Feige, 2016a, p. 5)—a statement he later repeats in his rejoinder (Feige, 2016b).⁵ In response, Schneider (2016) argues that Feige’s informal economy estimates relied heavily on OECD data, and that results for 27 countries were obtained from personal interviews and correspondences, which cannot be replicated.

Given the challenges of analysing something that is difficult to measure, we merely note at this point the observation of Albrecht von Haller, that ‘we are all wrong, but everyone is wrong differently’—something accepted by Schneider (2016), who noted drawbacks with all methods, and in particular an estimation error of about $\pm 15\%$. Medina and Schneider (2018, 2021) introduced ‘light intensity’ into the mimic model in attempt to address one of the critiques, but the drawbacks associated with electricity consumption method also hold here. We return below to MIMIC models and aspects of the Feige–Schneider debate, in the context of a fuller reflection on the limitations of *all* methods.

4 | ANALYSING INFORMAL ECONOMIC ACTIVITY: A CRITICAL REFLECTION

The foregoing introduces the principal approaches to obtaining data for economic analyses of the informal economy: direct versus indirect; official versus non-official; single-indicator versus multiple-indicator. In this section, we reflect on the criticisms, concerns, and shortcomings of these different approaches to data-gathering, starting with indirect methods and recalling that our primary interest is their application in developing countries. Before that, however, we revisit the theories introduced earlier and reflect further on each in the context of specific data requirements.

Table 1 summarises the main features of the key theories. What is missing from these theories is any reference to the composition of activities within the informal economy, and the specific data needs of each theory—which will in turn *also* be influenced by the composition of informal activities in the economy or economies under investigation. There is an implicit hierarchy in the understanding of data shaping much economics research and the methods of analysis that are then chosen. Thus secondary data, collected ‘directly’ by official processes, is perceived as having greater robustness than, say, data collected directly by non-official processes (by which we mean here data collected by academic researchers from surveys they have designed and distributed themselves).

The IMF (2021) refers to official bodies collecting data, with advice given to each of ‘International Organizations’ and ‘National Statistical Agencies’ as to how to enhance the data they collect on the informal economy. That said, the number of developing countries utilising this method of data collection remains limited. Further, the authors’ personal experiences have found that even when developing countries have collected official data, they are often either reluctant to share such data with academics, or simply refuse all requests. Thus direct data collection methods based on researchers designing and disseminating survey instruments themselves is not just a relevant source of data: it might be the only viable source.

4.1 | Analyses based on indirect data collection

We now offer a critical reflection on the limitations of the different data collection methods available for research on the informal economy. Dealing first with those approaches that utilise proxies for informal economic activity, electricity demand is highly problematic. First, and most directly, not all informal economic activity needs electricity. The single largest informal activity globally is street hawking, much of which will be mobile and most of which will have no need for electricity. Further, many developing countries suffer from sporadic formal electricity supplies, and it may not even be available in some parts of the country. Individuals (including those in the formal economy) might generate off-grid electricity (witness the shops that sell diesel generators). Recent advances in wind turbine and solar panel technologies are facilitating more and more local community energy schemes, providing power without grid coupling being needed. In short, official electricity consumption data may no longer capture either formal or informal activities.

The currency demand approach is also problematic (inter alia, OECD, 2002; Restrepo-Echavarria, 2014; Schneider & Enste, 2000). Notably, the assumption that informal activities are cash-based is increasingly questionable. Growing mobile phone ownership and the development of app-based banking services, even in remoter parts of the poorest countries, is reducing the use of cash—a shift encouraged by governments as a result of the COVID-19 pandemic (IMF, 2021, p. 12; see Phiri et al., 2023, for an analysis of branchless banking innovations). Casual observation on a recent visit to Lagos found a growing number of informal traders offering electronic payment options. Even beggars are not left out. When asking for money, if someone says they do not have cash, they offer their bank details to transfer tips into. Reflecting on institutional incongruence and the significance of informal institutions in some societies, there may well be transactions that are not even paid for, let alone paid in cash. In short, the informal economy now thrives with(out) cash and, as such, any estimate of the informal economy that is based on the currency approach is increasingly likely to be flawed.

These two cases illustrate a wider development that requires further research. Technological advances in recent years have begun to challenge fundamentally some of the key assumptions underpinning these methods of analysing the informal economy. At the very least, new research is required to determine the extent of the decoupling of formal electricity consumption from total economic activity, let alone informal activities specifically; and to analyse the impact of growing cashless transactions—again in the entire economy, let alone in undertaking informal transactions.

The foregoing approaches to modelling the informal economy also fail to reflect key features of informality. As noted earlier, neither the theories discussed earlier, nor these approaches to data collection, address different structural compositions of informality. For example, electricity consumption will vary between locations depending on the shares of activities such as small-scale manufacturing and street hawking. Second, individual motivations for economic activity are not only 'economic' (as seen for example in post-structuralist theory); but even where they are economic, it can be a sweeping assumption to assign a single driver to them all.

Turning to discrepancy-based methods, the transaction method is data-intensive. 'The problem is that [Feige] has to assume a base year where the assumption is that there is no informal economy. Also, the assumption that the ratio of transactions to official GNP is constant over time is quite strong. Additionally, obtaining accurate estimates of the total number of transactions is difficult' (Restrepo-Echavarria, 2014, p. 468; see also Schneider & Enste, 2000, pp. 93–94). Feige's application was to the United States, so one might plausibly assume these

challenges apply *a fortiori* in developing countries, with weaker institutions of governance and data collection capacity.

Moreover, whilst one alternative to estimating money velocity is reported by the OECD (2002, p. 189), this was based on using cash withdrawals and banknote-use instead, which returns us to criticisms of cash-based estimates of the informal economy. That said, this is one instance where technological developments, for example regarding branchless banking technologies, might provide more, and more robust, data in future, even in developing countries. This therefore also warrants further investigation in terms of possible future estimates of informal economic activity, as uses of branchless banking technologies expand.

The discrepancy approach based on labour force statistics, assuming constant labour force participation, is questionable as it ignores aggregate (formal plus informal) labour market entry and exit. This might be less significant in countries where social safety-nets are weak or absent, creating incentives to engage in economic activity for necessity and survival reasons. This approach to modelling informality also fails to capture multiple job-holding. This applies most directly where a person has jobs in both the formal and informal economies, but even if a multiple job-holder takes on or gives up jobs just within the formal economy, interpreting those as representing movements into or out of the informal economy is misleading. It also undermines the definition and empirical observations of opportunity-seeking entrepreneurs, some of whom start a business in the informal economy whilst maintaining full-time formal employment. In sum, more research is needed to determine the significance of labour-market entry and exit in different contexts, before discrepancies in labour market participation could be taken as a credible indicator of informal economic activity.

The third discrepancy approach, based on NIA calculations, raises one fundamental practical issue. Standard NIA approaches incorporate GDP balancing across the income, output and expenditure estimations. Attributing such discrepancies to, for example, the presence of informal economic activity and that alone is, therefore, highly questionable, when all countries routinely encounter inconsistencies across different NIA methods. Moreover, this discrepancy can only be used as the basis for modelling the informal economy if total expenditure can be calculated independently as the comparator.

In sum, even this short review highlights the presence of significant limitations and challenges when undertaking research based on any of these methods. That said, it also highlights areas for future research, in particular in relation to modelling approaches where new technologies are overturning old assumptions, such as those concerning electricity consumption or the demand for cash.

4.2 | Analyses based on direct data collection

In the literature on direct data collection, the focus is on the use of surveys as part of the collection of official statistics. Moreover, 'Conducting surveys is the main activity of the statistical office' (OECD, 2002, p. 106). What is striking, however, is the extent to which this discussion is wrapped in caveats about potential downsides and advice on how to limit them. We reflect first on some of the specific comments about different types of survey, before considering what this means for the study of the informal economy. We then link these discussions back to the theories outlined earlier, and how the data demands underpinning different theories align with direct methods of data collection.

As noted above, the IMF (2021) only discusses ‘official’ surveys as a means of data collection. The OECD (2002) mentions ‘independent surveys’, but they do not explain what they mean by this term. That said the implication (on pages 172–173) is that they are describing ‘official’ surveys, but where the survey is stand-alone, rather than having questions pertaining to the informal economy added into other existing surveys; labour force, household income and expenditure, enterprise surveys, and so forth.

Looking first at labour force surveys (OECD, 2002, especially pages 170–171; ILO, 2013), in the first instance definitions are required as to what constitutes informal economy employment; and its relation to the formal economy, such as in the context of multiple job-holding (which has already been highlighted for its challenges to data-collection). There is also the issue of survey duration, with informal employment varying over the year. There is thus no agreement over the frequency with which such surveys should be conducted and the timeframe for which information should be gathered (OECD, 2002, pp. 170–171).

With household income and expenditure surveys, ‘the source data are *generally inadequate* in terms of coverage, scope, and timeliness’ (IMF, 2021, p. 25, emphasis added). Further, whilst information can be gathered on ‘household final consumption expenditure for informal sector products’, this is only part (albeit an important part) of ‘total demand for informal sector products’ (OECD, 2002, p. 171).

Survey data on enterprises have been exploited by economists for some time and, more recently, by management and entrepreneurship scholars (Galdino et al., 2018, p. 234), but this raises questions about defining and identifying enterprises. First, ‘An enterprise survey presupposes the availability of a sampling frame of informal sector enterprises or establishments ... [and] even where a business register exists, it does not usually cover informal sector enterprises’ (OECD, 2002, p. 171; see also Benjamin et al., 2014, p. 11; IMF, 2021, p. 25). Thus, even the availability of business registers is not assured to deliver adequate data.

Investment Climate Assessment (ICA) surveys define such enterprises by size: ‘usually firms with fewer than five employees’ (Benjamin et al., 2014, p. 11). Without significant effort, informal activities undertaken in people’s homes or on the move with no fixed location are simply ‘likely to be omitted’ (OECD, 2002, p. 171). Even going to locations where informal economic activity is known to occur, the best Benjamin et al. (2014, p. 11, emphasis added) can offer is that ‘the firms surveyed are *quite probably* informal [... but ...] there remain issues of representativeness of the overall informal sector’. This is particularly significant given that the most numerous activity in the global informal economy is street hawking—precisely the sort of ‘ambulant trade’ that the OECD (2002: 171) fears will simply be omitted from survey data-gathering—but which has been the specific focus of surveys designed and conducted by academics, including the present authors.

To illustrate these issues in the specific context of Nigeria, only two formal firm surveys have been carried out (2007 and 2014), which immediately raises a crucial issue with the use of enterprise surveys, especially in developing countries. Recalling the importance of street hawking in informal economic activity globally, whilst these surveys include ‘retail’ in the disaggregation of services data, the smallest units included are those with 5–19 employees. This categorisation appears to omit street hawkers. Moreover, and to emphasise the significance of this omission, one recent survey (PwC, 2020) found that enterprises smaller than five employees constitute 73% of all micro, small, and medium enterprises in Nigeria (in direct contrast to the ICA surveys as reported by Benjamin et al., 2014).

Galdino et al. (2018, p. 234, emphasis added) advise researchers that they ‘should also rely on other sources that provide *comprehensive and accurate data* on informal activities and the

informal economy. The World Bank Enterprise Surveys, for example ... [tackle] *some questions* related to informality'. As noted, however, even international organisations identify multiple serious concerns about survey data, whilst the breadth of the questions on informality in the World Bank Enterprise Surveys is limited. It therefore seems a stretch to suggest that World Bank Enterprise Surveys provide the 'comprehensive and accurate' data that Galdino et al. (*op cit*) call for. That said, the foregoing suggests *all* official surveys possess weaknesses and limitations.

A different approach is discussed, in particular, by Benjamin et al. (2014, p. 12)—surveying households to obtain information on enterprises. Even so, they admit that 'employees are unlikely to reliably report firm details'; and that it is 'unlikely that these surveys can offer sufficient coverage of large firms', a point inconsistent with the ICA surveys, where informal firms are defined by their *small* size. The OECD (2002, p. 172), via what they refer to as 'Informal Sector Mixed Household-Enterprise Surveys', suggest that the survey may be completed by household members 'other than the enterprise owners themselves', thus 'it is not normally possible to obtain good quality data relating to the informal sector criteria. Instead, the focus is on ensuring good coverage of the informal sector by attempting to identify the owners of all enterprises that *may* belong to the informal sector' (emphasis in original). Another challenge with household surveys based on home visits is created by the strong links between informal economic activity and informal economy workers living in informal settlements (*inter alia*, Grant, 2015, p. 135), where a systematic approach to sampling will likely be more complex. This includes the basic challenge of even finding these workers at home to survey.

4.3 | Researcher-led direct data collection and research credibility

In all of these sources, there is one category of survey that is not discussed—the survey that is designed and distributed by (academic) researchers themselves. What is striking about all of the cited sources above, but especially the reports from the OECD (2002), ILO (2013), and IMF (2021), is the detail they provide on how to conduct robust surveys. This is not surprising. It does, however, raise a concern over data from any survey ever being, to quote Galdino et al. (2018, p. 234) once again, 'comprehensive and accurate'.

It is also worth reflecting on some of the other observations made in the cited literature regarding survey-based data collection. We are told that 'the results are going to depend directly on the questions asked by the survey' (Restrepo-Echavarria, 2014, p. 468); that 'results from these surveys are sensitive to the way the questionnaire is formulated, and as with all surveys, precision and results depend greatly on the respondents' willingness to cooperate' (Schneider & Enste, 2000, pp. 91–92); bearing in mind also that 'many informal sector entrepreneurs have a relatively low level of education and do not keep usable written records of their activities. They are not used to participating in surveys and are often unwilling or unable to devote much time to doing so' (OECD, 2002, p. 175); and that 'people are not usually willing to admit that they are not reporting taxes or that they are having a fraudulent behavior, either because they feel afraid of getting caught or because they feel ashamed since they know this is a moral issue' (Restrepo-Echavarria, 2014, p. 468).

The key point from the foregoing is that direct methods of data collection are demanding in terms of their design. There is more than a suggestion in this that whilst the collection of data that are comprehensive and accurate can be an aspiration, its attainment would be challenging even for well-resourced public organisations in the Global North with pools of highly skilled

staff. But if we return to the theories of the informal economy, we see more clearly still the limitations of many of the aforementioned methods of data collection (Table 2).

Three key features emerge from Table 2. First, it is possible to collect data both directly and indirectly to explore the different issues highlighted by each theory. Second, within the 'direct' and 'indirect' categories, different specific techniques of data collection can be utilised, to capture informality along at least one dimension (with regard to labour markets, electricity, consumption, etc.). Third, and crucially for what follows, there is an expectation from the ways in which the theories have been formalised that their application in practice will involve an exploration of individual motivations.

Individual motivations are, therefore, central to understanding the informal economy, but finding out about them is not straightforward. They can be determined indirectly, for example by analysing changing tax rates over time in a country, different tax rates in different countries at any point in time, or a panel combining both dimensions, to seek to identify whether any statistical link can be found with the decision to participate in the informal economy. But, as noted earlier, this also assumes that, in this example, tax rates are the only, or primary, motivation for individuals to engage in informal activity. By extension, cash-based methods then rely on the assumption of a predictable relationship between tax rates, participation in the informal economy and the demand for cash (recalling also the point noted earlier, about new technologies affecting what might have been assumed or determined previously about the demand for cash). In short, this approach to analysing the informal economy does not capture individuals' actual motivations, but an *estimate* of *assumed* motivations.

Turning to direct survey methods, the approaches to, and issues with, official surveys have been summarised above. In principle, they offer access to data collected from a large number of individuals, firms, and so forth. Moreover, if the research undertaken and research questions asked by researchers align with the specific details of a given survey, they represent an excellent source of information. If, for example, a study focuses on key choices made by enterprise owners, then well-aligned informal economy-related questions incorporated into an enterprise study will ensure alignment between research questions and the nature of the data

TABLE 2 Analytical data needs of theories of the informal economy.

Theory	Key data implications
Dualist	<ul style="list-style-type: none"> Analyse the size of the IE, to determine diminution over time Analyse skills gaps, education gaps, etc. Analyse constraining factors within the FE
Structuralist	<ul style="list-style-type: none"> Analyse motivations of FE actors towards IE Analyse FE policies that might lead FE actors to exploit IE Therefore a need to analyse individual motivations
Legalist	<ul style="list-style-type: none"> Analyse motives of actors to move from the FE to the IE Analyse FE policies that might lead FE actors to move to the IE Therefore a need to analyse individual motivations
Voluntarist/ neoliberal/ post-structuralist	<ul style="list-style-type: none"> Analyse motives of actors to move from the FE to the IE Analyse how FE markets can be liberalised to match the attractiveness of the IE Therefore a need to analyse individual motivations

Abbreviations: FE, formal economy; IE, informal economy.

collected, from which appropriate methods of data analysis can then be chosen. The same principle applies to all survey types noted earlier.

Reflecting further on some of the caveats noted previously, all cross-country studies will be subject to qualifications over the consistency of questions, sampling, and other issues fundamental to data collection in multiple countries. Given the challenges of conducting robust surveys, the example given earlier, of Nigeria's enterprise surveys, highlights the issue of their infrequency. From this, we can suggest that research that focuses specifically on one aspect of informality (e.g., enterprises, household income and expenditure) can appropriately utilise official survey-based data, *ceteris paribus*. Research that focuses on cross-country analysis, focusing on a single year chosen for data availability and addressing one of these aspects, can appropriately utilise official survey-based data, *ceteris paribus*. Research that focuses on developed countries can reasonably be expected, *a priori*, to have more data to access than developing countries. That is a lot of Latin *caveats*.

This discussion is not intended to dismiss or denigrate research undertaken using one of these data-collection methods. Rather, our intention is simply to highlight the questions that arise when analysing the informal economy utilising even the most robust-looking data sources. Given the likely shortfall from any attempt to analyse something that is difficult to measure below the benchmark of 'comprehensive and accurate data', at the very least there needs to be explicit recognition of such data-related issues in any research. This should be provided not only explicitly, but also reflected in the research design, to ensure full alignment between research questions, data, methods/sources of data collection, and methods of data analysis.

The flip-side of these issues, however, is research where those *ceteris paribus* assumptions do not hold. This is more likely the more limited are the official surveys, the more infrequent are the official surveys, and the more limited are the resources to undertake regular, extensive, and robust official surveys. These challenges will also be compounded, the more extensive and diverse is the informal economy. In such countries, the scale of informality will likely, *ceteris paribus*, involve a greater range of motivations for informal economic activity, more fluidity between the formal and informal, where the informal is the 'everyday' and the norm, and where capturing the scale of informality will be a greater challenge; however, extensive are the official surveys used that try to capture it. In other words, official surveys will offer a more limited picture of informality in developing countries, a problem compounded by the greater challenges faced by developing countries, as noted above, around conducting official surveys.

This brings us to the one method of data collection that the international organisations cited throughout this review do not engage with—surveys designed and conducted by academic researchers. These are, first and foremost, subject to the same caveats as official surveys, with the added observation that research teams will typically lack the financial resources (and probably the staffing levels) available to those bodies conducting official surveys. This might limit the number of responses obtained, although it cannot be assumed that they lack the expertise of the official bodies. Research based on such data collection methods must therefore be designed with certain key principles in mind. It cannot seek to answer research questions based on cross-country analyses. It can only seek to answer research questions based on time-series analysis for very carefully designed research, where the researchers have had the resources to undertake multiple surveys, possibly for a short run of years, possibly for a series of non-consecutive years (for which the methods of data analysis must also be chosen carefully, to match these features of the data collected).

Such methods are therefore most suited to research where the research questions address the breadth of concerns captured by the theoretical model(s) chosen; where official informal economy survey data are not available or are limited for the purposes of the research; where weaknesses or limitations might be manifest in official data that could be utilised for indirect, especially discrepancy-based, analyses; and where the limitations around currency and electricity consumption approaches created, for example, by recent technological advances, have yet to be addressed. In short, researcher-led surveys will be most appropriate for in-depth analyses of informality in single developing countries.

There is still a potential trade-off between analyses that focus on a country in its entirety (that can also incorporate a comparative element, between different regions) and analyses that focus on one city or one region. This trade-off will be greater, the more limited are the time and financial resources to undertake the data collection, and the larger the focal area. Decisions must be made at the research design stage as to how best to address and accommodate these trade-offs, to ensure consistency with the specified research questions.

What the informal economy literature has also shown is that the data debate is incomplete without an in-depth consideration of data analysis, with many studies utilising MIMIC models, the earlier debates and criticisms notwithstanding; recognising also that all approaches to analysing the difficult-to-measure concept of informality will, as discussed, have limitations. As such, the first duty of researchers is to be open and transparent as to what they have done, how, and why. This will enable a shift in focus from the sort of debate cited earlier, between Feige and Schneider, to considerations of whether research is well-designed and justified on its own terms.

As for why the MIMIC method remains popular, Table 2 highlights that most theories of the informal economy encompass the potential for multiple drivers. In developing countries in particular, as discussed earlier, the larger and more diverse the informal economy, the more likely there are to be multiple, interacting, drivers to participation in the informal economy. Further, it is in developing countries where official data limitations are otherwise likely to be at their greatest. As a result researcher-led surveys, combined with MIMIC methods, offer an approach that enables multiple drivers of informality to be explored in unison, *absent* adequate official data.

From this, two uses for MIMIC models can be identified. The first, as seen for example in Schneider et al. (2010, p. 14), utilises multiple series of official statistics. MIMIC models can also be run using researcher-led survey data on multiple variables, to enable identification and analysis of the motives for participation in the informal economy (inter alia, Igudia et al., 2016).⁶ Such approaches are subject to the same criticisms as official surveys. Thus another area for further research is whether individuals are more likely to be honest about, for example, their income to an official (from whom they might wish to hide such information) or an independent researcher (to whom they might not be willing to give as much of their time to reflect on their answers).

On this point, we note the recent experience of the current authors in conducting a survey in Lagos, Nigeria. As part of a research project funded with a modest budget, there were sufficient resources to enable the recruitment, training and support of a team of undergraduate students to disseminate the survey, across all 20 Local Government Areas of Lagos. Further, resources were included to reimburse the participants (including street hawkers and motorcycle taxi riders) for their time.⁷ As a result, in a short period of time we obtained over 1700 completed responses.

5 | CONCLUDING THOUGHTS: WHERE DOES THIS LEAVE STUDIES OF THE INFORMAL ECONOMY?

There is a very famous quote, cited by Sir Josiah Stamp in ‘Some Economic Factors in Modern Life’, from an anonymous source:

The government are very keen on amassing statistics. They collect them, add them, raise them to the *n*th power, take the cube root and prepare wonderful diagrams. But you must never forget that every one of these figures comes in the first instance from the village watchman, who just puts down what he damn well pleases.

It might be exaggeration for effect, but one aim of the foregoing has been to set out just how many questions there are over the data available with which to analyse the informal economy. Against the benchmark of comprehensive and accurate data, any theory that focuses on specific drivers, and any proposed method of data collection outlined earlier that focuses on one indicator of informality, will fall short of that benchmark. The notion of comprehensive data collection depends on the definition of population to be sampled. As for accuracy, the demanding standards set by the OECD (2002, pp. 172–174) for the design of what they call ‘independent surveys’ would be a challenge for any survey. Their design:

entails fairly complex survey operations and sample design and estimation procedures. It requires a team of qualified survey staff, sound training of interviewers, constant supervision and control of all survey operations, and care in keeping records of the listing operation, sample selection and sample outcome for each sample area.

Schneider and Enste (2000, p. 105) argue that ‘one should be very careful when interpreting the size of the shadow economy in a country using only one method’. They go on (on pages 105–106) to compare estimates of the informal economy in Canada, Germany, Great Britain, Italy, and the United States, from a range of studies, calculated using the different methodologies and methods outlined above for different time periods. The lowest lower-bound estimates came from surveys (an informal economy estimate of 1.5%–4.5% of GNP). Monetary approaches produced estimates of 15%–35% of GNP (transactions approach), 10%–30% of GNP (Gutmann method), and 4%–20% of GNP (currency demand approach). Lying within these estimates, calculations based on electricity as a physical input average 12.7%, whilst modelling approaches produce an overall average of 7.9%. Discrepancy methods produce higher than average estimates for some countries, lower than average for others.

These are estimates for five developed countries, for which data can be expected to be much more reliable and consistent than developing countries, where the informal economy will be larger and more important, economically and socially. The advice to seek to analyse informal economies using more than one method is sound, but it is not clear what can be done when estimates vary as much as those above—even if the resources were available to undertake multiple parallel analyses. That said, when ‘agencies may refrain from publishing estimates of the informal economy, given the absence of a well-established statistical framework with specific guidance for compiling data on informal production’ (IMF, 2021, p. 24), we learn nothing.

Galdino et al. (2018, p. 243), in their survey of 102 studies on the informal economy, found that ‘Most of the studies in this review that use surveys as the main method of data collection

do not provide detailed information that can attest to the quality of the data and consequently the results obtained'. Given that surveys continue to be a popular and important method of data collection, and given the likelihood that most academic studies will not be able to undertake multiple diverse analyses, we return to the argument that journal editors and their reviewers need to recognise and accept; first, that comprehensive and accurate data are unachievable benchmarks, that all data collection methods in studies of the informal economy have limitations, that there needs to be recognition of the implications of this for philosophical and methodological approaches to economics research, and that transparency around all of these questions is the *sine qua non* to enable an evaluation of each study on its own terms.

It is important in this review also to recognise some examples of studies where the research does reflect one or other of the key issues described above. To give just a few examples, consider the following research from Colin Williams and co-authors, where the titles conveniently include reference to 'evaluating competing theories' (Williams et al., 2012; Williams & Kedir, 2018; Williams & Windebank, 2015). In these three papers, clear information is provided about data sources—respectively, researcher-led face-to-face interviews, and two analyses using European Union or World Bank data. In all three papers, the theories being tested are outlined, hypotheses developed on the basis of each, details provided on the additional data collected, discussion of how each dataset links to the underlying theory being tested, and where the data are obtained from (although the source(s) of data for the control variables in Williams & Kedir, 2018, are not clear).

Crucially, the research questions being answered in these studies, especially with the two studies based on pre-existing databases, are specific to the data available and therefore are consistent with the analyses that can reasonably be conducted with the data. These three papers, therefore, demonstrate clarity and coherence regarding research design and choices around data, sources and methods of data collection, and methods of data analysis, all linked to the theories of the informal economy under investigation.

Returning to our initial research question, we have set out in detail the data challenges facing those wishing to study the informal economy. The data available for something that is difficult to measure are replete with caveats. In many cases, the narrowness of the data to be used contrasts with the breadth of many developing countries' informal economies. As reported above, well-designed studies are available, but they are unusual at least insofar as the authors are transparent about the key features of their research design, to enable this judgement to be made. Many studies do not match this. Areas for future work that we identify from this are, first, to reflect on how our analysis of the data demands of economic theories designed specifically to understand the informal economy might apply to other theories, such as entrepreneurship, that are general but imported into studies of the informal economy. Second, within Economics there are different philosophical underpinnings to research, between orthodox and heterodox, that might also influence choices around data, and data collection and data analysis. It would be interesting to explore how these differences affect such data choices.

Ultimately, it is essential that we continue to analyse and understand better the informal economy, given its importance to billions of people worldwide. In the absence of a single unifying theory, in this review, we have sought to argue that all methods of data collection are flawed, but the need to continue studying the informal economy means we must accept the current (flawed) data until better alternatives are found. Whatever data are employed, they must satisfy the theoretical underpinnings and be aligned with the design of the research, and we must be uncompromising on transparency. We have also identified several areas for which further research is desperately needed. This includes recognising that some of the longstanding

economic theories of the informal economy are increasingly being overtaken by technological developments. These developments also, however, offer new opportunities for data collection that can enhance the clarity and accuracy of analyses of the informal economy. We have much to do.

ACKNOWLEDGEMENTS

The authors are extremely grateful to the reviewers for their helpful comments directed at improving the paper; and to the editors for their guidance, support and encouragement. The usual disclaimer applies.

CONFLICT OF INTEREST STATEMENT

We report no potential conflicts of interest.

DATA AVAILABILITY STATEMENT

Data sharing not applicable—no new data generated. The article utilises purely conceptual research.

ORCID

Rob Ackrill  <https://orcid.org/0000-0002-0739-1812>

Eghosa Igudia  <https://orcid.org/0000-0002-2566-3199>

ENDNOTES

- ¹ <https://www.ilo.org/employment/units/emp-invest/informal-economy/lang--en/index.htm> (last accessed 19 April 2023).
- ² <https://www.imf.org/en/News/Articles/2021/07/28/na-072821-five-things-to-know-about-the-informal-economy#:~:text=The%20informal%20economy%20is%20a%20global%20phenomenon%2C%20but%20there%20is,15%20percent%20in%20advanced%20economies> (last accessed 19 April 2023).
- ³ A comprehensive review of such theories is beyond the scope of this article. For a recent review, see Dell'Anno (2022).
- ⁴ The OECD (2002: 189–190) also identifies a method based on the ratio of cash to deposits. For brevity, we focus just on the more widely applied method of modelling informality using currency demand.
- ⁵ In late July 2023, data on Google Scholar indicate well over 1000 citations for Schneider et al. (2010). From Feige's perspective, that is a lot of research using flawed data.
- ⁶ In the unpublished doctoral research underpinning this paper, estimates were made of the size of Nigeria's informal economy, over several years, using both the MIMIC method (utilising secondary data), and the currency approach. Comparing the results over a 16-year period, the MIMIC results were more variable, but there was no consistency in terms of one set of estimates being higher than the other. Further, the average of each came to 52% (currency method) and 52.6% (MIMIC model). Estimates for multiple years therefore provide richer information and, arguably, a more reliable basis for analysing and interpreting the practical meaning of such estimates. This is strengthened further by the utilisation of two separate estimation methods, following which such comparisons as these become an important part of the analysis. It is also, arguably, less important to wonder if Nigeria's informal economy is 52%, 52.6%, 50%, or 55% of GDP. What matters for policy is that the informal economy is at least as large as the formal economy.
- ⁷ A small sum consistent with the obligations of research ethics and integrity.

REFERENCES

Becker, K. F. (2004). *The Informal Economy*. Sida Fact Finding Study. <https://www.sida.se/en/publications/the-informal-economy>

- Benjamin, N., Beegle, K., Recanatini, F., & Santini, M. (2014). *Informal economy and the World Bank*. World Bank policy research working paper WPS6888. The World Bank.
- Bruton, G. D., Ireland, R. D., & Ketchen, D. J., Jr. (2012). Toward a research agenda on the informal economy. *Academy of Management Perspectives*, 26, 1–11.
- Cagan, P. (1958). The demand for currency relative to the total money supply. *Journal of Political Economy*, 66, 302–328.
- Castells, M., & Portes, A. (1989). World underneath: The origins, dynamics, and effects of the informal economy. In A. Portes, M. Castells, & L. Benton (Eds.), *The informal economy: Studies in advanced and less advanced developed countries* (pp. 11–40). John Hopkins.
- Chen, M. (2012). *The informal economy: Definitions, theories and policies*. WIEGO Working paper 1, <http://www.wiego.org>
- Chen, M., Jhabvala, R., & Lund, F. (2002). *Supporting workers in the informal economy: A policy framework*. Working paper on the informal economy 2002/2. International Labour Office.
- De Soto, H. (1989). *The other path: The economic answer to terrorism*. Harper and Row.
- Dell'Anno, R. (2008). What is the relationship between unofficial and official economy? An analysis in Latin American countries. *European Journal of Economics*, 12, 185–203.
- Dell'Anno, R. (2022). Theories and definitions of the informal economy: A survey. *Journal of Economic Surveys*, 36, 1610–1643.
- Dell'Anno, R., & Halicioglu, F. (2010). An ARDL model of unrecorded and recorded economies in Turkey. *Journal of Economic Studies*, 37, 627–646.
- Fairlie, R., & Fossen, F. (2018). Opportunity versus necessity entrepreneurship: Two components of business creation, IZA discussion paper, DP No. 11258, Bonn, Germany.
- Feige, E. (1979). How big is the irregular economy? *Challenge*, 12, 5–13.
- Feige, E. L. (2016a). Reflections on the meaning and measurement of unobserved economies: What do we really know about the “shadow economy”? *Journal of Tax Administration*, 2, 5–41.
- Feige, E. L. (2016b). Professor Schneider's shadow economy (SSE): What do we really know? A rejoinder. *Journal of Tax Administration*, 2, 93–107.
- Feld, L. P., & Schneider, F. (2010). Survey on the shadow economy and undeclared earnings in OECD countries. *German Economic Review*, 11, 109–149.
- Galdino, K. M., Kiggundu, M. N., Jones, C. D., & Ro, S. (2018). The informal economy in pan-Africa: Review of the literature, themes, questions, and directions for management research. *Africa Journal of Management*, 4, 225–258.
- Grant, R. (2015). *Africa: Geographies of change*. Oxford University Press.
- Gutmann, P. (1977). The subterranean economy. *Financial Analysts Journal*, 33, 26–27.
- Hart, K. (1973). Informal income opportunities and urban employment in Ghana. *Journal of Modern African Studies*, 11, 61–89.
- Henry, S. (1978). *The hidden economy*. Martin Robertson.
- Hyden, G., & Williams, D. C. (1994). A community model of African politics: Illustrations from Nigeria and Tanzania. *Comparative Studies in Society and History*, 36, 68–96.
- Igudia, E., Ackrill, R., Coleman, S., & Dobson, C. (2016). Determinants of the informal economy of an emerging economy: A multiple indicator, multiple causes (MIMIC) approach. *International Journal of Entrepreneurship and Small Business*, 28, 154–177.
- Igudia, E., Ackrill, R., & Machokoto, M. (2022). Institutional incongruence, the everyday, and the persistence of street vending in Lagos: A demand-side perspective. *Environment and Planning A*, 54, 1256–1276.
- ILO. (1972). *Employment, incomes and equality: A strategy for increasing productive employment in Kenya*. International Labour Office.
- ILO. (2013). *Measuring informality: A statistical manual on the informal sector and informal employment*. International Labour Office.
- ILO. (2018). *Women and men in the informal economy: A statistical picture*. International Labour Office. https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/documents/publication/wcms_626831.pdf
- IMF. (2015). *Current challenges in revenue mobilization: Improving tax compliance*. International Monetary Fund staff report. IMF. <https://www.imf.org/en/Publications/Policy-Papers/Issues/2016/12/31/Current-Challenges-in-Revenue-Mobilization-Improving-Tax-Compliance-PP4944>

- IMF. (2017). *Sub-Saharan Africa: Restarting the growth engine*. World Economic and Financial Surveys Regional Economic Outlook, April. IMF. <https://www.imf.org/en/News/Articles/2017/05/05/na050917-sub-saharan-africa-restarting-the-growth-engine#:~:text=Policy%20priorities%20to%20restart%20growth%20engine&text=%E2%80%9CSub%2DSaharan%20Africa%20remains%20a,this%20potential%2C%E2%80%9D%20Selassie%20said>
- IMF. (2021). *Measuring the informal economy*. IMF policy paper. International Monetary Fund.
- Kaufmann, D., & Kaliberda, A. (1996). Integrating the unofficial economy into the dynamics of post-socialist economies: A framework of analysis and evidence. In B. Kaminski (Ed.), *Economic transition in the newly independent states* (pp. 81–120). Sharpe.
- Ketchen, D., Ireland, D., & Webb, J. (2014). Toward a research agenda for the informal economy: A survey of the strategic entrepreneurship journal's editorial board. *Strategic Enterprise Journal*, 8, 95–100.
- Medina, L., & Schneider, F. (2018). Shadow economies around the world: What did we learn over the last 20 years? IMF working paper WP/18/17. <https://www.imf.org/en/Publications/WP/Issues/2018/01/25/Shadow-Economies-Around-the-World-What-Did-We-Learn-Over-the-Last-20-Years-45583>
- Medina, L., & Schneider, F. (2021). The evolution of shadow economies through the 21st century. In C. Delechât & L. Medina (Eds.), *The global informal workforce: Priorities for inclusive growth* (pp. 10–69). International Monetary Fund.
- Moser, N. (1978). Informal sector or petty commodity production? Dualism or independence in urban development? *World Development*, 6, 1041–1064.
- Neuwirth, R. (2013). Life and livelihood, on the street. *SAIS Review of International Affairs*, 33, 65–71.
- OECD. (2002). *Measuring the non-observed economy: A handbook*. Organisation for Economic Cooperation and Development.
- Phiri, Y., Ackrill, R., Amissah, E., & Makhoulouf, Y. (2023). Determinants of remote and physical branchless banking innovations: Evidence from Malawi. Unpublished Mimeo, Department of Economics, Nottingham Business School, Nottingham Trent University.
- Pratt, A. (2019). Formality as exception. *Urban Studies*, 56, 612–615.
- Putnins, T. L., & Sauka, A. (2015). Measuring the shadow economy using company managers. *Journal of Comparative Economics*, 43, 471–490.
- PwC. (2020). *PwC's MSME survey 2020: Building to last*. Nigeria Report. Pricewaterhouse Coopers. <https://www.pwc.com/ng/en/assets/pdf/pwc-msme-survey-2020-final.pdf>
- Restrepo-Echavarría, P. (2014). Macroeconomic volatility: The role of the informal economy. *European Economic Review*, 70, 454–469.
- Roberts, A. (2014). Peripheral accumulation in the world economy: A cross-national analysis of the informal economy. *International Journal of Comparative Sociology*, 54, 420–444.
- Schneider, F. (2016). Comment on Feige's paper "Reflections on the meaning and measurement of unobserved economies: What do we really know about the 'shadow economy'?". *Journal of Tax Administration*, 2, 82–92.
- Schneider, F., Buehn, A., & Montenegro, C. (2010). *Shadow economies all over the world: New estimates for 162 countries from 1999 to 2007*. World Bank policy research paper 5356. The World Bank.
- Schneider, F., & Enste, D. (2000). Shadow economies around the world: Sizes, causes, and consequences. *Journal of Economic Literature*, 38, 77–114.
- Sindzingre, A. (2006). The relevance of the concepts of formality and informality: A theoretical appraisal. In B. Guha-Khasnobis, R. Kanbur, & E. Ostrom (Eds.), *Linking the formal and informal economy: Concepts and policies* (pp. 58–74). UNU-WIDER studies in development economics). Oxford University Press.
- Tanzi, V. (1980). The underground economy in the United States: Estimates and implications. *Banca Nazionale del Lavoro*, 135, 427–453.
- Tanzi, V. (1983). The underground economy in the United States: Annual estimates, 1930–1980. *IMF Staff Papers*, 30, 283–305.
- Webb, J. W., Bruton, G. D., Tihanyi, L., & Ireland, R. D. (2013). Research in entrepreneurship in the informal economy: Framing a research agenda. *Journal of Business Venturing*, 28, 598–614.
- Webb, J. W., Tihanyi, L., Ireland, R. D., & Sirmon, D. G. (2009). You say illegal, I say legitimate: Entrepreneurship in the informal economy. *Academy of Management Review*, 34, 492–510.
- Williams, C., & Nadin, S. (2010). Entrepreneurship and the informal economy: An overview. *Journal of Development Entrepreneurship*, 15, 361–378.

- Williams, C. C. (2023). *A modern guide to the informal economy*. Edward Elgar.
- Williams, C. C., & Kedir, A. (2018). Evaluating competing theories of informal sector entrepreneurship; a study of the determinants of cross-country variations in enterprises starting-up unregistered. *The International Journal of Entrepreneurship and Innovation*, *19*, 155–165.
- Williams, C. C., Nadin, S., & Rodgers, P. (2012). Evaluating competing theories of informal entrepreneurship: Some lessons from Ukraine. *International Journal of Entrepreneurial Behaviour & Research*, *18*, 528–543.
- Williams, C. C., & Windebank, J. (2015). Evaluating competing theories of informal employment: Some lessons from a 28-nation European survey. *International Journal of Business and Globalisation*, *15*, 45–62.

How to cite this article: Ackrill, R., & Igudia, E. (2023). Analysing the informal economy: Data challenges, research design, and research transparency. *Review of Development Economics*, 1–23. <https://doi.org/10.1111/rode.13070>