KNRajMNF (DEPR): 01

Prof. K.N. Raj Memorial National Fellowship Scheme

Informality: Causes, Consequences and Policy Responses

Dr. Ravi Kanbur
Professor of Economics
Cornell University
Ithaca, NY, USA

DEPARTMENT OF ECONOMIC AND POLICY RESEARCH
JUNE 2014

The Reserve Bank of India (RBI) initiated the Prof. K.N. Raj Memorial National Fellowship Scheme in 2012. The Fellowship is awarded to a distinguished scholar of social sciences from India or abroad to pursue research for a short period in an academic institution in India. The Studies under the Scheme present the research work of the scholar during the period of the fellowship. The views expressed in the Study are those of the author and do not represent the views of the RBI.

Informality: Causes, Consequences and Policy Responses

Ravi Kanbur[*]
www.kanbur.dyson.cornell.edu

June 2014

Contents

[*] Cornell University: T.H. Lee Professor of World Affairs, International Professor of Applied Economics and Management, and Professor of Economics. Work on this paper was conducted while I was the inaugural Reserve Bank of India K.N. Raj National Fellow during 2013-2014.

1. Introduction

A stylized prediction of the development economics discourse is that informality will disappear with development. Much of our analytical and policy discussion on informality is framed by an expectation of its imminent demise. And yet in the last twenty years conventional measures of informality, far from declining, have either remained stagnant or have actually increased. This includes countries like India where economic growth has been at historically high levels. These trends led the OECD to ask in a recent publication: "Is Informal Normal?"[1] Their answer was in the affirmative.

What exactly is informality and what are its magnitudes and trends? What are the causes of informality and why is it not decreasing as predicted by standard theories of development? What are the consequences for inclusive economic growth of a large and increasing informal sector? What are feasible and desirable policy responses to informality? These are the questions which motivate this broad based overview of informality. The questions will be addressed based on recent and ongoing research on India and globally.

The paper comprises six sections including this introduction. Section 2 begins the analysis by considering the concept of informality, or rather the alternative conceptualizations of informality that abound in the literature. This section will also summarize the latest empirical work on levels and trends in informality globally, including in India, paying particular attention to data and measurement difficulties. It will establish the basic contention that informality, no matter how measured, is not on the decrease worldwide.

Why does informality exist? And why is it increasing? Section 3 develops a formal model to provide a framework in which to examine a range of hypotheses, from regulatory burden to technological change, focusing in particular on trying to understand why informality is increasing when development theory predicts that it should decline with growth and development. Section 4 uses the framework of the model of section 3 as an entry point to the debate on whether informality is bad for inclusive economic growth. The consequences of informality for inclusion will also be assessed in this section, and it will be argued that research shows strong association between informality and poverty, especially of women.

Given the central role for policy towards informality in an inclusive growth strategy, what are the specifics of such a strategy? Section 5 considers a range of policy interventions, including deregulation and addressing the low productivity of small scale production. It will discuss in particular credit based interventions, for example encouraging or requiring formal sector banks to engage with informal enterprises, or supporting small scale financial institutions in the informal sector itself.

Finally, the concluding section 6 draws together the threads and summarizes the main findings and policy messages.

[1] Jutting and de Laiglesia (2009).

2. Measurement

The discourse on "informality" and "formality" is an old one in development economics, and is closely linked to notions of "dualism" or "dualistic development." One cannot get very far in the study of development without encountering these terms. And yet, as Guha-Khasnobis, Kanbur and Ostrom (2006) observe:

"Given the prominence of the formal-informal dichotomy in the development discourse, one might expect to see a clear definition of the concepts, consistently applied across the whole range of theoretical, empirical and policy analyses. We find no such thing. Instead, it turns out that formal and informal are better thought of as metaphors that conjure up a mental picture of whatever the user has in mind at that particular time." (pp 2-3).

Let us start with a brief historical excursion into thinking on formality and informality, beginning with discussions and debates in the context of dualism. It is commonly acknowledged that the term dual economy was coined by the Dutch colonial administrator and academic Julius Herman Boeke in his writings on the Dutch East Indies as they were then, Indonesia as it is now (Boeke, 1949). The dualism he wrote of was that between the "imported high capitalism" and the "native economy," the argument being that modes of economic and social organization were very different between the two, and the former came into the ambit of formal colonial law and regulation, while the latter did not.

For development economists the idea of dualism was crystallized by Arthur Lewis's (1954) Nobel Prize winning articulation of a two-sector model of development in which one sector is "modern/capitalist" ("industrial", "urban","formal") and the other is "traditional" ("agricultural", "rural", "informal"). The key difference is that enterprises in one sector maximize profits while in the other division of output is through traditional norms because in this latter case the marginal product of labor is zero ("surplus labor"). However, as the modern sector grows through investment, more and more labor is sucked into this sector from the traditional sector—the natural trajectory is thus for the traditional (informal) sector to fall in size relative to the modern (formal) sector.

Another notion of dualism, one that is related more to Boeke than to Lewis, is captured in the work of Keith Hart, the anthropologist who is credited with having coined the term "informal" in contradistinction to "formal" in his study of a slum area in Accra, the capital of Ghana in West Africa (Hart, 1973). Here is how he describes his insight in a later retrospective (Hart, 2006):

"Following Weber, I argued that the ability to stabilize economic activity within a bureaucratic form made returns more calculable and regular for the workers as well as their bosses. That stability was in turn guaranteed by the state's laws, which only extended so far into the depths of Ghana's economy. "Formal" incomes came from regulated economic activities, and "informal" incomes, both legal and illegal, lay beyond the scope of regulation." (p. 25)

It should be striking how similar this conceptualization is to that of the distinction between the regulated realm of colonial activities and the unregulated realm of "native" enterprise in Boeke (1949). This distinction between a regulated and an unregulated sector also played a key role in the paper of

Harris and Todaro (1970), which helped structure much of the analytical and policy discourse in development economics in the 1970s and 1980s, and which shows its influence even today. In this model the economy is divided into a sector which has a minimum wage regulation, the formal sector, and another where there is no regulation so the wage is determined in a competitive labor market. Here the size of the informal sector is closely related to the intensity of the regulation since it is the regulation which creates excess supply of labor in the formal sector which is then absorbed in the informal sector. Thus in this model informality will decline with deregulation.

The central idea in the above contributions is that "formality" is to do with an activity coming under the purview of the state, in the form of coming under the ambit of a law or a regulation, while informality is that which is outside this domain. As Guha-Khasnobis, Kanbur and Ostrom (2006) argue, this is the conceptualization that stands out as a common strand in a mass of literature that attempts different definitions, based on size of enterprise, degree of competition, coincidence of ownership and control, etc. However, there are still a number of steps to go before the concept can be operationalized for statistical measurement using national data sources.

An important issue is whether the perspective taken is that of the enterprise or that of the worker. India's National Commission on Employment in the Unorganized Sector (NCEUS) reflects the debate in distinguishing between the informal sector (consisting of enterprises) and informal employment (consisting of workers), drawing on the guidelines of the International Convention of Labour Statisticians (ICLS, 2003) but applying them to India's specificities:

"*Informal Sector*: '…..all unincorporated private enterprises owned by individuals or households engaged in the sale and production of goods and services operated on a proprietary or partnership basis and with less than ten total workers'.
Informal worker/employment: '….those working in the [informal] sector or households, excluding regular workers with social security benefits provided by the employers and [including] the workers in the formal sector without any employment and social security benefits provided by the employers'.
Informal economy: The informal sector and its workers plus the informal workers in the formal sector constitute the informal economy." (NCEUS, 2009, p. 3)

With these definitions, the NCEUS found that in 2004-05, out of a total employment of 455.7 million workers in the Indian economy, 393.3 million were in the informal sector. Further, of the 62.6 million workers in the formal sector, 28.9 million were informal workers according to the above definition. Thus the informal sector employed 86.3% of all workers in India in 2004-05, the figure being 92.3% for the informal economy. The NCEUS also showed that these patterns had hardly budged since 1999-2000, the figures for that year being 86.1% for the informal sector and 91.5% for the informal economy. If anything, then, there has been a slight increase in informality during this period of high growth in the Indian economy. These broad trends in the Indian economy are confirmed by the independent work of

Ghani, Kerr and O'Connell (2013), who find that the employment share of the organized sector in Indian manufacturing has remained at around 81% between 1989 and 2005.[2]

When one moves across countries the institutional and legal framework changes and the specific statistical definition of formality and informality can also change, even when they derive from a common conceptual framework. Thus, for example, although in India size of enterprise is a key aspect of defining informality, for Mexico the institutional legal structures are somewhat different. As Levy (2008) clarifies, Articles 20 and 21 of the Federal Labor Law define the relationship between an employer and an employee (what is known as "subordinated work" in exchange for a wage, *la relacion obrero-partonal* in Spanish). Articles 12 and 13 of the Social Security Law then specify the obligations of the employer towards the employee as defined in the Labor Law. Notice that there is no reference to size of establishment. Formality is defined by Levy (2008) as being those workers who are registered for social security—he estimates that in 2006 these workers constituted 38% of the work force. Thus formality as measured is much higher than in India, but still below 50%.

Country specific studies can of course use country specific definitions of formality and informality. But cross-country comparisons of level and trends need a more uniform statistical approach, as illustrated by the recent ILO-WIEGO (2013) publication. A common feature of cross-country compilations is that agriculture is excluded, among the reasons being that "usual data collection systems do not often distinguish formal and informal (or modern and traditional) agriculture." (Charmes, 2009, p. 32). Despite comparability and other issues, such cross-country compilations give us a handle on the global picture interms of levels and trends:

"…on average, informal employment accounts for more than 47 per cent of total non-agricultural employment in West Asia and in North Africa, and more than 70 percent in sub-Saharan Africa, more than 50 per cent in Latin America, nearly 70 per cent in South and Southeast Asia and 24 per cent in transition economies….Albeit not uniform, the data show an upward-oriented trend of informal employment in all regions." (Charmes, 2009, p. 32)

The conclusion on trends needs to be treated with caution not only because of country specific variations but because different studies can produce different results as new and more recent data become available. For example, the ILO-WIEGO (2013) study presents trends for 13 countries over a five to ten year period. While trends for Mexico (not much change in informality rates) and India (informality rates of around 85% up to 2011/12) are consistent with other studies, it finds declines in informality in several Latin American countries, and in South Africa. However, we can safely say that the uniform and significant reduction predicted by development theory has not materialized, especially in a country like India.

[2] There now appears to be broad consensus on these trends in Indian informality, in academic and in official circles—see for example Economic Survey of India 2013 (Government of India, 2013), Box 2.5.

3. Causes

If formality and informality are to do with relationship of economic activity to state regulation and laws, then the causes of informality must be sought in the nature of these regulations relative to the structure of economic activity, and the evolution of both in relation to each other. It will be helpful to set up a simple conceptual schema. Imagine first of all a world without any laws or regulation. There will then be a "natural" pattern of economic activity. For example, there will be a size distribution of enterprises, from the very small to the very large, which is the outcome of economic forces untrammeled by state intervention. Now let the state introduce a regulation affecting economic activity. Conceptually this could be any sort of regulation, but to fix ideas suppose the regulation states that all enterprises employing more than a certain number of workers have to register with the authorities and have to in turn provide certain benefits to their workers. Such "size-dependent regulation" is very common around the world. For example, India's Factories Act (1948) requires such registration of all enterprises in manufacturing who use electricity and who employ 10 or more workers (20 or more workers if they use electricity).

The basic ideas can be developed with a simple model.[3] Let output $y$be given by

$$y = al - \left(\frac{1}{2}\right) bl^2$$

where l is labor and$a$ and $b$are production parameters. We will focus particularly on a as a productivity parameter. If the wage is w then profit is

$$\pi = al - \left(\frac{1}{2}\right) bl^2 - wl$$

The unconstrained profit maximizing choice of l and the maximized profitπ are given respectively by:

$$l = (a - w)/b$$

$$\pi = (a - w)^2/2b$$

Thus for the optimal unconstrained firm, size and profit increases with productivity a.

Now suppose that there is a regulation which bites at $l > \hat{l}$, and the effect of the regulation is to introduce a fixed cost c. This regulation is clearly irrelevant for $a \leq \hat{a}_1 = b + w\hat{l}$. For those with $a > \hat{a}_1$ the choice is between complying, not complying, or adjusting out of the regulation by keeping employment at $\hat{l}$. If the firm complies,employment and profit are given by is given by:

$$l_A = (a - w)/b$$

[3] A more general formulation is presented in Chatterjee and Kanbur (2013); a related model focusing on taxation and informality is presented in Kanbur and Keen (2014). There are now many such models, of different degrees of sophistication—for example, de Paula and Scheinkman (2007), Gourio and Roys (2012) and Garicano, Le Large and Van Reenen (2013). But all models recognize their debt to the initial formulation by Lucas (1978).

$\pi_A = [(a-w)^2/2b] - c$

Suppose that if the firm does not comply there is a probability p of getting caught and, if caught, there is a fine of f per worker employed. Then employment and profits are given by:

$l_B = (a-w-pf)/b$

$\pi_B = [(a-w-pf)^2/2b] - pc$

If instead a firm chooses to stay at an employment level of $\hat{l}$ its employment and profit is:

$l_C = \hat{l}$

$\pi_C = a\hat{l} - \left(\frac{1}{2}\right) b\hat{l}^2 - w\hat{l}$

Finally, returning to those with $a \leq \hat{a}_1 = b + w\hat{l}$, since they are outside theregulatory net their employment and profits are given by the unconstrained values for this range:

$l_D = (a-w)/b$

$\pi_D = (a-w)^2/2b$

A firm with $a > \hat{a}_1$ compares profit in the three regimes to decide which regime to be in—comply, evade, or avoid. It can then be shown that there exist values of a, $\hat{a}_1 < \hat{a}_2 < \hat{a}_3$, such that[4]

$$\hat{a}_1 < a < \hat{a}_2 \Rightarrow \pi_C > \pi_B > \pi_A$$

$$\hat{a}_2 < a < \hat{a}_3 \Rightarrow \pi_B > \pi_C > \pi_A$$

$$\hat{a}_3 < a \Rightarrow \pi_A > \pi_B > \pi_C$$

To these regimes is to be added the category of those who do not come under the regulation at all—in other words, those with $a \leq \hat{a}_1 = b + w\hat{l}$. These might be termed the outsiders, with profits denoted π_D . The values of $\hat{a}_1 < \hat{a}_2 < \hat{a}_3$ are defined by the following equations:

$$\hat{l} = (\hat{a}_1 - b)/w$$

$$\pi_C(\hat{a}_2) = \pi_B(\hat{a}_2)$$

$\pi_B(\hat{a}_3) = \pi_A(\hat{a}_3)$

The introduction of regulation thus creates four categories of enterprises:

[4] This will hold for p and f small enough.

A. Those who are covered by the regulation and comply. For example, in the case of India's Factories Act, enterprises with 10 or more workers which register.
B. Those who are covered by the regulation but do not comply. This is, quite simply, illegality—an enterprise with 10 workers or more which does not register.
C. Those who adjust out of the coverage of the regulation. This would be an enterprise whose "natural" size would be 12 workers, say, but which chooses to stay at 9 workers to avoid registration costs. These enterprises have reacted perfectly legally to the regulation.
D. Those who are outside the coverage of the regulation. These would be enterprises whose "natural" size would be less than 10 workers.

The above categorization can be used to address a number of issues in the causal factors behind informality. If formal is defined as being covered by regulation and complying, category A is formal. Then if informality is the complement of formality, B + C + D is informal. Indeed this is how informality is often measured from national statistical sources. But it is immediately clear that informality is itself composed of different sub-categories, with very different economic causes (and consequences). Category B would not exist if enforcement of the regulation were perfect. Its size depends on the nature of the enforcement regime. Category C is one which worries many economists as representing the distortionary costs of the regulation—in the example of size-dependent regulation, these enterprises would be larger and would employ more workers in the absence of regulation. Category D is outside the regulatory net altogether, but is the special worry of many civil society organizations, which see low productivity and low incomes in these very small scale enterprises as a major barrier to inclusive growth.

Before proceeding to a detailed discussion of the causes of informality, let us put some empirical flesh on this theoretical skeleton. How big are the different categories A, B, C, D? We can answer this question for the case of India's Factories Act, the law we have been using as an example throughout the theoretical development. We draw on the work of Chatterjee and Kanbur (2013). India's Factories Act (1948) applies to manufacturing and requires registration of all enterprises with 10 or more workers if they use electricity and 20 or more workers if they do not. Chatterjee and Kanbur (2013) argue that the second requirement has become increasingly irrelevant since there are fewer and fewer establishments that employ 10 workers or more but do not use electricity. We follow them in focusing on the "10 or more workers" criterion (they show that the patterns are not much affected using the broader criterion).

Registered enterprises comprise the "organized sector" in official terminology, or the formal sector in terms of the development economics discourse. This is our category A above—those who come within the ambit of the state and comply with the defining regulation. Thus all unregistered enterprises would constitute informality, B+C+D. But what about categories B, C and D taken separately? How do we get an estimate of the number of enterprises and the employment in each of these categories? Registered enterprises are surveyed every year under the Annual Survey of Industries (ASI). But every five years the National Sample Survey Organisation (NSSO) conducts a survey precisely of those enterprises which are not registered under the Factories Act. It collects a large amount of information on these enterprises, including employment.