The Agricultural Productivity Gap: Informality Matters

Rajveer Jat*
rjat001@ucr.edu

and

Bharat Ramaswami#
bharat.ramaswami@ashoka.edu.in

September 2024

Abstract

The literature has debated whether the productivity gap between agriculture and non-agriculture reflects mobility barriers or selection. Non-agriculture is not a homogenous category. In developing countries, most of non-agricultural employment is informal. Could it be that the productivity gap is driven by formal sector firms that are numerically small but economically substantial? This paper compares the productivity of agriculture to the informal and formal non-farm sectors in India. The comparison controls for sectoral differences in hours worked, human capital and labor share of value added. The paper finds substantial productivity gaps with the formal sector but small and negligible gaps with the informal non-farm sector. Between 40-50% of non-farm workers are in sectors not more productive than agriculture. These findings suggest that the primary dualism in development is between the formal non-farm sector and the informal sector including agriculture.

Keywords: Agricultural productivity gap, Informal sector, India

JEL Code: O11, O13, O41

*University of California, Riverside, USA
Ashoka University, India

The Agricultural Productivity Gap: Informality Matters

1. Introduction

A robust stylized fact about the process of development is that the share of agriculture in employment is greater than the share of the sector in income. The gap between agriculture's employment and income share means that a worker in the agriculture sector is less productive than her counterpart in the non-agricultural sector. Calculations based on national income and product accounts suggests that, across countries, productivity in the non-agricultural sector is, on average, three times higher than the productivity in the agricultural sector (Gollin, Lagakos, and Waugh, 2014). In the current literature, this gap in productivity has been called the agricultural productivity gap (APG).[1] For the poorest quartile of countries, the productivity gap rises nearly to six compared to about two for the richest quartile of countries. The literature has debated whether the APG reflects mobility barriers or whether it is because of self-selection into the high productivity sectors or whether it is due to measurement errors.

The discussion on the APG has, however, for the most part, ignored substantial heterogeneity in the non-farm sector. This is the point of departure for this paper. Developing country non-farm sectors are typically characterized by a large number of small firms with an employer and few or no employees. However, large firms do exist and worker productivity is higher in large firms and, therefore, share of large firms in income is higher than their share in employment (Ciani et.al, 2020, OECD, 2014). Such heterogeneity prompts the question whether the observed agriculture productivity gap is driven by the larger firms in the nonfarm sector that are numerically small but economically substantial.[2]

[1] In older work, the differences in agricultural productivity across countries was called the agricultural productivity gap (Hayami, 1969).

[2] Agriculture might be heterogenous as well. In particular, there might be productivity differences between staple crops and cash crops (Rivera-Padilla, 2020). This feature is not considered in this paper.

In this paper, we pursue the implications of such heterogeneity in the non-farm sector for the agricultural productivity gap in India by utilizing the distinction between formal and informal segments. The International Labour Office (ILO) defines the informal economy to consist of unincorporated household enterprises that are not registered with the government (for taxes or social security) or do not keep accounts. Informal employment includes all employers of informal enterprises plus all workers that have an informal relation with the employer (Bonnet, Van and Chen, 2019).[3] Since not every country collects relevant data, proxy data (such as small size of enterprise) are often used to measure the informal economy. For 2016, the ILO estimated that informal employment accounted for 73% of non-agricultural employment in low-income countries, 59% of non-agricultural employment in middle-income countries and 17% of non-agricultural employment in high-income countries (Bonnet, Van and Chen, 2019). From World Bank Surveys, La Port and Shleifer (2008, 2014) estimated that informal firms may account for 35% of GDP in low-income countries. They also report large productivity differences between formal and informal firms.

In the Indian context, informal enterprises accounted for 43% of non-farm GDP in 2017 (Murthy, 2019).[4] In the same year 68% of all non-farm employment was informal (Nagaraj and Kapoor, 2022, Murthy, 2019). This is strikingly similar to the disparity between agriculture's share of employment and its share of GDP. At first glance, it would, therefore, seem that the APG could depend on whether the farm sector is compared with the formal non-farm segment or with the informal non-farm segment. These comparisons therefore merit a nuanced investigation of the productivity gap.

[3] Examples of these are lack of social security contributions by the employer and the lack of benefits such as annual leave and paid sick leave.

[4] The share of the informal sector in non-farm GDP is derived from Murthy's estimates. According to those estimates, the informal sector accounted for 52.4% of all GDP in 2017/18. Agriculture contributes 17% of GDP almost all of which is informal (97% of agricultural GDP). Hence the non-farm informal sector is 36% of GDP and therefore 43% of non-farm GDP.

In this paper, we probe these disparities further. Following the methods of Gollin, Lagakos and Waugh (2014), (abbreviated to GLW, henceforth) we adjust the gap in value added per worker for sectoral differences in human capital and in hours worked. However, unlike them, we also adjust for differences in the labour share of value added. We use a disaggregation of the non-farm sector into 24 sub-sectors to identify (primarily) formal and (primarily) informal segments and to estimate their productivity relative to agriculture. In a second approach, we use the same data to non-parametrically estimate, by sub-sector, the relation between informality and the agricultural productivity gap. Finally, we supplement these approaches by a comparison of sectoral wage gaps adjusted for differences in human capital and hours worked. While this has the virtue of being a direct measure of productivity gaps, we offer it here as a robustness check rather than as a principal result. Identification of wage gaps rely on individual level panel data that captures migration across sectors (e.g., Herrendorf and Schoellman, 2018, Alvarez, 2020, Hamory et.al, 2020). Such data is not available for India. But even if it were, its coverage would be incomplete as much of the workforce in the informal sector is self-employed and do not report wage data. Indeed, La Porta and Shleifer (2008) use the percentage of the non-agricultural labor force that is self-employed as an indicator of informality. [5]

A preview of our results is that the productivity gap between the farm sector and the informal segment of the non-farm sector is low or negligible. However, there is a sizeable productivity gap between the formal and informal segments of the non-farm sector and also between the formal segment of the non-farm sector and the farm sector. The non-parametric analysis shows that the results depend on the extent of informality. When informal workers account for more than 83% of a non-farm sub-sector, we are unable to reject the null hypothesis

[5] In the Indian context, the self-employed accounted for 43% of male employment and 51% of female employment in the informal economy in 2004 (India. National Commission for Enterprises in the Unorganized Sector, 2008).

of a zero productivity gap. Such non-farm activities account for 40-50% of non-farm employment. The implication is that the APG debate may benefit from focussing on the gap between the formal and informal segments rather than only on the dualism between agriculture and non-agriculture. Like GLW, our paper is agnostic about the source of this productivity gap – whether selection or whether mobility barriers. While our finding is specific to India, it may have wider applicability because of the substantial presence of the informal segment in many low income countries (Bonnet, Van and Chen, 2019). Since small relatively unproductive unincorporated enterprises are characteristic of the typical developing country (La Porta and Shleifer, 2014), the findings here suggest that similar results may obtain for other countries too.

In the next section, we place this paper and its methods in the context of previous literature on the subject. In section 3, the two sector agricultural productivity gap (GLW, 2014) is extended to a heterogenous non-farm sector with informal and formal components. Data sources are discussed in section 4. Section 5 describes the procedures for identifying the (primarily) formal and (primarily) informal segments of the non-farm sector. It also presents estimates of the agricultural productivity gap for the formal and informal segments. Section 6 examines the robustness of our findings in several ways. First, it shows that the estimates are not sensitive to the assumptions used to identify formal and informal segments. Second, we conduct a similar analysis of wage gaps between agriculture and the informal and formal segments. Third, we present an altogether different approach where we move away from the idea of identifying informal and formal segments. Instead, we consider agriculture's productivity gap with each of the 24 sub-sectors that constitute the non-farm sector. Each of these sub-sectors contain, to a varying degree, an informal component. Hence, we estimate a non-parametric relation between the agricultural productivity gap and the extent of informality.

We formally test the null hypothesis that the productivity gap is zero and invariant to the extent of informality. We gather our conclusions in Section 7.

2. Relation to Literature

Previous work has argued that low agricultural productivity in the poor countries is one reason for aggregate productivity differences between rich and poor countries (Restuccia Yang and Zhu, 2008, Vollrath, 2009). McMillan and Rodrik (2011) emphasized that labor flows from low productivity to high productivity sectors could be an important way of increasing overall productivity.

Could the productivity gap be mostly due to measurement errors? In a major contribution, GLW re-measured the productivity gap for 151 countries after taking into account two salient features: lower hours of work in agriculture and lower levels of human capital in agriculture relative to other sectors.[6] They showed that these adjustments reduce the productivity gap but do not eliminate it - it is about two on average for the combined sample of rich and poor countries and is about three for the poorest quartile. On the face of it, such a large productivity gap between sectors is puzzling and is suggestive of frictions and barriers that lock too much labor in agriculture.

This finding has been challenged in, at least, couple of ways. Using micro time use data for four African countries, McCullough (2017) showed that productivity measures based on categorizing individuals by their primary sector of occupation (typically used in macro measures of productivity) overstates agricultural labor measured in hours of work. As a result, the per-hour productivity gaps are much smaller. In a similar vein, Fuglie et.al (2020) cite

[6] The paper also considered other factors such as under-estimation of agricultural home production, mismeasurement of agricultural work, and urban-rural differences in cost of living

micro evidence from China and India to argue that productivity gaps or wage gaps are close to zero.

Another strand of the literature does not deny productivity gaps but questions the interpretation that labor is mis-allocated across sectors. Explanations for the agricultural productivity gap (APG, henceforth) have been proposed in terms of self-selection of human capital into the high and low productivity sectors (Young, 2013, Herrendorf and Schoellman, 2018, Alvarez, 2020, Hamory et.al, 2020). This implies that there are no large gains from re-allocating labor from the farm to non-farm sector. Consistent with this view, these papers find only modest wage gains to those who switch occupation from one sector to another.

These findings are, however, not supported by the literature that finds large returns to migration across sectors (Beegle, De Weerdt and Dercon (2011), Imbert and Papp (2020)). This has been confirmed by experimental evidence that finds large returns to migration induced by modest incentives (Bryan, Chowdhury, and Mobarak, 2014). Recent surveys of this literature point to a middle ground and assess a role for both sorting and labor mobility frictions in accounting for the APG (Lagakos, 2020; Donovan and Schoellman, 2021).

Relative to this literature, our paper is closest to GLW (2014) in its objectives and methods. We adopt some of their principal methods to examine the productivity gap with the formal and informal non-farm segment. We extend their methods in two directions. We do away with their assumption that labor shares in value added are the same across sectors. This is particularly restrictive in our case as the formal segment comprising of large production units typically employ technologies different from the small units (including one-person firms) in the informal sector. Second, we utilize the disaggregation of the non-farm sector to estimate the relation between the APG and the extent of informality.

Alvarez (2020) examined the productivity gap between formal sector workers in agriculture and those in non-agriculture but not between informal and formal segment workers.

Herrendorf and Schoellman (2018) examine agriculture's wage gap with different sub-sectors of the nonfarm sector – industry and services and as well for two components of services – skilled services and unskilled services. Wage gaps are large for all non-farm sectors but are smallest for unskilled services – a sector that probably bears the greatest resemblance to the unorganized segment in our data. It should, however, be noted that workers in their unskilled sectors can have as many as 13 years of schooling. Typically, schooling accomplishments are much less in the informal segment (La Porta and Shleifer, 2014).

3. The Agricultural Productivity Gap with Informal and Formal Non-Farm Components

Table 1 displays the Indian APG defined as the ratio of value added worker in non-agriculture relative to agriculture in the two decades spanning the 1990s and 2000s. With Cobb-Douglas production functions (in labor and capital) and common factor shares across sectors, GLW show that this ratio ought to be unity whenever labour is mobile across sectors. However, the APG in India has ranged between 3 and 4 during this period – a value that is close to the global average of this variable (GLW, 2014).

Table 1 : Agricultural Productivity Gap (APG)

Estimate	1999-00	2004-05	2011-12
APG	3.91	4.32	4.03
Corrected APG	1.83	2.21	1.77

Notes: APG is the ratio of value added in non-agriculture to value added in agriculture, both measured in constant rupees. The data source is the India KLEMS project, https://www.rbi.org.in/Scripts/PublicationReportDetails.aspx?UrlPage=&ID=1158

As GLW point out, non-agricultural workers typically work more hours and are better educated. This applies to India as well. For instance, in India, during 2004-05, non-agricultural work hours were, on average, 56% higher than in agriculture. Similarly, human capital in non-agriculture was 21% higher than in agriculture (following the GLW methods to estimate human capital). We follow the methods of GLW to correct the APG for these sectoral differences in effective labor input. The second row of Table 1 displays the corrected APG. The corrected figure accounts for about half of the unadjusted gap. However, even the corrected APG is well above 1. While these results replicate the finding of GLW that APG survives corrections for errors in measuring effective labor, they do not take cognizance of the substantial informal segment in non-agriculture that is likely to be less productive than the formal segment.[7]

To address this, let us consider an economy with two sectors, agriculture (called the A sector) and a non-farm sector broken up into two segments: informal and a formal non-segment. The production function for agriculture is Cobb-Douglas and is

$$Y_a = B_a L_a^{\theta_a} K_a^{1-\theta_a}$$

where B_a is total factor productivity in agriculture, L_a and K_a are the labour and capital inputs, and θ_a is the labor share of agricultural value added.

The production functions in the non-farm sector are also Cobb-Douglas but the parameters are different across the informal and formal segments

$$Y_j = B_j \left(\alpha_{1j}\alpha_{2j}L_j\right)^{\theta_j} K_j^{1-\theta_j} \qquad j = i,f$$

where i and f subscript informal and formal sectors, B_j is total factor productivity in segment j, L_j and K_j are the labour and capital inputs, and θ_j is the labor share of non-farm value added in segment i. Workers in the non-farm sector may work longer hours and may have higher human capital. $\alpha_{1j}\alpha_{2j}L_j$ is the labor input in the informal sector measured in terms of the efficiency of agricultural labor. The efficiency parameter α_{1j} captures the sectoral differences in working hours while α_{2j} denotes the relative efficiency of non-farm labor because of greater human capital. If labor is freely mobile across sectors, all workers receive the same wage equal

[7] The formal/informal distinction is not meaningful for agriculture where formal enterprises are negligible.

to the value of marginal product in each of the sectors and the segments within them. Thus, marginal value products are equalized and we get the result

$$\frac{Y_j/L_j}{Ya/L_a}\,\frac{\theta_j}{\theta_a}\,\frac{1}{\alpha_{1j}\alpha_{2j}} = 1, \text{ for segments } j = i,f \qquad (1)$$

where j indexes the non-farm sector according to whether it is formal (*f*) or informal (*i*), and *a* subscripts the variables of the agricultural sector.

The first term on the left hand side of (1) is the ratio of value added per worker in the non-farm sector (formal or informal) to the value added per worker in agriculture. This is the `raw' agricultural productivity gap because it does not take into account the sectoral differences in the labor share of value added (second term) and the sectoral differences in effective labor input (third term). A productivity gap exists if the left hand side of (1) is larger than unity. GLW computed the left hand side of (1) for 151 countries. They adjust the raw APG for sectoral differences in effective labor input stemming from differences in work hours and human capital. However, they assumed the labor shares of value added to be equal across sectors in their analysis of 151 countries. This assumption holds up in the Indian case, when comparing agriculture to all of non-farm sector. In our case, however, it would be unwise to assume similarly. The formal segment consists of larger production units (by employment) and are typically associated with greater access to credit and greater use of capital. Hence we adjust the raw APG for effective labor input (like in GLW) and also for differences in labor share of value added.

A more direct measure of the productivity gap would be to look at the wages in the non-farm sector (formal and informal) relative to the farm sector. This ratio should be one if labor is fully mobile across sectors. To verify this, we do not need to assume a functional form for the production function and neither do we need to compute sectoral labor shares. However, as noted in the introduction, even if we have the ideal panel data to permit identification, wage data do not capture the productivity of the self-employed – a characteristic of the informal sector. Therefore, our approach is to estimate the left hand side of equation (1). For supporting evidence, we also look at sectoral wage gaps corrected for sectoral differences in human capital and hours worked.

4. Data Sources

For the macro data, we primarily rely on the India KLEMS database.[8] This give us a time series on value-added per worker for 26 sectors that include agriculture and 24 non-farm sectors[9]. However, in this data base, the non-farm sectors are not divided into formal and informal segments. To do that, we supplement this with employment data from nationally representative surveys of the National Sample Survey Organization (NSSO). As a result, even though the KLEMS data runs from 1990 to 2022, we are constrained by the availability of NSSO surveys and report results for three years: 1999-00, 2004-05, and 2011-12.

The macro data is adjusted for sectoral differences in human capital and hours worked. The former is drawn from the NSSO surveys while the latter is sourced from the nationally representative household-level Indian Human Development Surveys (IHDS)[10] on social and economic issues.[11]

[8] The India KLEMS project measures productivity at the industry level (27 industrial divisions). The information is reported on gross value added, total labor employed, labor quality index, and labor share in value-added. https://www.rbi.org.in/Scripts/PublicationReportDetails.aspx?UrlPage=&ID=1158

[9] The data set contains 26 non-farm sectors, but we use only 24 out of them. We drop two outlier sectors: "Public *Admin, Defense and Social Security* "and "*Petroleum and Nuclear* Fuel". The former has 82% labor share in value-added and the later has only 7% labor share in value added, these are two extremes. The inclusion of these two sectors does not qualitatively change our results, however their peculiarities lead us to exclude them from the analysis.

[10] Indian Health and Demographic Survey (IHDS) data is a detailed household-level micro-data. Link: https://ihds.umd.edu

[11]. Workers in the NSS dataset and the IHDS dataset are mapped to the sector of employment using the National Industry Classification (NIC) codes. We use the mapping to compute the sectoral labor inputs (human capital and hours worked) relative to agriculture.

The wage gaps are also estimated using NSSO employment surveys and the adjustments for differences in hours worked and human capital rely on the same sources as the macro data.

Table 2 summarizes the variables and their data sources.

Table 2: Data Sources

Value added per worker in Agriculture	KLEMS
Value added per worker in non-farm sub-sector	KLEMS
Labour share of value added in Agriculture	KLEMS
Labour share of value added in non-farm sub-sector	KLEMS
Percentage of employment that is informal, by non-farm sub-sector	NSS Employment Survey
Percentage of employment that is formal, by non-farm sub-sector	NSS Employment Survey
Human capital in agriculture and in non-farm sub-sector	NSS Employment Survey
Hours worked in agriculture and in non-farm sub-sector	Indian Human Development Survey
Average weekly wages in agriculture and in non-farm sector (formal/informal)	NSS Employment Survey

5. APG Estimates

Table 3 (next page) presents the proportion of sectoral employment that is informal (or formal). To compute the numbers in Table 3, we classify all employment into the industries described in the KLEMs data set.[12] Within each industry, an individual is classified as working in the formal sector if the enterprise of employment offers retirement benefits (provident fund) or if the enterprise has at least 10 employees. This follows from the definition of the formal sector by a government commission (Report on Conditions of Work and Promotion of Livelihoods in the Unorganized Sector, 2007).[13] We are then able to compute the proportion of total employment within each non-farm sector that is either formal or informal.

From the results in Table 3, we can see that there is no sector that is either fully formal or fully informal. Thus, there is a formal-informal continuum with sub-sectors employing varying proportions of informal labor. In this section, we consider the heterogeneity at the two ends of this continuum. Informal labor dominates in some sectors (e.g., Trade, Wooden Products and Hotels and Restaurants). Similarly, some sectors exist mainly as formal segments (e.g., Chemicals, Motor Vehicle Manufacturing; Electricity, gas and water supply). We begin by adopting a thumb rule that a sector is `primarily formal' if more than two-thirds (66%) of employment within it is formal. A similar definition applies to `primarily informal' sectors. By our procedure, 11 non-farm sub-sectors are identified as informal and 5 as formal.[14] The remainder are neither primarily formal nor primarily informal. What is the APG for sectors that are primarily formal and the sectors that are primarily informal?

[12] The data comes from the household level all-India employment-unemployment surveys of the National Sample Survey Organization (NSSO).

[13] https://dcmsme.gov.in/Condition_of_workers_sep_2007.pdf

[14] The industries identified as primarily informal are Trade, Wood and Cork, Hotel and Restaurants, Transport and Storage, Textile Leather and Footwear, Paper Print and Publishing, Construction, Food Beverages Tobacco, Electric Equipment Manufacturing, Metal Manufacturing, and Individual Professional Services. The industries qualifying as primarily formal are Motor Vehicle Manufacturing, Chemicals, Electricity Gas Water Supply, Rubber and Plastic Manufacturing, and Finance & Banking Services.

Table 3: Distribution of Employment in Non-farm sectors. (Source: NSS EUS).

	Informal Employment (%)			Formal Employment (%)		
Sector	**1999-00**	**2004-05**	**2011-12**	**1999-00**	**2004-05**	**2011-12**
Trade	95.0	93.1	96.3	4.1	3.1	3.7
Wooden Products and Cork	94.6	93.6	94.7	5.1	3.9	5.3
Other Individual Services	90.7	92.4	91.5	6.4	4.8	8.2
Hotel and Restaurants	89.5	87.5	88.5	10.1	10.7	11.3
Other Manufacturing\Recycling	86.6	82.6	85.1	13.3	16.1	14.8
Transport and Storage	77.7	80.9	83.6	21.9	17.1	16.3
Textile Leather and Footwear	76.0	78.5	79.4	23.7	19.4	20.6
Construction	84.0	78.7	77.2	15.2	19.4	22.5
Food Beverages and Tobacco	79.5	76.9	74.8	20.3	20.8	25.1
Mining and Quarrying	40.0	32.4	67.6	58.3	66.2	32.0
Electrical Equip. Manufacturing	32.3	49.8	66.1	67.4	48.5	33.8
Basic Metal Manufacturing	63.1	71.2	62.0	36.6	27.5	37.9
Business Services	80.0	74.5	60.0	19.1	23.2	39.8
Education	45.9	52.0	54.5	53.1	45.8	45.2
Pulp Paper Printing\Publishing	62.9	58.4	54.4	36.3	40.8	45.6
Health Social-work	52.9	57.0	53.8	45.9	41.0	46.0
Post and Telecommunication	52.7	62.2	52.2	46.7	37.7	47.8
Machinery Manufacturing	49.1	52.7	51.2	50.6	45.8	48.8
Finance and Banking Services	32.6	38.9	46.2	66.8	60.0	53.7
Non-Metallic Manuf.\Minerals	63.5	57.6	41.3	35.4	40.8	58.5
Rubber and Plastic Manufacturing	39.5	42.4	34.4	60.2	57.6	65.6
Motor Vehicle Manufacturing	36.5	25.7	25.8	63.2	73.5	74.2
Chemicals	39.9	34.0	21.9	59.7	65.4	78.1
Electricity Gas and Water Supply	18.6	17.0	18.8	80.3	81.9	80.9

Notes: The industrial classification in the table is as according to the KLEMS database. The division of employment into informal and formal segments uses the NSS data and the definition of organized sector discussed in the text. Sometimes, the informal and formal numbers may not add up to 100 because we could not classify some observations into either segment due to missing information.

Two concerns arise. First, the definition of what is primarily formal/informal seems arbitrary. As we shall see in the next section, we can consider other thresholds as well. Second, the analysis ignores the information contained in the sectors that are neither primarily formal or primarily informal. This is remedied in section 7 where the analysis exploits the heterogeneity arising from the entire spectrum of formality/informality.

Panel A of Table 4 (displayed next page) displays the raw APG (the ratio of value added per worker across non-farm and farm sector) for the non-farm segments identified as primarily formal and primarily informal. While the value added per worker relative to agriculture is much greater in either of these non-farm segments, the raw or uncorrected productivity gap is 6 to 8 times greater with the formal non-farm sector. This is suggestive of a productivity gap between the formal and informal segments of the non-farm sector as well.

We compute the labor share of value added for the primarily formal (informal) sector as the weighted average for the industries that constitutes this sector. Panel B of Table 4 reports the results. The table also reports the labor share of value added in agriculture – a figure that is readily available in the KLEMS data base.[15] As expected, the labor share in the informal part of the non-farm sector is much greater than in the formal segment. Note that labor shares in agriculture and the informal segment are similar.

Panel C of Table 4 displays α_1 −the parameter in equation (1) that defines the extent to which a worker in the non-farm sector (formal or informal) is more productive than a worker in the farm sector because of greater work hours. Annual average hours of work for an individual worker are about similar magnitudes in the formal and informal segments of the non-farm sector. However, they are substantially greater than the average work hours in the farm sector by 72-98% depending on the year. Individual annual hours of work are sourced from the nationally representative Indian Human Development Survey (IHDS). To estimate the numbers, we compute a population weighted average of labor hours of workers in the farm sector, the primarily informal non-farm sector, and the primarily formal non-farm sector.[16] The IHDS data are not available for 1999-00. In the computations, the adjustment factor for that year is assumed to be the same as that for 2004-05.

Table 4: APG Calculations for Primarily Informal and Primarily Formal Non-Farm Sectors

Panel A: Raw APG			
Sector / Segment	1999-00	2004-05	2011-12

[15] Formal employment in agriculture is negligible.

[16] The labor hours refer to the primary occupation of the individual – whether in agriculture or in non-agriculture.

Primarily Informal Non-Farm Sector	2.88	3.28	2.53
Primarily Formal Non-Farm Sector	16.67	21.49	20.27
Panel B: Labor share of value-added			
Sector / Segment	1999-00	2004-05	2011-12
Agriculture	0.53	0.51	0.52
Primarily Informal Non-Farm Sector	0.5	0.51	0.49
Primarily Formal Non-Farm Sector	0.31	0.32	0.29
Panel C: Adjustment Factors for Differences in Labor Hours			
Sector / Segment	1999-00	2004-05	2011-12
Primarily Informal Non-Farm Sector	1.72	1.72	1.85
Primarily Formal Non-Farm Sector	1.98	1.98	1.84
Panel D: Average Years of Education			
Sector / Segment	1999/00	2004-05	2011-12
Agriculture	2.51	3.24	4.18
Primarily Informal Non-Farm Sector	4.81	5.91	6.51
Primarily Formal Non-Farm Sector	8.1	9.53	10.01
Panel E: Adjustment Factors for Human Capital			
Segment	1999-00	2004-05	2011-12
Primarily Informal Non-Farm Sector	1.17	1.21	1.18
Primarily Formal Non-Farm Sector	1.48	1.56	1.5
Panel F: Corrected APG			
Segment	1999-00	2004-05	2011-12
Primarily Informal Non-Farm Sector	1.35	1.58	1.09
Primarily Formal Non-Farm Sector	3.33	4.37	4.10
Panel G: Productivity gap between primarily formal and primarily informal			
	1999-00	2004-05	2011-12
	2.46	2.77	3.75
Panel H: Corrected APG for 2004 - heterogenous returns case			
	1999-00	2004-05	2011-12
Primarily Informal Non-Farm Sector	--	1.43	--
Primarily Formal Non-Farm Sector	--	3.62	--

Labor input is also adjusted for sectoral differences in human capital (the α_2 parameter in equation (1)). The nationally representative employment data contain information on individual level years of schooling.[17] The sector wise differences in average years of education is given in panel D of Table 4. Workers in agriculture are typically poorly educated relative to workers in the non-farm sector. We also see that the education gaps are much larger relative to the formal non-farm sector than with the informal non-farm sector.

We follow GLW in converting years of education to human capital. To convert it to human capital, we assume a constant marginal rate of return on an additional year of schooling equal to 7% as estimated by Montenegro and Patrinos (2013) for South Asia. Using the Mincerian form, our formula for human capital estimation for a worker *i* who has attained n_i years of school can be given as follows:

$$Human\ Capital_i = e^{\{0.07*n_i\}}$$

Relative to agriculture, the human capital in the informal and formal segments of the non-farm sector is given in Panel E of Table 4. As expected, the gap in human capital is larger relative to the formal segment than for the informal segment.

Panel F of Table 4 brings together the information in the preceding panels and computes the corrected APG according to equation (1). The corrected APG of the informal and formal segments are greater than 1 but the departure from unity is much larger for the formal sector. The value added per worker in the (primarily) informal sector is about 9-36% greater than in agriculture. The value added per worker in the (primarily) formal sector is 233-337% greater than in agriculture. The informal sector productivity gap is at most 10% of the formal sector productivity gap. This is simply a consequence of a productivity gap between the informal and

[17] The NSSO employment survey contain information about levels of schooling. This is converted to years of schooling.

formal segments (see panel G of Table 4) – almost as large as the productivity gap between the formal sector and agriculture.

From the table it can also be inferred that the observable sectoral differences explain a substantial part of the formal sector raw productivity gap. All together, they accounted for 80% of the raw productivity gap in 2011/12 Adjusting for the difference in labor share makes the biggest difference as it accounts for nearly half of the raw productivity gap (10 percentage points). The difference in working hours also makes a substantial contribution – accounting for 5 percentage points of the gap or about one-fourth of it. The informal sector, on the other hand, is similar to the agricultural sector in terms of labor share. The human capital gap is also small. Hence, most of the difference in corrected productivity gap is because of difference in working hours.

6. Robustness

Our calculations above assumed an uniform return to human capital for all agents. The rates of return to human capital may, however, vary between the rural and the urban sector (Agarwal, 2012). Estimates of such heterogeneity in returns are available for 2004-05. Taking this into account, we recompute human capital and the corrected APG for 2004-05.[18] The results, in Panel H of Table 4, closely resemble the findings in Panel F.

A more direct measure of the productivity gap is the gap in wages between sectors. With unrestricted labor mobility, wages and, thus, the marginal productivity of labor would equalize across sectors and this can be directly checked. The APG measure, by contrast, relies

[18] We use the estimates of returns in Agrawal (2012). As Agrawal computes the returns to levels (rather than years) of learning, the expression of human capital modifies to Human Capital = exp $(\sum_{k=1}^{K} I(k) R_k Y_k)$ where k is a categorical variable depending on individual attainment of education (illiterate, informally schooled, primary schooled, middle schooled, secondary schooled, high school, and graduate), R_k are the returns to a particular level of education, and Y_k is the number of years in category k.

on a form for the production function and the benchmark level requires estimates of the output labor share. However, wage is an incomplete measure because it does not capture the productivity of one-person proprietor firms which dominate in agriculture and in the informal non-farm sector. With this caveat in place, we examine wage gaps as a supplementary measure of productivity gap.

Table 5 uses the data from the employment surveys to present the average weekly wages relative to the informal and formal segments of the nonfarm sector relative to the farm sector. In Table 5, these gaps are corrected for differences in labor inputs (weekly) and for differences in human capital in exactly the same way as the corrections for the APG This table tells the same story that we found with the APG measure, i.e., the large gap is between the entire informal sector (including the farm sector) segment and the formal segment of the non-farm sector. However, the wage gap is noticeably smaller than the APG measure.

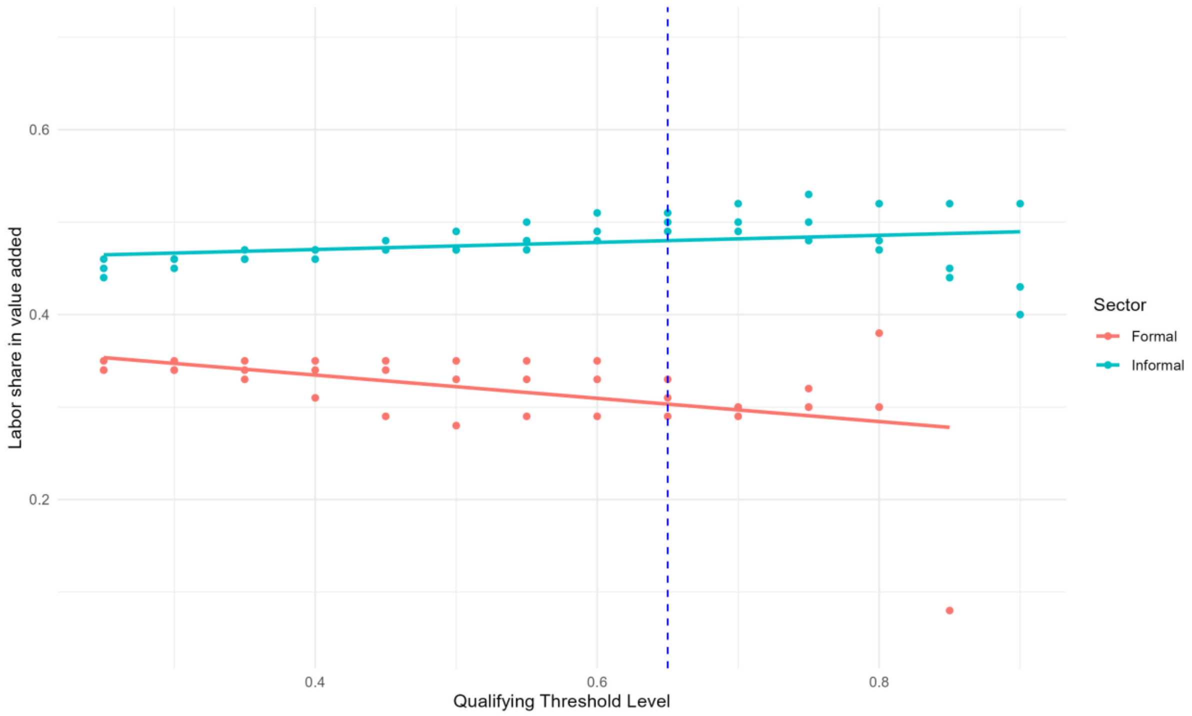
Table 5: Wage Gap Relative to Agriculture Sector. Source: NSS EUS dataset.

Segment	**Wage Gap**	1999-00	2004-05	2011-12
Primarily Informal	Raw	1.90	1.83	1.48
	Corrected	0.94	0.88	0.68
Primarily Formal	Raw	5.52	5.92	4.20
	Corrected	1.88	1.92	1.58

Notes: Computation of wage-gap doesn't require disaggregation of macro non-farm sectors into organized and unorganized segments, NSS EUS directly provides that information. The benchmark wage-gap level is equal to one or all sectors. Wages or marginal product of labor should be equalized across all sectors.

In identifying primarily formal (informal) sectors, we considered all the industrial sub-sectors where formal (informal employment) was at least two-thirds of total employment. How do our results depend on this assumption? Since our procedure is critical to the computation of labor share of value added, Figure 1 assesses how labor shares vary with different assumptions about threshold values varying in the range of 20-80%. The qualitative pattern of labor shares is almost the same for 1999-00, 2004-05, and 2011-12. For the sake of simplicity, we present the graph for the year 2011-12. The vertical line represents the threshold of 66%. Relative to the ideal of 100%, our procedure may be under-estimating labor share of value added in the informal sector (θ_i) and over-estimating the labor share of value added in the formal sector (θ_f) is over-estimated. If that is so, our computations are not correcting the raw APG for the full extent of the difference in labor share. However, it can also be seen that the labor share lines have such gentle gradients beyond the 66% threshold that the bias is negligible. A choice of threshold higher than 66% unlikely to alter the results.

Figure 1: Relationship between Labour share in value-added and choice of threshold level



Notes: The horizontal axis measures the threshold proportion of employment that is formal (or informal). All non-farm sub-sectors that are above the threshold are considered as primarily formal (or primarily informal). The vertical axis plots the labor share of value added of the sub-sectors that make up the primarily formal (or informal) segment.

7. APG and its relation to informality

In this section, we exploit the heterogeneity in the proportion of informal employment in all the 24 sub-sectors to estimate a relation between APG and the extent of informality. Since the information in panels A to E are available for each of the non-farm sub-sectors identified in the KLEMS data base (names displayed in Table 3), we compute the APG i.e., for every sub-sector k = 1,2,….24, define

$$APG_k = \frac{Y_k/L_k}{Ya/L_a}\,\frac{\theta_k}{\theta_a}\,\frac{1}{\alpha_{1k}\alpha_{2k}} \qquad (2)$$

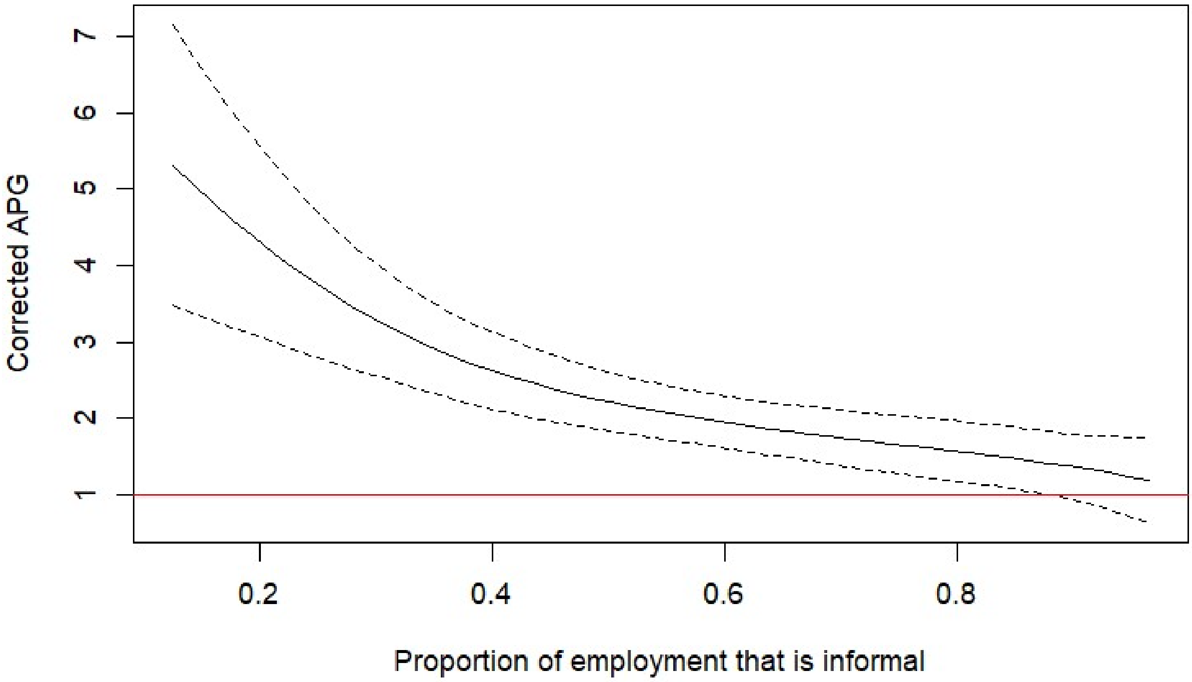
Under perfect labor mobility, (2) should be equal to 1. But for each of these sub-sectors, we also know the proportion of labor force that is informal. We estimate a non-parametric regression of APG_k on the proportion of informal employment in sub-sector k using local linear least-squares (Pagan and Ullah, 1999). To do this, we use the KLEMS data for 24 non-farm sub-sectors and the agriculture as explained in the earlier section. This data is used to compute the APG for each of the sub-sectors. The proportion of employment that is informal is obtained from the NSS data. The regression is estimated using data pooled from 1999/00, 2004/05 and 2011/10. We represent the econometric relation as the following:

$$y_{kt} = m(\,x_{kt}) + \;u_{kt} \qquad (3)$$

Where y_{kt} denotes the APG in sector k in time period t, x_{kt} is the proportion of sector k labor force that is informal, u_{kt} is an error term and $m(.)$ is the possibly non-linear functional form of the relation to be estimated by the data. The null hypothesis in the case of perfect mobility is $m\big(\,x_{jt}\big) = 1$ for all values of x, i.e., the APG does not depend on the proportion of labor force that is informal.

The results displayed in Figure 2 correspond to the bandwidth that minimizes the integrated mean squared error. The confidence intervals are computed by drawing repeated bootstrapped samples. For a given x, we draw 1000 samples and compute m(x). The average of these 1000 samples is our estimate from m(x). The percentiles at 2.5 and 97.5 form the 95% confidence interval.

Figure 2: Agricultural Productivity Gap and Employment Proportion in Informal Sector



Notes: For each of the 24 non-farm sub-sectors, the horizontal axis plots the proportion of employment that is informal. The vertical axis plots the APG for the sub-sector corrected for differences in labor shares, hours worked and human capital.

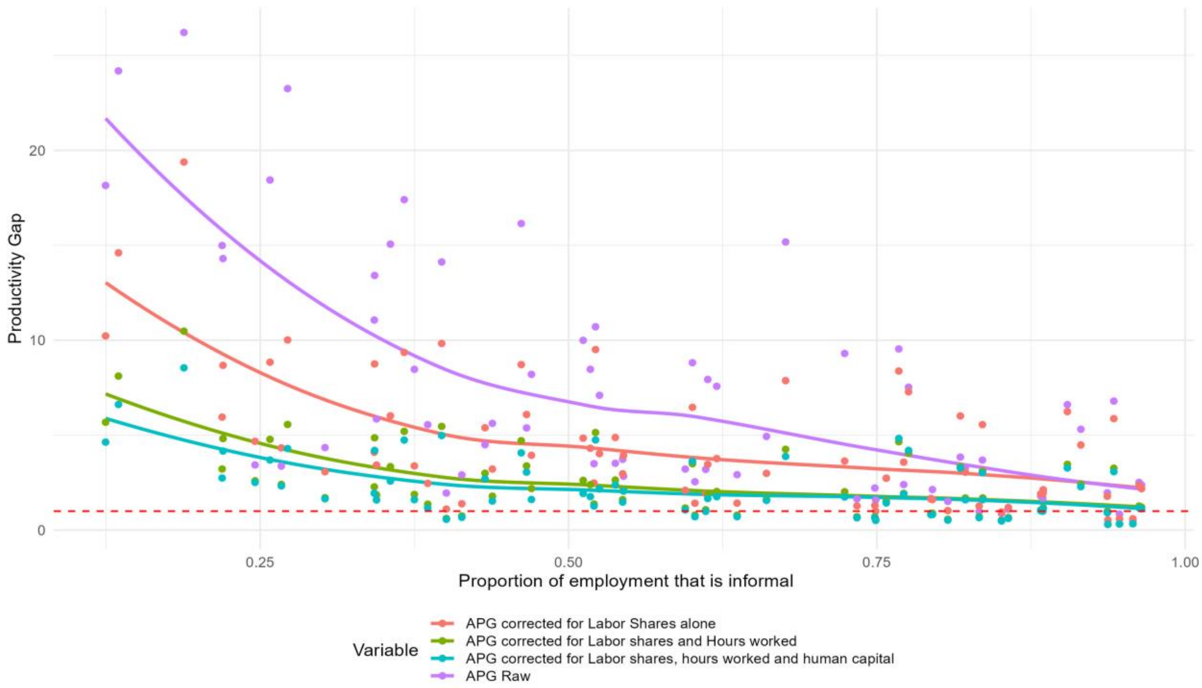
The estimated function *m*(.) is downward sloping – the APG declines as the proportion of informal employment increases. The figure also displays the 95% confidence interval. The null hypothesis of perfect mobility between agriculture and the non-farm sectors is rejected for all values of $x \leq 0.83$. On the other hand, the hypothesis is not rejected for sectors where the proportion of employment that is informal is greater than 0.83. These sub-sectors account for 40-50% of all non-farm employment (Table 6). These results confirm the findings in the earlier section – that the APG is primarily driven by the formal component of the non-farm sector. Figure 2 also demonstrates the marked heterogeneity within the non-farm sector.

Table 6: Percentage of Non-farm Employment in Sectors that are not more productive than Agriculture

Year	1999-00	2004-05	2011-12
Percentage of non-farm employment accounted by sectors where more than 83% of employment is informal	50.50	38.70	43.15

The role of sectoral differences in labor share, hours worked, and human capital in correcting the raw productivity gap is illustrated in Figure 3 which displays non-parametric regressions of raw APG, APG corrected for labor shares alone, APG corrected for labor shares and hours worked and the fully corrected APG. Like in the tabular analysis of Section 5, Figure 3 shows the major role played by labor shares and hours worked in correcting the productivity gap. It can also be seen that human capital differences do not matter much in the non-farm sub-sectors where informal workers exceeds 50% of employment

Figure 3: The Role of Sectoral Differences in Correcting APG



Notes: For each of the 24 non-farm sub-sectors, the horizontal axis plots the proportion of employment that is informal. Corresponding to it, the vertical axis plots four graphs as described by the legends above.

8. Conclusions

In this paper, we question the standard view of the two-sector agricultural productivity gap. It is well known that, in developing countries, a large part of the non-farm sector (by employment) is informal. The literature has also found informal enterprises to have lower productivity relative to formal enterprises. Acknowledging the heterogeneity in the non-farm sector, we construct agricultural productivity gap measures for the formal and informal non-farm segments of the Indian economy. Because of data limitations however, we observe, at best, sectors that are primarily formal or primarily informal or neither. We also consider the productivity gap of agriculture with respect to 24 non-farm sub-sectors that vary with regard to the extent to which they are informal.

The value added per worker in the primarily informal sector is about 2.5 to 3 times the value added per worker in agriculture. The informal sector productivity gap almost vanishes when the sectoral differences are adjusted. We find that about 40-50% of non-farm workers are in sectors whose productivity is statistically indistinguishable from agriculture. In these sectors, at least 83% of workers are in informal employment. Thus, the informal sector is not substantially more productive than the agriculture sector. The small or negligible productivity gap between agriculture and the informal non-farm sector is consistent with free labor mobility between these sectors. It should be noted that much of the uncorrected productivity gap between them arises because of greater hours of employment in the informal sector. Therefore, even if the corrected productivity gap is low, the seasonality of agricultural activity is a reason to migrate.

The value added per worker in the primarily formal sector is as much as 17-20 times the value added per worker in agriculture. Corrections for the observable sectoral differences ((labor share of value added, human capital, and working hours) explain a substantial proportion of the formal sector raw productivity gap. Nonetheless, the corrected productivity gap remains substantial. In sectors where informal employment is low (less than 40% of total), the corrected non-farm productivity is 4 to 5 times more than that of agriculture. The non-farm formal sector is much more productive than either agriculture or the informal component of the non-farm sector. Our analysis supports the view that the dualism in developing economies is primarily between its formal and informal components including agriculture (La Porta and Shleifer, 2014).

References

Alvarez, J. A. (2020). The agricultural wage gap: Evidence from Brazilian micro-data. *American Economic Journal: Macroeconomics*, *12*(1), 153-173.

Bonnet, F., Vanek, J., & Chen, M. (2019). Women and men in the informal economy: A statistical brief. *International Labour Office, Geneva*, *20*.

Beegle, K., De Weerdt, J., & Dercon, S. (2011). Migration and economic mobility in Tanzania: Evidence from a tracking survey. *Review of Economics and Statistics*, *93*(3), 1010-1033.

Bryan, G., Chowdhury, S., & Mobarak, A. M. (2014). Underinvestment in a profitable technology: The case of seasonal migration in Bangladesh. *Econometrica*, *82*(5), 1671-1748.

Ciani, A., Hyland, M. C., Karalashvili, N., Keller, J. L., & Tran, T. T. (2020). *Making It Big: Why Developing Countries Need More Large Firms*. World Bank Publications.

Djido, Abdoulaye I., and Bekele A. Shiferaw. "Patterns of labor productivity and income diversification–Empirical evidence from Uganda and Nigeria." *World Development* 105 (2018): 416-427.

Donovan, K., & Schoellman, T. (2023). The role of labor market frictions in structural transformation. *Oxford Development Studies*, 1-13.

Fuglie, K., Gautam, M., Goyal, A., & Maloney, W. F. (2019). *Harvesting prosperity: Technology and productivity growth in agriculture*. World Bank Publications.

Gollin, D., Lagakos, D., & Waugh, M. E. (2014). The agricultural productivity gap. *The Quarterly Journal of Economics*, *129*(2), 939-993.

Hamory, J., Kleemans, M., Li, N. Y., & Miguel, E. (2021). Reevaluating agricultural productivity gaps with longitudinal microdata. *Journal of the European Economic Association*, *19*(3), 1522-1555.

Hayami, Y. (1969). Sources of agricultural productivity gap among selected countries. *American Journal of Agricultural Economics*, *51*(3), 564-575.

Hayfield T, Racine JS (2008). “Nonparametric Econometrics: The np Package.” Journal of Statistical Software, 27(5), 1-32. doi:10.18637/jss.v027.i05, https://www.jstatsoft.org/index.php/jss/article/view/v027i05.

Herrendorf, B., & Schoellman, T. (2018). Wages, human capital, and barriers to structural transformation. *American Economic Journal: Macroeconomics*, *10*(2), 1-23.

India. National Commission for Enterprises in the Unorganised Sector, and Academic Foundation (New Delhi, India). *Report on conditions of work and promotion of livelihoods in the unorganised sector*. Academic Foundation, 2008.

Imbert, C., & Papp, J. (2020). Costs and benefits of rural-urban migration: Evidence from India. *Journal of Development Economics*, *146*, 102473.

La Porta, R., & Shleifer, A. (2008). The Unofficial Economy and Economic Development. *Brookings Papers on Economic Activity*, 275-352.

La Porta, R., & Shleifer, A. (2014). Informality and development. *Journal of economic perspectives*, *28*(3), 109-126.

Lagakos, D. (2020). Urban-rural gaps in the developing world: Does internal migration offer opportunities?. *Journal of Economic perspectives*, *34*(3), 174-192.

McCullough, E. B. (2017). Labor productivity and employment gaps in Sub-Saharan Africa. *Food policy*, *67*, 133-152.

McMillan, M. S., & Rodrik, D. (2011). *Globalization, structural change and productivity growth* (No. w17143). National Bureau of Economic Research.

Murthy, S. R. (2019, November). Measuring informal economy in India–Indian experience. In *Seventh IMF Statistical Forum, Washington, DC*.

Nagaraj, R., & Kapoor, R. (2022). What is 'Formalisation'of the Economy? India's quest to shrink its informal sector. In *The India Forum*.

OECD (2014), "Productivity by enterprise size", in *Entrepreneurship at a Glance 2014,* OECD Publishing, Paris, https://doi.org/10.1787/entrepreneur_aag-2014-11-en.

Pagan A R and Ullah Aman. (1999). "Non-parametric econometrics'', Cambridge University Press

Restuccia, D., Yang, D. T., & Zhu, X. (2008). Agriculture and aggregate productivity: A quantitative cross-country analysis. *Journal of monetary economics*, *55*(2), 234-250.

Rivera-Padilla, Alberto. "Crop choice, trade costs, and agricultural productivity." *Journal of development economics*146 (2020): 102517.

Vollrath, D. (2009). How important are dual economy effects for aggregate productivity?. *Journal of development economics*, *88*(2), 325-334.

Young, A. (2013). Inequality, the urban-rural gap, and migration. *The Quarterly Journal of Economics*, *128*(4), 1727-1785.