The Agricultural Productivity Gap: Informality Matters*

Rajveer Jat
rjat001@ucr.edu
University of California, Riverside, USA

and

Bharat Ramaswami
bharat.ramaswami@ashoka.edu.in
Ashoka University, India

August 2025

Abstract

The measured agricultural productivity gap (APG) in developing countries typically compares agriculture with the entire non-farm economy, implicitly treating the latter as homogeneous. In developing countries, most non-farm employment is informal, concentrated in small, unregistered enterprises with low productivity. This paper compares the productivity of agriculture to the informal and formal non-farm sectors in India. Using Indian sectoral data from the India KLEMS database linked with nationally representative labor surveys, we decompose the non-farm economy into formal and informal segments and adjust productivity measures for differences in hours worked, human capital, and labor's share of value-added. We find that the APG is almost entirely driven by the small formal non-farm sector. The gap with the informal sector is negligible. Between 63-75% of non-farm workers are in informal employment dominated industries that are not more productive than agriculture.. These results reframe the APG as a formal–informal divide.

Keywords: Agricultural productivity gap, Informal sector, India

JEL Code: O11, O13, O41

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 demonstrates 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 non-farm 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. There might be productivity differences between staple crops and cash crops (Djido and Shiferaw, 2018, 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 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 Porta 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. The low productivity associated with the informal sector is one reason why policies have been proposed to formalize the informal sector and thereby shrinking it (e.g., Benhassine, 2018, Campos et.al, 2023).[4]

In the Indian context, informal enterprises accounted for 43% of non-farm GDP in 2017 (Murthy, 2019).[5] 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,

[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] For surveys of this literature, see La Porta and Shleifer, 2014, Ulyssea, 2020.
[5] 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.

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.

In this paper, we probe these disparities further using sectoral aggregates of output, employment and labor shares. 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. The contribution of this paper to the literature, the specific relation to the GLW paper and data sources are discussed in Sections 2, 3 and 4.

We use a disaggregation of the non-farm economy into 24 industry divisions. It turns out that there is no industry division that is fully informal or fully formal. Workers are employed in both formal and informal enterprises. However, sectors vary according to the proportion of its workers that are employed in informal enterprises. We use this information to construct an index of informality. In a first exercise (in Section 5), we use thresholds to come with a binary index to classify an industry division as either formal or informal. Aggregating over the industries that make up each category, we compute the labor productivity of the formal sector (or the informal) sector relative to agriculture. This corresponds to a three sector model (agriculture, informal non-farm and formal non-farm).

In a second and more disaggregated approach (in Section 6), we take the proportion of workforce in each industry division that is informal as the index of informality. Through non-parametric means, we estimate the relation between informality and the agricultural productivity gap. This corresponds to a 25 sector model (agriculture and 24 non-farm sectors) where the sectors vary by the extent of informality. Finally, we supplement these approaches by looking at sectoral wage gaps in micro data. Here too, sectoral wage gaps are

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, wage data would be missing for the self-employed who form a large proportion of the informal sector. Indeed, La Porta and Shleifer (2008) use the percentage of the non-agricultural labor force that is self-employed as an indicator of informality. [6]

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 segment of the non-farm sector and the farm sector. The two findings reflect a third: that there is a sizeable productivity gap between the informal and formal segments of the non-farm sector. The non-parametric analysis, disaggregated at the industry level, shows that the results depend on the extent of informality. When informal workers account for more than 76% of employment in a non-farm industry division, we are unable to reject the null hypothesis of a zero productivity gap. The employment in such industries account for 63-75% of all non-farm employment – the range indicating the variation across the sample years.

Like GLW, our paper is agnostic about the source of the productivity gap with the formal sector – whether selection or whether mobility barriers. The implication of our findings 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. The dual economy view of informality sees it as a cause and consequence of

[6] 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).

under-development. This is corroborated (in Section 7) by the negative correlations of informal employment with various indicators of development (urbanization, banking access, education, and lower availability of low skilled labour).

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.

2. Relation to Literature

Sectoral differences in productivity and wages have long been associated with features of developing countries (Lewis, 1954; Harris and Todaro, 1970). More recent 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 re-allocating labor 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.[7] 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

[7] 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

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 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.[8] We extend their methods in two directions. We do away with their assumption that labor shares in value-added are the same across sectors because the formal segment comprising of larger production units are unlikely to use the same technologies as the smaller 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 non-farm 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 first two decades of this century. With Cobb-Douglas production functions (in labor and capital) and common factor shares across sectors, GLW argue that this ratio ought to be unity whenever labour is mobile

[8] Alvarez (2020) examined the APG between formal sector workers in agriculture and those in non-agriculture but not between informal and formal sector workers. Herrendorf and Schoellman (2018) examine agriculture's wage gap with different sub-sectors of the non-farm sector – industry and services and as well for two components of services – skilled services and unskilled services.

across sectors. However, the APG in India (first row of Table 1) has fluctuated around 4 during this period – a value that is close to the global average of this variable (GLW, 2014).

As GLW point out, non-agricultural workers typically work more hours and are better educated. This applies to India as well. Over the period 1999 to 2022, non-agricultural work hours 79% higher than in agriculture. Similarly, human capital in non-agriculture was 27% 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 more than 50% of the unadjusted gap. Despite the adjustments, 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.[9]

To address this, let's consider an economy with three sectors, agriculture (called the A sector) and a non-farm sector broken up into two segments: informal and a formal non-segment. The sectoral earnings per worker (which we call wage) is defined by the identity

$$w_j = \theta_j \frac{Y_j}{L_j} \quad , j = a, i, f$$

where a, *i* and *f* subscript the agricultural, informal non-farm and formal non-farm sectors respectively, θ_j is the labour share of value-added, Y_j is value-added, and L_j is labour employed in the sector. If labor is freely mobile across sectors, all workers receive the same wage. Hence,

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

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

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.

To account for the sectoral differences in human capital and in hours worked, the labour input in the non-farm sectors can be measured in terms of the efficiency of agricultural labor. Let the efficiency parameter α_{1j} denote the differences in working hours between non-farm sector *j* and the agricultural sector. Similarly, let α_{2j} denote the efficiency of non-farm labor in sector *j* (relative to agriculture) because of greater human capital. In efficiency inputs, the labour input in non-farm sector *j* is $L_j^e = \alpha_{1j}\alpha_{2j}L_j$.

Replacing L_j by its efficiency counterpart, (1) can be rewritten as

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

The first term on the left hand side of (2) 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 correcting 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. While approximately true when comparing agriculture with all of non-farm sector, it is worth relaxing the assumption when comparing the larger production units in the formal sector with the smaller production units in the informal sector. 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 micro-economic data on 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. 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 and Definitions

Our primary data set is the India KLEMS database.[10] For 27 industry divisions of the economy (including agriculture), the database offers time series on gross output, value-added, the primary inputs of capital and labour and the intermediate inputs of energy, materials and services. To ensure international comparability, the industrial classification and the methodology are aligned with the European Union KLEMS research project.

We use the KLEMS data for industry level magnitudes relating to value-added, labor input and share of labor in value-added.[11] For value-added, the KLEMS data draws upon the national accounts statistics. However, the sectoral disaggregation in national accounts is more coarse than the classification in KLEMS. This is especially true for the manufacturing sector. Hence, KLEMS splits the national accounts figures for value-added in manufacturing sector into a number of industries based on output proportions obtained from survey data on manufacturing. For industry division employment, KLEMS sources the data from employment surveys of individuals. These surveys reports the industry where the employed individuals work. In this paper, we rely on KLEMS for estimates of the industry division

[10] https://rbi.org.in/Scripts/KLEMS.aspx

[11] 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.

value-added per worker. These are available from 1980-81 to 2022-23 at the country level. Estimates are not available at lower levels of spatial disaggregation.

The other key variable that this paper takes from KLEMS is the sectoral labor shares of value-added. The national accounts statistics reports the compensation of employees and the income of the self-employed. The total labor compensation is, however, the sum of employee compensation and the labor income component of self-employed income. The latter is not available in national accounts and is, therefore, computed by KLEMS for each of the industry divisions. The labor income component is estimated based on wage rates of hired labor and the survey based estimates of employment (number of workers and days of worker) of different categories of workers (self-employed and hired workers)

The KLEMS data do not distinguish between formal and informal components within each industry division's magnitude of value-added, employment and labor share. To do that, we supplement the KLEMS data with employment data from nationally representative surveys of the National Sample Survey Organization (NSSO). As the data contain industry of employment, we classify the sampled individuals into the industry divisions defined by KLEMS. The data also provides information about the characteristics of the enterprises where the surveyed individuals work.

Following the guidelines by a government commission (Report on Conditions of Work and Promotion of Livelihoods in the Unorganized Sector, 2007) [12], we consider individuals who satisfy any one of the following conditions as formally employed: if they work for an enterprise that belongs to either the government, public sector enterprises, private limited company, cooperative societies or non-profit institutions; if they work for an

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

enterprise that offers some kind of social security benefits (pension, gratuity or health care benefits); if they work for an enterprise that consists of, at least, 10 workers.

The remainder workforce is informal. They are characterized by the intersection of above three conditions: they are not employed in either the government, public sector enterprise, private limited company, cooperative societies or non-profits, they do not receive any kind of social security benefits and they are employed in enterprises with less than 10 employees.

These criteria also correspond to a sub-set of operational measures of employment in the informal sector proposed by International Labor Organization or ILO (Bonnet, Vanek and Chen, 2019). The close relation of our measure to other notions of informal employment (e.g., proprietary or partnership enterprises, lack of work benefits such as paid leave and the absence of a job contract) is discussed in Appendix A1.

Even though the productivity data is available for all of the years in the period 1980-2022, we are constrained by the limited availability of employment surveys. We use data at 5 time points: 1999/00, 2004/05, 2011/12, 2018/19 and 2022/23.[13] Starting in 2018/19, employment surveys began to be reported annually. However, because of the Covid shock, we chose the last pre-Covid year (2018/19) and the first post-Covid year (2022/23) and omit the intervening years.

The productivity data is adjusted for sectoral differences in human capital and hours worked. The NSS0 employment survey contain information about levels of schooling. This is converted to years of schooling. Individual annual hours of work are sourced from the nationally representative Indian Human Development Survey (IHDS). [14] The labor hours refer to the primary occupation of the individual – whether in agriculture or in non-

[13] Estimates are reported for the financial year – April of year t to March of year (t+1).
[14] Indian Health and Demographic Survey (IHDS) data is a detailed household-level micro-data. Link: https://ihds.umd.edu

agriculture. Workers in the NSSO dataset and the IHDS dataset are mapped to the KLEMS industry of employment using the National Industry Classification (NIC) codes. We use the mapping to compute the industry division labor inputs (human capital and hours worked) relative to agriculture.

The wage gaps are estimated using NSSO employment surveys. As we have data on daily wages, we adjust the data for differences in human capital alone along the same lines as the macro data. The correlates of the share of informal employment are primarily obtained from the population censuses in 2001 and 2011.

Table 2 summarizes the variables and their data sources.

5. APG Estimates

Following these definitions of informal employment discussed earlier, the paper computes the degree of informality in an industry division as the proportion of the industry workforce that is employed in informal enterprises. This is shown in Table 3.

From Table 3, it can be seen that there is no industry division 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). If more than two-thirds (66%) of employment in a particular sector is formal, then we call the sector as `primarily formal'. A similar definition applies to `primarily informal' sectors. In this section, we shall consider the APG for sectors that are primarily formal and the sectors that are primarily informal. By our procedure, 9 non-farm

industry divisions are identified as primarily informal and 11 as primarily formal in 2022/23.[15] 4 sectors cannot be identified as either formal or informal.

Two concerns arise. First, the definition of what is formal/informal seems arbitrary. The sensitivity of the findings to the threshold is examined later in this section. Second, the analysis ignores the information contained in the intermediate sectors that are neither formal or informal. Both of these concerns are fully resolved in section 6 where the analysis exploits the heterogeneity arising from the entire spectrum of formality/informality and is, therefore, not reliant on threshold definitions for when a sector is formal (informal).

Table 4 displays the sector (formal/informal) weighted averages of the KLEMS industry divisions that constitute it. Sectoral averages of value-added per worker, years of education and hours of work are employment weighted. The sectoral average of labor share is weighted by each industry's value-added. In addition, industries that are further away from the threshold receive greater weight.[16] Panel A of Table 4 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, primarily informal and the intermediate sector. While the raw APG is much greater than 1 in in both of the non-farm segments, the formal sector APG is between 2 to 5 times larger than the informal sector productivity gap. This implies a productivity gap between the formal and informal segments of the non-farm sector as well.

[15] It can be seen from Table 2 that in 2022/23, the industries identified as primarily informal are Trade, Wood and Cork, Individual Services, Hotel and Restaurants, Construction, Transport and Storage, Recycling, Textile Leather and Footwear, Food Beverages and Tobacco. The industries identified as primarily formal were Electric Equipment Manufacturing, Mining and Quarrying, Electricity Gas and Water Supply, Health-Social Work, Machinery Manufacturing, Finance and Banking Services, Rubber and Plastic Manufacturing, Education, Post and Telecommunications, Chemicals and Motor Vehicle and Manufacturing. Industries that could not be identified as primarily formal or informal were Pulp Paper Printing/Publishing, Business Services, Basic Metal Manufacturing and Non-Metallic Manufacturing.

[16] We apply the squared distance from the threshold as an additional weight. Abstracting from the employment weights, this implies that industries with a higher degree of formality (or informality) receive proportionately greater weight than industries closer to the threshold, provided both fall on the same side of it. The detailed construction of these measures is presented in Appendix A2.

Panel B of Table 4 reports the labor share of value-added in the primarily formal and primarily informal sectors. The table also reports the labor share of value-added in agriculture – a figure that is readily available in the KLEMS data base.[17] As expected, the labor share in the informal part of the non-farm sector is greater than in the formal segment and closer to the labor share in agricultural value-added.[18]

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 80% or more (depending on the year). The labor hours 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.

Labor input is also adjusted for sectoral differences in human capital (the α_2 parameter in equation (1)). 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. 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

[17] Formal employment in agriculture is negligible.
[18] The labour share takes into account the quantum of employment (and hence the hours of work) and the compensation (implicitly accounting for human capital differences). Hence the labour share is not adjusted for these variables.

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 between agriculture and the formal sector is larger than the human capital gap between agriculture and the informal sector.

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. Depending on the year, the value-added per worker in the (primarily) informal sector is about 8-38% greater than in agriculture. On the other hand, agriculture's productivity gap relative to the formal non-farm sector is much greater. This is simply a consequence of a substantial productivity gap between the informal and formal segments (see panel G of Table 4).

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 the adjustments make a difference and collectively they account for about 80% of the raw productivity gap in all of the years. 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.

A micro 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 on a macro estimate of sectoral earnings and therefore requires estimates of the labor share of value-added. 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.[19] 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 earnings in the informal and formal segments of the non-farm sector relative to the farm sector. The first and third rows contain the formal (and informal) wage gaps for daily earnings – hence the gaps already correct for differences in labor inputs. The second row and third row add the additional corrections for human capital.[20] With corrections, the Table shows no difference in value-added per worker between agriculture and the informal non-farm sector while the corrected formal sector wage gap is large.

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 the APG varies with the threshold when it varies between 50 to 100%. The formal sector APG is an increasing function of the threshold. The informal sector APG is a declining function of the threshold. Both of these results are consistent with the hypothesis that the APG is increasing in the

[19] In 2022/23, 73% of non-farm employment was accounted by single proprietor enterprises (Periodic Labour Force Surveys, NSSO).

[20] To measure labor input, we use the responses that describe the number of days worked in the reference week of the survey.

formality index while it is decreasing in the informality index. The APG is minimum at a threshold level of 50% and increases monotonically for higher threshold levels.

6. APG and its relation to informality

In this section, we exploit the heterogeneity in the proportion of informal employment in all the 24 industry divisions to estimate a relation between APG and the extent of informality. Since the information in panels A to E (Table 4) are available for each of the non-farm industry divisions identified in the KLEMS data base (names displayed in Table 3), we compute the APG i.e., for every industry division k = 1,2,….24, i.e.,

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

where a is the index for agricultural variables and parameters, and k is the index for the variables and parameters that belong to industry division k.

Under perfect labor mobility, (2) should be equal to 1. For each of these industry divisions, 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 industry division k using local linear least-squares (Pagan and Ullah, 1999). The regression is estimated using data pooled from 1999/00, 2004/05, 2011/12, 2018/19 and 2022/23. 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 $m(x_{jt}) = 1$ 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.

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 is rejected for all values of $x \leq 0.76$. The non-farm industry divisions that exceed this threshold can be read off Table 2. These industry divisions account for 63-75% 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 demonstrates the marked heterogeneity within the non-farm sector.

The role of sectoral differences in labor share, hours worked and human capital in correcting the raw productivity gap is illustrated in Figure 3 that 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 industry divisions where informal workers exceeds 50% of employment.

The non-parametric regression uses pooled data from 120 points – 24 industry divisions at 5 time points. By interpolating missing observations on informal employment for

the intervening years, the sample size can be increased substantially to 600 observations. The non-parametric regression in this case is displayed in the appendix (Appendix A3).

7. **Correlates of informal employment**

Informal employment has been seen through different lenses by researchers (La Porta and Shleifer, 2014; Ulyssea, 2020). In the `exclusion' view, informality happens when enterprises are excluded from formality by regulation that denies them access to public goods and formal sector inputs (e.g., credit, clear property titles). In the `exit' view, informality is the outcome of choice by enterprises for whom avoiding the burden of taxes and regulation outweigh the benefits of formalization. The `dual' view sees informal sector as a characteristic of under-development. The informal and formal sector produce different goods or goods of different qualities with different technologies. The informal sector goods are low quality and meet the needs of the poor. With development, the informal sector declines. The supply of workers to the low productivity informal sector shrinks just as the demand for low quality goods contracts.

A registered business enterprise in India is subject to direct taxes, payment of indirect taxes collected from customers, safety and security regulations and labor regulations.[21] Using state level variation, Besley and Burgess (2004) showed that stronger labor legislation biases outcomes towards a smaller formal manufacturing sector and a larger informal manufacturing sector. Similarly, Chaurey (2017) found that formal sector firms in manufacturing respond to temporary demand shocks by hiring more contract labor because it does not enjoy protections mandated by labor regulation. However, most informal enterprises such as single proprietor households may be too small to be affected by the regulations that govern formal enterprises (Kanbur, 2017).

[21] Vij, Khanna and Srivastava (2017).

Figures 4 plots the per worker usage of intermediate inputs in each KLEMS industry division against the proportion of industry employment that is informal. The correlations of each type of intermediate input (energy, materials, services) are displayed in Table 7. The correlations are consistent the `dual' view that the informal sector technology is different from that of the formal sector. Greater use of intermediate inputs is associated with increasing returns at the industry level and with greater propagation of shocks (e.g., Ciccone 2002, Carvalho and Tahbaz-Salehi, 2019). The limited use of intermediate inputs limits productivity in the informal sector while also keeping its economic activity partly segregated from the formal sector (La Porta and Shleifer, 2014).

A second source of variation in informal employment is spatial. Figure 5 displays the spatial variation in informal employment (as share of non-farm employment) across India's districts. The histogram pools the informal employment share data from employment surveys in 1999/00 and 2011/12. We extract the data for these two years because they can be correlated with district level variables from the population censuses of 2001 (nearest to 1999/00) and 2011.[22] The variables we consider measure the trend (across the two periods), landless agricultural labor (proxying availability of low-skilled workers), human capital (average years of education), urbanization, banking access and social group composition measuring the presence of disadvantaged castes and tribes. The coefficients and their 5% significance levels are plotted in Figure 6.

The 2011 dummy is negative – controlling for other variables, informal employment has declined – a trend that's consistent with the dualism hypothesis that informal sector declines with development. The share of agricultural landless labor in the agricultural work force is positively associated with informal employment. This is also consistent with the dual view. Landless agricultural labor does not own land, and they are characterized by higher

[22] The 2021 population census was not carried out because of the Covid pandemic.

levels of poverty relative to other population groups. Poverty increases the supply of low-skilled workers to the informal sector while also increasing the demand for the low-quality products of informal firms. The result is consistent with easy entry into low skilled informal sector occupations by agricultural labor.[23]

Expectedly, the average years of education of the workforce, the urban share of the population and the per capita availability of banks are negatively correlated with informal employment. `Scheduled tribes' and `scheduled castes' are social groups recognized as so historically disadvantaged that they are constitutionally guaranteed affirmative action policies, especially in terms of representation in Parliament, public sector jobs, and education. Their access to jobs, education and public goods is regarded as poor. That being the case, it is not much of a surprise that the proportion of scheduled castes in the district is positively correlated with the informal employment share. It is a surprise, however, the proportion of scheduled tribes is negatively associated with informal employment. This could be driven by their presence in forest districts that tend to be the location of formal sector mining. The inclusion of a variable for mining activity cuts the magnitude of the scheduled tribe coefficient by half but the coefficient continues to be significant.

The association of informal employment with state-level measures of labor regulation (not reported in Figure 6) were both economically small and statistically insignificant.[24] This may well reflect the dominance of single proprietor enterprises that are too small to be affected by regulation. Figure A4 in the appendix displays regression coefficients after controlling for state-fixed effects. The state-fixed effects control for state specific features

[23] As noted earlier, the informal sector consists of small enterprises, mostly single proprietors, and mostly without an office outside their home (or the street). Even when they hire workers, they tend to be without a written contract or social security benefits. Hence, these enterprises are likely to be characterized by low fixed costs and therefore, low barriers to entry.

[24] The measures of labor regulation coded states as pro-worker, pro-employer and neutral. These were sourced from Chaurey (2017).

such as regulatory environment and other variables such as natural endowments, industrial policy and state GDP. The major change compared to Figure 6 is the sharp drop in the partial correlation of per capita availability of banks and of the proportion of scheduled castes. Other variables continue to remain significant predictors.

8. Conclusions

This paper revisits the conventional two-sector view of the agricultural productivity gap by recognizing the deep heterogeneity within the non-farm economy. In India, as in many developing countries, the majority of non-farm employment is informal — and informal enterprises operate with much lower productivity than the formal segment. By separating the non-farm sector into formal and informal components, we find that the standard APG narrative requires important qualification.

Our results show that the productivity gap between agriculture and the informal non-farm sector is small or negligible once differences in labor share, human capital, and hours worked are accounted for. In fact, 63–75% of non-farm workers are in industries whose productivity is statistically indistinguishable from that of agriculture. By contrast, the formal non-farm sector is substantially more productive: even after adjustments, its value-added per worker is two to three times higher than that of agriculture. The large aggregate APG is therefore driven primarily by the small but economically dominant formal segment of the non-farm economy.

These findings have two important implications. First, labor mobility between agriculture and the informal sector appears relatively unconstrained — the negligible productivity gap suggests that workers can move freely between them. 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.

Second, the real structural dualism in the Indian economy is between the high-productivity formal sector and the low-productivity informal sector (including agriculture), not simply between "farm" and "non-farm." While our analysis is specific to India, the presence of large informal segments is common in many low-income economies, suggesting that our findings may have wider relevance.

References

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

Benhassine, N., McKenzie, D., Pouliquen, V., & Santini, M. (2018). Does inducing informal firms to formalize make sense? Experimental evidence from Benin. *Journal of Public Economics*, *157*, 1-14.

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.

Campos, F., Goldstein, M., & McKenzie, D. (2023). How should the government bring small firms into the formal system? Experimental evidence from Malawi. *Journal of Development Economics*, *161*, 103045.

Carvalho, V. M., & Tahbaz-Salehi, A. (2019). Production networks: A primer. *Annual Review of Economics*, *11*(1), 635-663.

Chaurey, R. (2015). Labor regulations and contract labor use: Evidence from Indian firms. *Journal of Development Economics*, *114*, 224-232.

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.

Ciccone, A. (2002). Input chains and industrialization. *The Review of Economic Studies*, *69*(3), 565-587.

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.

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

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

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.

Kanbur, R. (2017). Informality: Causes, consequences and policy responses. *Review of Development Economics*, *21*(4), 939-961.

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.

Lewis, W. A. (1954). "Economic Development with Unlimited Supplies of Labor." *The Manchester School* 22(2): 139–91.

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.

Ulyssea, G. (2020). Informality: Causes and consequences for development. *Annual Review of Economics*, *12*(1), 525-546.

Vij, J., Khanna, A., & Srivastava, P. (2017). Informal economy in India: Setting the framework for formalisation. *New Delhi: Konrad Adenauer Foundation and Federation of Indian Chambers of Commerce and Industry*.

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.

Table 1 : Agricultural Productivity Gap (APG)

Estimate	1999-00	2004-05	2011-12	2018-19	2022-23
APG	3.83	4.24	3.82	3.83	3.86
Corrected APG	1.47	1.57	1.41	1.38	1.42

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. In the second row, APG is corrected for sectoral differences in labor hours and human capital. The labor hours adjustment factors are derived from the India Human Development Surveys. The human capital adjustment factors are derived from NSSO employment surveys.

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	NSSO Employment Survey & Periodic Labour Force Survey
Percentage of employment that is formal, by non-farm sub-sector	NSSO Employment Survey & Periodic Labour Force Survey
Human capital in agriculture and in non-farm sub-sector	NSSO Employment Survey & Periodic Labour Force Survey
Hours worked in agriculture and in non-farm sub-sector	Indian Human Development Survey
Average weekly wages in agriculture and in the non-farm sector (formal/informal)	NSSO Employment Survey & Periodic Labour Force Survey
District level Informal Employment as share of Non-Farm Employment	NSSO Employment Survey
District Level Average Years of Education of Labor Force	NSSO Employment Survey
Share of Population that is Urban	Population Census
Number of Bank Branches Per Capita	Population Census
Agricultural Landless Labour as share of Agricultural Labor Force	Population Census
Share of Population that belongs to Scheduled Castes	Population Census
Share of Population that belongs to Scheduled Tribes	Population Census

Table 3: Informal Employment (%) Employment in Non-farm Industry Divisions.

	Informal Employment (%)				
Sector	**1999-00**	**2004-05**	**2011-12**	**2018-19**	**2022-23**
Trade	95.84	96.89	96.11	90.68	91.88
Wooden Products and Cork	94.85	90.59	94.90	88.84	91.52
Other Individual Services	93.35	95.26	92.62	84.90	86.17
Hotels and Restaurants	89.82	82.93	88.11	82.47	85.48
Construction	84.77	82.36	78.63	79.18	82.50
Transport and Storage	78.46	89.21	81.31	79.06	82.25
Textile, Leather, and Footwear	76.42	81.10	79.25	74.30	77.21
Other Manufacturing and Recycling	86.73	80.77	77.43	76.00	72.49
Food, Beverages, and Tobacco	79.21	80.68	75.32	70.49	71.83
Pulp Paper Printing and Publishing	63.27	60.70	55.34	55.88	51.43
Business Services	80.81	79.89	63.53	43.54	47.72
Basic Metal Manufacturing	63.84	73.35	63.13	48.45	46.42
Non-Metallic Manuf. and Minerals	64.76	60.53	49.08	50.58	44.79
Electrical Equip. Manufacturing	32.38	52.46	25.06	22.11	16.01
Mining and Quarrying	42.57	36.89	30.90	28.53	24.42
Electricity, Gas, and Water Supply	14.29	13.77	21.04	23.14	21.85
Health Social-work	46.74	47.86	43.55	23.67	21.35
Machinery Manufacturing	49.63	52.76	35.80	28.51	20.81
Finance and Banking Services	23.62	34.73	37.62	19.22	20.50
Rubber and Plastic Manufacturing	40.00	44.91	33.59	20.54	20.49
Education	27.90	28.03	24.57	16.56	19.35
Post and Telecommunication	42.37	52.74	35.74	24.50	17.15
Chemicals	39.29	35.01	21.67	12.22	13.09
Motor Vehicle Manufacturing	37.58	27.43	21.39	15.56	9.18

Notes: The industrial classification in the table is as according to the KLEMS database. The sectors "Public Administration, Defense and Social Security" and "Petroleum and Nuclear Fuel" are omitted. The division of employment into informal and formal segments uses the NSS employment data (the employment-unemployment surveys for 1999/00, 2004/05 and 2011/12 and the Periodic Labour Force Survey for 2018/19 and 2022/23) and the definition of formal sector discussed in the text.

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	2018-19	2022-23
Primarily Informal Non-Farm Sector	3.02	3.38	2.91	2.92	2.65
Primarily Formal Non-Farm Sector	14.43	12.45	7.54	6.53	8.24
Panel B: Labor share of value-added					
Sector / Segment	1999-00	2004-05	2011-12	2018-19	2022-23
Agriculture	0.56	0.55	0.55	0.56	0.56
Primarily Informal Non-Farm Sector	0.55	0.51	0.50	0.52	0.53
Primarily Formal Non-Farm Sector	0.32	0.32	0.37	0.39	0.38
Panel C: Adjustment Factors for Differences in Labor Hours					
Sector / Segment	1999-00	2004-05	2011-12	2018-19	2022-23
Primarily Informal Non-Farm Sector	1.89	1.90	1.94	1.90	1.84
Primarily Formal Non-Farm Sector	1.89	1.83	1.79	1.87	1.83
Panel D: Average Years of Education					
Sector / Segment	1999/00	2004-05	2011-12	2018-19	2022-23
Agriculture	2.51	3.24	4.18	5.14	5.22
Primarily Informal Non-Farm Sector	4.86	6.00	7.36	8.46	8.47
Primarily Formal Non-Farm Sector	10.06	11.15	12.03	13.13	12.80
Panel E: Adjustment Factors for Human Capital					
Segment	1999-00	2004-05	2011-12	2018-19	2022-23
Primarily Informal Non-Farm Sector	1.18	1.21	1.25	1.26	1.26
Primarily Formal Non-Farm Sector	1.70	1.74	1.73	1.75	1.70
Panel F: Corrected APG					
Segment	1999-00	2004-05	2011-12	2018-19	2022-23
Primarily Informal Non-Farm Sector	1.34	1.37	1.09	1.11	1.08
Primarily Formal Non-Farm Sector	2.55	2.28	1.62	1.40	1.78
Panel G: Productivity gap between primarily formal and primarily informal					
	1999-00	2004-05	2011-12	2018-19	2022-23
	1.90	1.66	1.48	1.25	1.66

Notes: The primarily formal sector consists of all industry divisions where the proportion of employment that is formal is greater than two-thirds. The primarily informal sector consists of all industry divisions where the proportion of employment that is informal is greater than two-thirds.

Table 5: The Agricultural Wage Gap

Segment	Wage Gap	1999-00	2004-05	2011-12	2018-19	2022-23
Primarily Informal	Corrected for labor input	1.85	2.29	2.13	1.89	1.77
	Corrected for labor input and human capital	0.83	0.99	0.88	1.14	1.05
Primarily Formal	Corrected for labor input	4.80	6.46	5.85	3.57	3.19
	Corrected for labor input and human capital	1.50	2.03	1.89	1.59	1.37

Notes: The wage data is from the employment surveys of NSSO. The Primarily Informal and Primarily Formal sectors are defined in the same manner as in Table 4. Sectoral wages are averaged using employment weights and weights for squared distance from the thresholds that define formal and informal sectors. The first and third rows compare daily wages –therefore they are adjusted for labour hours differences by construction. The wage gap with corrections for labor input and human capital are in the second and fourth row.

Table 6: Percentage of Non-Farm Employment in Industry Divisions whose APG is not statistically different from 1.

Year	Proportion of Non-Farm Employment
1999-00	73.46
2004-05	74.93
2011-12	68.68
2018-19	61.09
2022-23	66.74

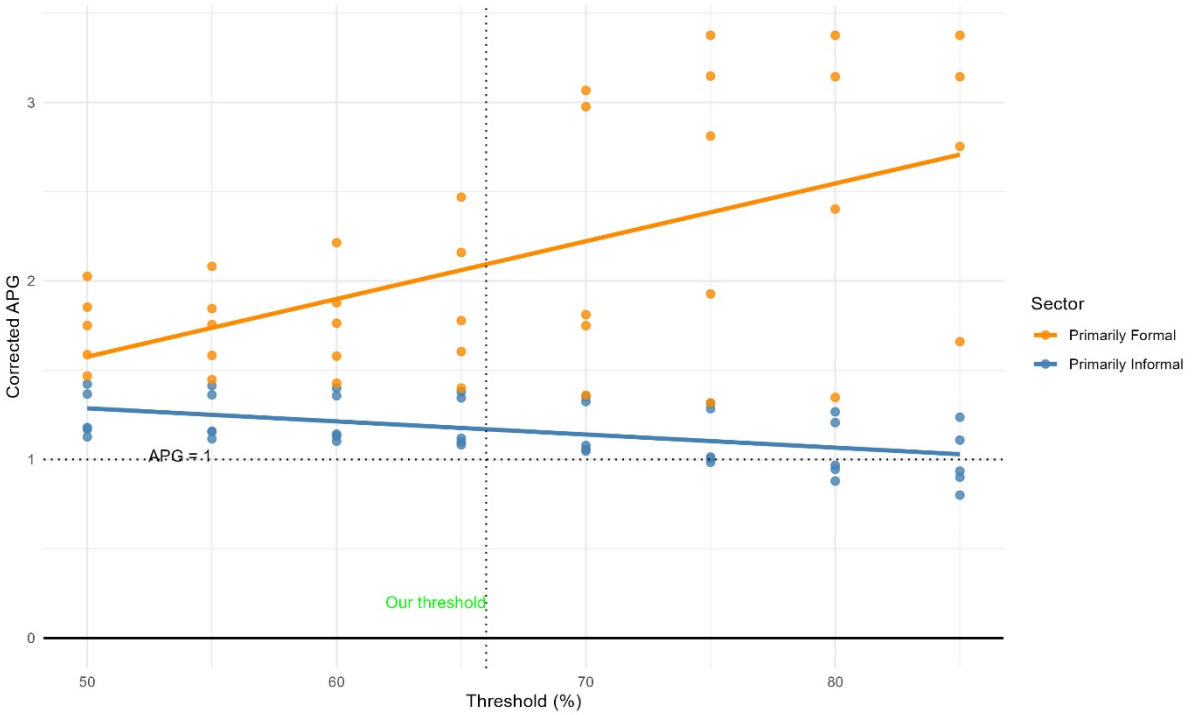
Notes: The industry divisions whose APG is not statistically different from 1 are identified in Figure 2 as those industries that have informal employment exceeding 76% of their total employment.

Table 7: Correlation Between Intermediate Input Per Worker and Informality

Intermediate Input per worker	Correlation Coefficient	p-value
Energy	-0.44	0.0000
Services	-0.57	0.0000
Material	-0.55	0.0000
Sum of Intermediate Inputs	-0.61	0.0000

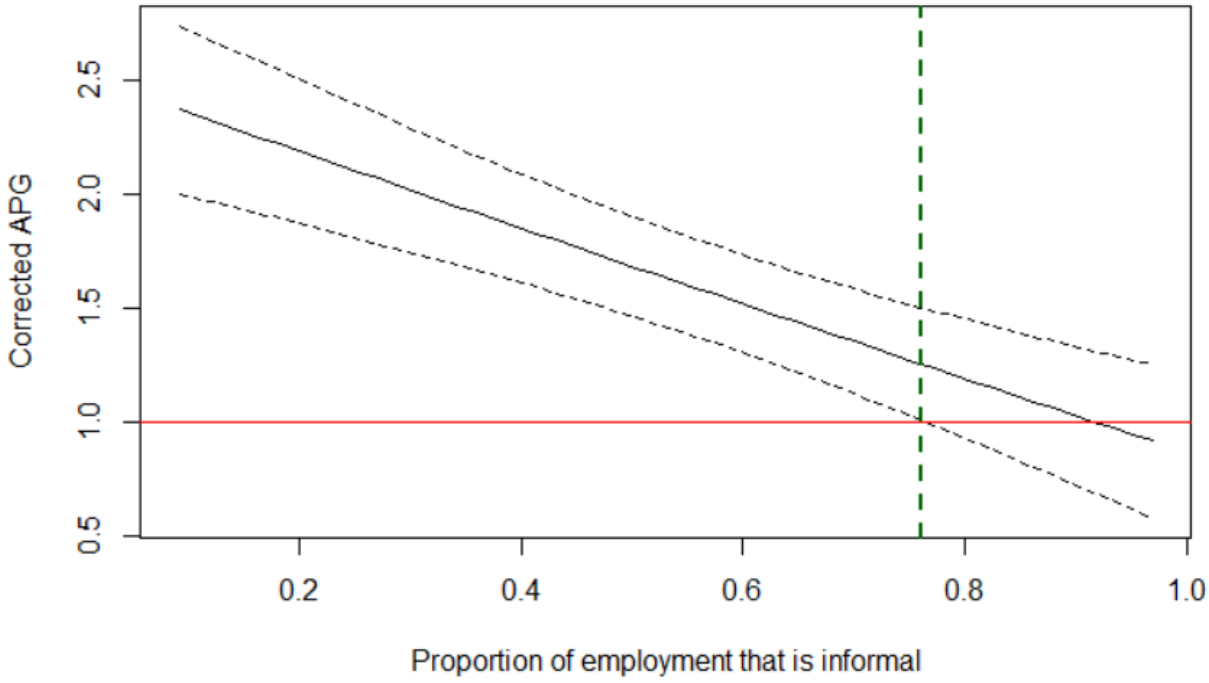
Notes: The intermediate inputs per worker use is computed in their value terms at constant prices (real terms) from pooled 24 sectors over 5 years of analysis (as in Figure 2).

Figure 1: Corrected APG for Non-farm Sector and Choice of Threshold



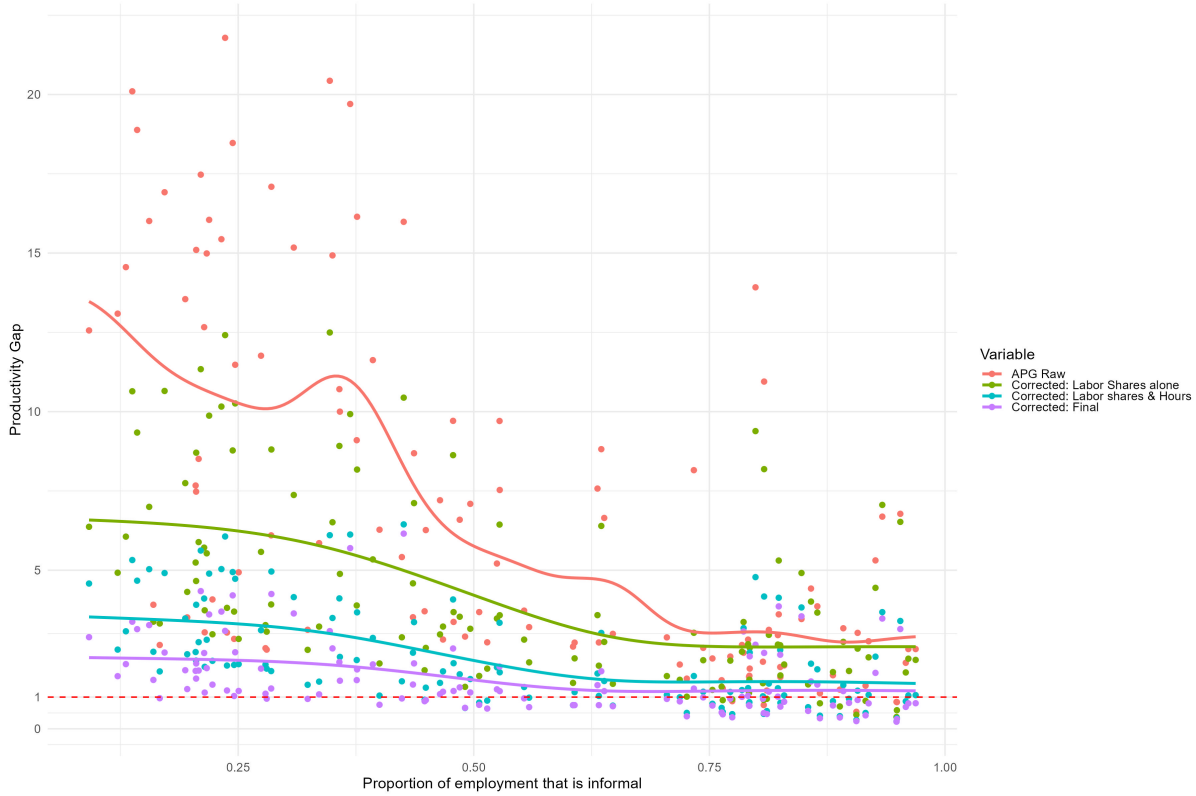
Notes: The corrected APG is computed in the same manner as the calculations in Table 4. The primarily formal sector consists of all industry divisions where the proportion of employment that is formal is greater than the threshold on the horizontal axis. The primarily informal sector consists of all industry divisions where the proportion of employment that is informal is greater than the threshold on the horizontal axis.

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



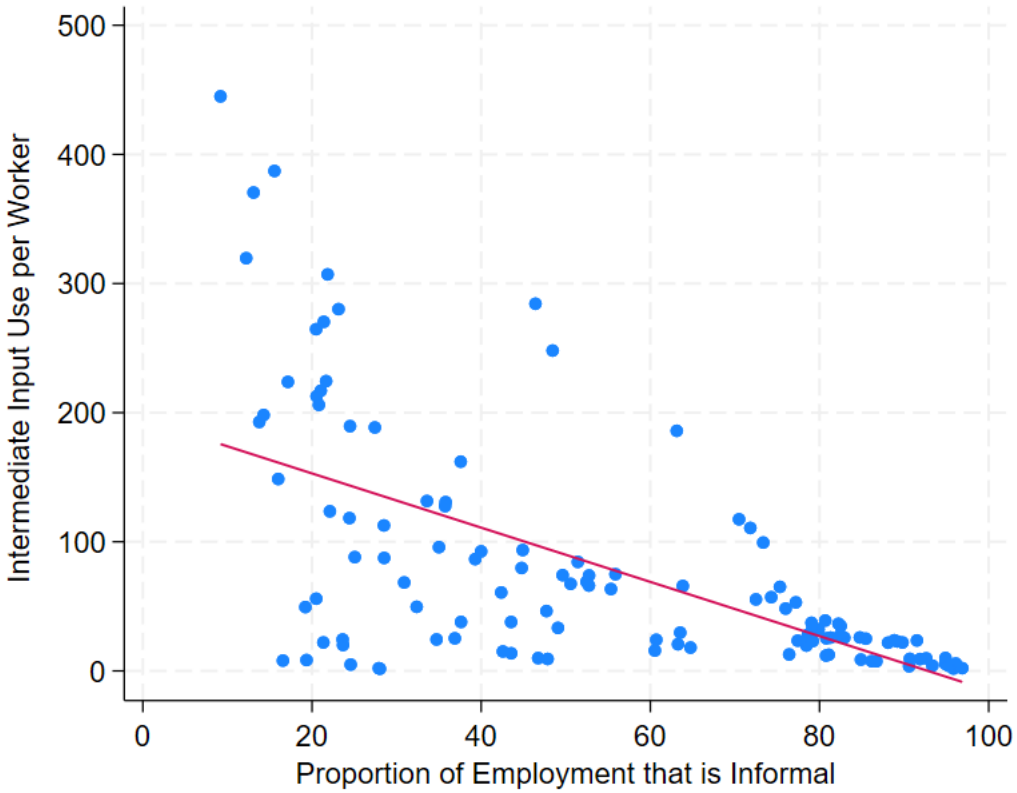
Notes: This is a graphical representation of a non-parametric regression of the industry division corrected APG on the proportion of industry division employment that is informal. 95% confidence bands are drawn.

Figure 3: The Role of Sectoral Differences in Correcting APG



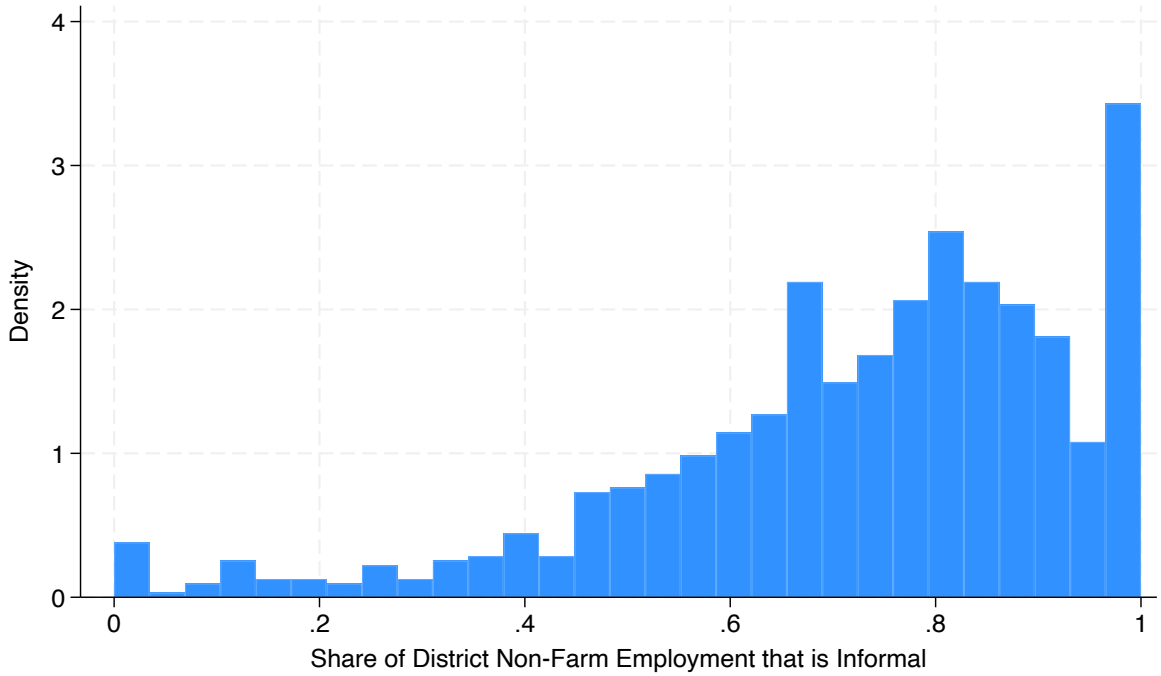
Notes: The scatter is the raw industry APG at different levels of the proportion of employment that is informal. The lines are non-parametric regressions of raw APG, APG corrected for labor shares alone, APG corrected for labor shares and hours worked and the fully corrected APG on the proportion of employment that is informal.

Figure 4. Intermediate Input Use per Worker and Informal Employment



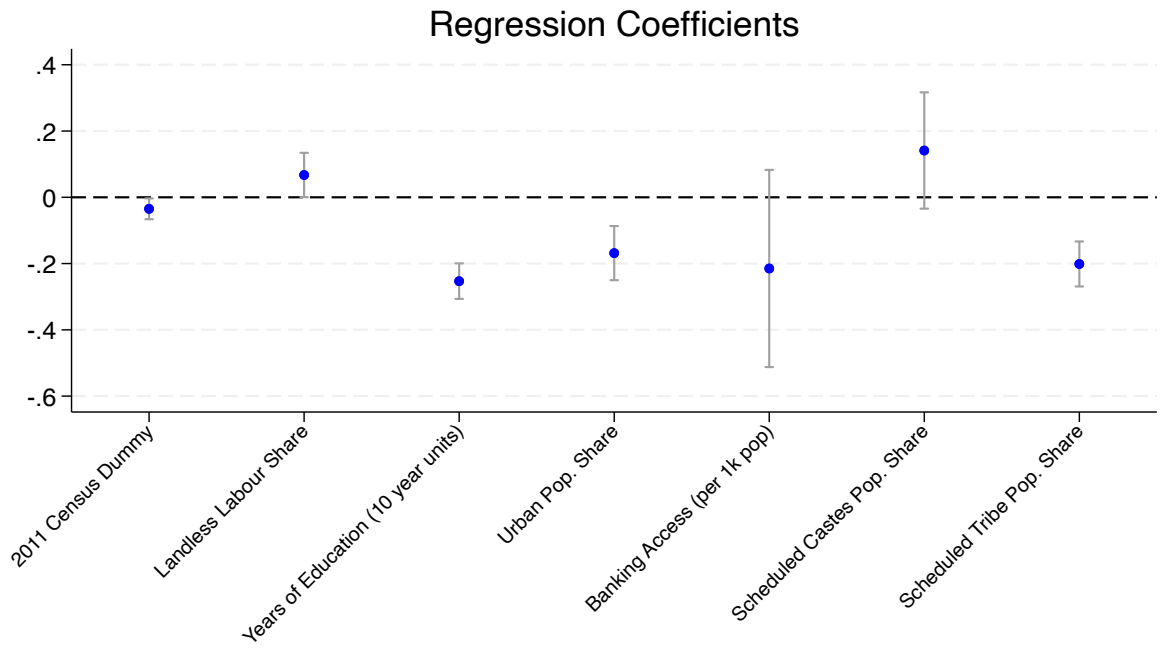
Note: The data is pooled for 1999/00, 2004/05, 2011/12, 2018/19 and 2022/23. Each scatter point represents a KLEMS industry division. The vertical axis measures all intermediate input use (energy, services, raw materials) in real terms of Rs. 10000 per annum. The horizontal axis is the proportion of employment in each industry division that is informal.

Figure 5: The District Level Distribution of the Share of Informal Employment in Non-Farm Employment, 1999/00 and 2011/12



Notes: The variable is computed from Employment Unemployment Surveys of the National Sample Survey Organization

Figure 6: Predictors of the District Level Share of Informal Employment in Non-Farm Employment



Notes: The dependent variable is the district level share of non-farm employment that is informal for 1999/00 and 2011/12. The variable `Education' measures the district level average years of education of the workforce (normalized in 10 year units), SC share is the share of scheduled castes in total population, ST share is the share of scheduled tribes in total population, Agricultural Labor is the share of agricultural landless workers in agricultural workforce, Banking Access measures the per capita availability of commercial and commercial banks, Urban population is the share of urban population in total population.

Appendix

A1. Notions of Informal Employment

Internationally, informal employment has been seen to be characteristic of unincorporated enterprises that do not offer contractual benefits to its workers. Here we map informal employment as defined in this paper with these related notions of informal employment. The results are displayed in Table A2 for the year 2022/23. The results are similar for other years,

The row defines the denominator and the column the numerator. The first row says 100% of the informally employed are in enterprises with size < 10 (by definition) and have no social security (by definition), 98% have no written job contract, 77% work either at home or in the streets, 100% work in private households (as domestic help) or in enterprises that are proprietary or family partnerships and 94% are not eligible for paid leave.

From the second row, it can be seen that of those who have no social security, 95% are informally employed. The remainder 5% are considered to be part of the formal sector – principally because of the size criterion. A similar comment applies to the other rows. Thus While our definition substantially agrees with other notions of informal employment, it does not fully coincide with them.

Table A1: The Distribution of Informal Employment

	Informal employment definition used in the paper	Enterprise Size < 10 workers	No Social Security benefit (retirement, health)	No written job contract	No Office Outside of Home or Street	Enterprise is Proprietary, Family Partnership or Private Household	Not Eligible for paid leave
Informal employment definition used in the paper	100.00	100.00	100.00	98.40	76.58	100.00	94.11
Enterprise Size < 10 workers	95.19	100.00	90.71	89.02	72.97	95.64	84.87
No Social Security benefit (retirement, health)	73.83	78.55	100	94.21	54.79	85.58	90.27
No written job contract	70.96	75.43	91.93	100.00	53.82	83.30	87.31
No Office Outside of Home or Street	96.28	96.42	98.36	98.30	100.00	99.52	97.44
Enterprise is Proprietary, Family Partnership or Private Household	94.08	94.48	96.16	96.08	74.60	100.00	90.91
Not Eligible for paid leave	73.96	78.28	95.98	95.04	57.86	85.56	100.00

A2: Construction of Weighted Averages

(a) The informal sector magnitudes of value-added per worker, years of education and hours of work are computed as weighted (weights w_i) averages of the KLEMS industry divisions that constitute it. The weights w_i are defined as below.

Let L_i denote the employment in industry division *i*. Let x_i denote the proportion of L_i that is informal. And let τ denote the threshold level such that the industry division is said to be a part of informal sector if $x_i \geq \tau$.

Suppose $x_i \geq \tau$ for industry divisions 1, 2…n.

Define
$v_i = (\frac{L_i}{\sum_{i=1}^{n} L_i}) \frac{(x_i-\tau)^2}{\sum_{i=1}^{n}(x_i-\tau)^2}$, for $i = 1,2, \dots n$

and hence $w_i = v_i / \sum_{i=1}^{n} v_i$, for $i = 1,2, \dots n$

(b) The share of labor in value-added in the informal sector is computed as a weighted average (weights w_i^*) of the KLEMS industry divisions that constitute it. The weights w_i^* are defined as below.

Let V_i denote the value-added in industry division *i*. Following the notation in (a), define
$v_i^* = (\frac{V_i}{\sum_{i=1}^{n} V_i}) \frac{(x_i-\tau)^2}{\sum_{i=1}^{n}(x_i-\tau)^2}$, for $i = 1,2, \dots n$

and hence $w_i^* = v_i^* / \sum_{i=1}^{n} v_i^*$, for $i = 1,2, \dots n$

(c) Formal sector magnitudes are calculated in the same manner as in (a) and (b).

A3: Analysis with Interpolations

Value added per worker and Labour share in value added is available for all years through KLEMS data. These are not interpolated.

We linearly interpolate three variables: *Proportion of Employment Informal*, *Years of Education*, and *labor hours*.

1. Proportion of Informal Employment and Years of Education:

Suppose we have data on variable *X* for a year T_1 and T_2. Then for a year T (where $T_1 < T < T_2$), the interpolated value is

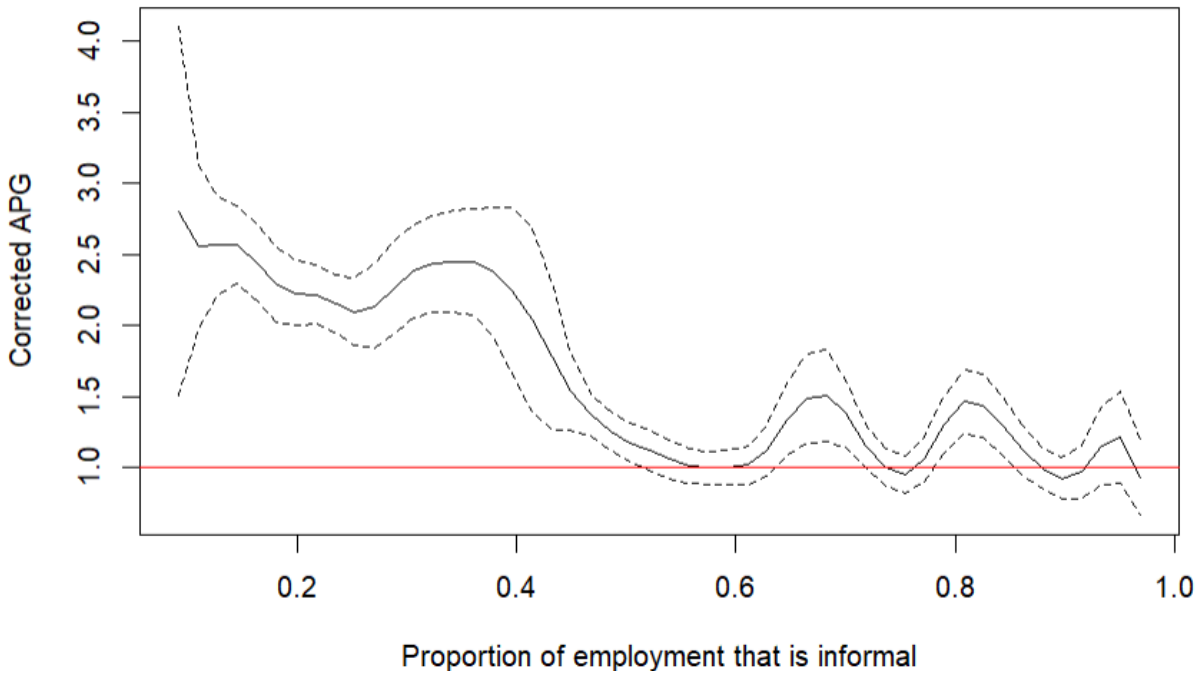
$$X_t = \frac{T_2 - T}{T_2 - T_1} X_{T_1} + \frac{T - T_1}{T_2 - T_1} X_{T_2}$$

2. Labor hours: For the labor hours there is no data available before 2004 and after 2012. Therefore:

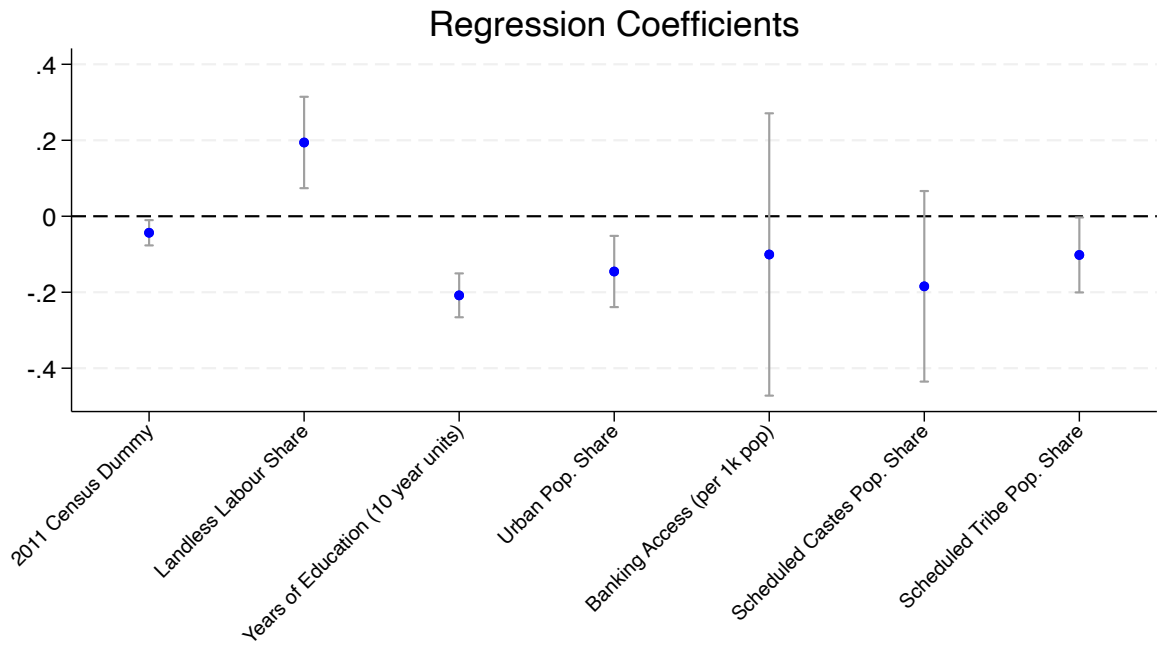
For years < 2004, labor hours= labor hours in 2004
years > 2012, labor hours = labor hours in 2012
2004 < years <2012, labor hours= Interpolated as discussed above

The non-parametric regression with interpolated data is represented below.

Figure A2: Agricultural Productivity Gap and Employment Proportion in Informal Sector (Interpolated all years: 600 datapoints)



A4: Correlates of Informal Employment: State Fixed Effects



Notes: The dependent variable is the district level share of non-farm employment that is informal for 1999/00 and 2011/12. The variable `Education' measures the district level average years of education of the workforce (normalized in 10 year units), SC share is the share of scheduled castes in total population, ST share is the share of scheduled tribes in total population, Agricultural Labor is the share of agricultural landless workers in agricultural workforce, Banking Access measures the per capita availability of commercial and commercial banks, Urban population is the share of urban population in total population. Besides these variables, the regression controls for state fixed effects.