Journal of Development Economics 168 (2024) 103270

Contents lists available at ScienceDirect

Journal of Development Economics

journal homepage: www.elsevier.com/locate/devec

Regular article

The agricultural wage gap within rural villages☆

Ceren Baysan[a,b], Manzoor H. Dar[c], Kyle Emerick[d,b], Zhimin Li[e,*], Elisabeth Sadoulet[f]

[a] University of Toronto, Canada
[b] CEPR, United Kingdom
[c] International Crops Research Institute for the Semi-Arid Tropics, India
[d] Tufts University, United States of America
[e] Peking University HSBC Business School, China
[f] University of California at Berkeley, United States of America

ARTICLE INFO

Keywords:
Wage gap
Agriculture
Structural change
Dual economy
Urbanization

ABSTRACT

We use unique data on daily labor-market outcomes for Indian casual workers to study labor reallocation between agricultural and non-agricultural activities within rural areas. Controlling for both individual time-invariant attributes and time-varying shocks, we find that workers who switch sectors across years or even within a week can obtain 23% higher wages by taking non-agricultural jobs. We then estimate a discrete choice model of daily labor allocation that decomposes preferences for jobs into two types of disamenities: (i) those associated with job characteristics and (ii) those associated with location. We find that the first type of disamenity is 23% of wages for men and 38% for women, and the second type is 36% of wages for men and 31% for women.

1. Introduction

Most of the poor live in rural areas and work in agriculture, earning lower wages than those working in the non-agricultural sector. Accordingly, theories and policies on economic development typically follow the dual-economy approach of explaining growth and are rooted in structural transformation: the shift of labor and other inputs from less productive activity, agriculture, to the more productive, non-agriculture (Fisher, 1939; Lewis, 1954; Clark, 1957; Kuznets, 1957; Johnston, 1970). While most literature on structural transformation focuses on the rural–urban productivity divide (Lagakos, 2020), the non-farm sector in rural areas has become an important source of employment in many low-income or lower middle-income countries (World Bank, 2017). This is especially true in India, where rural–urban migration is limited (Binswanger-Mkhize, 2013; Reddy et al., 2014). As of 2019, 66% of the Indian population was living in rural areas. In the same year, 45% of rural male workers were employed in non-agricultural work, while this number was only half as much in 1983 (Government of India, 2021).

In this study, we investigate sectoral employment transitions in rural Jharkhand, India. We do so by observing workers in the casual labor market who move between sectors within rural areas, and often within the same village. Using detailed panel data of daily labor market outcomes for these workers, we show that agricultural laborers can increase earnings by 23% when switching to non-agricultural work.[1] This is nearly the same magnitude as the urban–rural wage gap of 25% in India (Munshi and Rosenzweig, 2016) and comparable to findings from the broader literature on the rural–urban earnings gap (Lagakos et al., 2020).

We use a novel dataset to investigate which factors explain the sectoral wage gap in our setting. Our data allow us to control for both sorting on unobservable worker attributes and time-varying shocks in our analysis. In particular, the data generating process on income reflects the labor market structure unlike in other studies: we observe daily wages and labor supply choices for casual workers. This enables us to exploit the variation induced by workers changing sectors within a short time window of one to two weeks. The estimated

☆ We acknowledge financial support from the J-PAL & CEGA's Agricultural Technology Adoption Initiative and the Stress-Tolerant Rice for Africa and South Asia project of the CGIAR. Emerick is grateful to the Institute of Economic Development at Boston University where he was a visiting scholar while part of this research was carried out. Baysan acknowledges financial support from the NSF Graduate Research Fellowship program while part of this research was carried out.

* Corresponding author.
E-mail addresses: ceren.baysan@utoronto.ca (C. Baysan), m.dar@cgiar.org (M.H. Dar), kyle.emerick@tufts.edu (K. Emerick), zhimin.li@pku.edu.cn (Z. Li), esadoulet@berkeley.edu (E. Sadoulet).

[1] The data were collected as part of a randomized evaluation of the effects of a new drought-tolerant rice variety on labor markets. The technology was introduced in 2014 and we collected the six follow-up phone surveys during the planting and harvesting times for that season and the following two seasons.

https://doi.org/10.1016/j.jdeveco.2024.103270
Received 7 September 2023; Received in revised form 14 February 2024; Accepted 15 February 2024
Available online 17 February 2024

23% wage gap is based on our main regression in which we control for unobservable worker attributes with individual and survey fixed effects. Further, even if we use the variation from a smaller sample of workers switching sectors within one to two weeks for identification by including individual-by-survey fixed effects in our regression, we find that the estimated wage gap remains robust. Our estimate controls not only for individual time-invariant confounders but also components of unobserved ability that may vary over time. For example, a worker may gain additional skills between agricultural seasons, or they may migrate in response to a time-varying change in human capital. Though these changes may occur over several months, they are unlikely to happen over a period of just a few days. Our estimate also represents a specific phenomenon: switching sectors without migrating. This feature makes it more striking that an individual can earn much higher wages from casual non-farm labor without physically moving.

To shed light on possible explanations for the sectoral wage gap, we build on evidence based on a simple question we posed to workers. We asked them the top reason for working in agriculture if wages are a bit lower than in non-agricultural jobs. The two responses that stand out are being unable to find jobs that are close enough to home and the difficulty of non-farm work. To investigate the role of mobility constraints, we decompose the wage gap into the effect of switching sectors *within the same village* and the additional effect of leaving the village. We find that while part of the sectoral wage gap reflects mobility constraints, 78% of the wage gap remains even when focusing on within-village transitions.

We then consider the role of non-agricultural job attributes in explaining the remaining sectoral wage gap. Using year-to-year variation in rainfall, we show that agricultural productivity suffers and workers move to non-agricultural jobs when rain is deficient. An interpretation of these findings is that workers transition to non-farm jobs when forced to, which is consistent with rural non-agricultural work being less desirable. Non-agricultural jobs can be physically demanding in our context. They tend to involve construction, brick laying, and working in brick factories or coal mines. This observation matches our finding that workers from marginalized castes are more likely to engage in non-farm work and report lower disutility for it. In addition, we observe a strong positive correlation between working on one's own field and working in agriculture within one's own village, suggesting that there may be complementarities between working in agriculture and self-employment in agriculture. Altogether, we refer to attributes of non-agricultural work that decrease worker utility as non-agricultural job disamenities.

Finally, we estimate a discrete choice model of rural labor allocation. We use the model to quantify mobility-related disamenities, which are commonly explored sources of both the spatial and sectoral earnings gaps in the development literature, but within each sector. We use these disamenities as a benchmark to quantify the relative significance of the lesser studied non-agricultural disamenities within village. We are able to separate these two types of disamenities by observing people working in different sectors — both inside and outside their villages. The monetary value of mobility-related disamenities is estimated to equal about 75 rupees or 36% of the agricultural wage for male laborers. While previous work has shown the importance of moving and search frictions as barriers to internal migration (Bryan and Morten, 2019; Heise and Porzio, 2022), we find that mobility constraints also matter for the choice of whether to leave the village for casual labor even within small geographic space. Meanwhile, we show that other disamenities of non-agricultural work amount to 23% of the male wage. This result lines up with the self-reported explanation we got from workers that regardless of the location, rural non-agricultural work requires a compensating differential for the difficulty of the job. Additionally, we find that female laborers have a larger disutility from working in non-agricultural jobs compared with males.

Previous research has found urbanization to be an important source of the sectoral earnings gap (e.g., Lagakos et al., 2023; Bryan and Morten, 2019; Morten, 2019; Imbert and Papp, 2020; Baseler, 2023). On the other hand, several studies have shown that some of the large intersectoral gap in earnings can instead be explained by selection: more motivated and higher-ability people live in urban areas and work in the non-agricultural sector (Young, 2013; Herrendorf and Schoellman, 2018; Pulido and Swiecki, 2019; Alvarez, 2020; Hamory et al., 2021). In this case, no potential gains can be attained from switching sectors. With our data, we are able to control for both selection and time-varying shocks, and we find that the raw wage gap decreases from 38% to 23% after controlling for ability. Therefore, while sorting can explain part of the raw sectoral wage gap, it is less significant than that found in the urbanization literature, and the wage gap remains nearly as high as the rural–urban wage gap in India. Second, we find that the remaining sectoral wage gap is only partly (22%) attributed to mobility-related constraints. Instead, a lesser explored hypothesis in developing countries, albeit the difficult and precarious nature of casual labor in both rural and urban low-income settings, is whether the sectoral gap reflects compensating differentials (Smith, 1979; Duncan and Holmlund, 1983; Mas and Pallais, 2017). Under this explanation, sectoral differences in job attributes, and whether specific groups disproportionately face these differences, need to be better understood.

The remainder of the paper is organized as follows. Section 2 briefly describes our data and discusses the regression results showing that workers earn higher wages in non-agricultural jobs. It also shows survey evidence to explain sources of the wage gap. Section 3 outlines a model of daily labor allocation choices in the presence of mobility-related and non-agricultural job disamenities. It also presents evidence on how workers move between sectors in response to shocks. Section 4 estimates the parameters of the model. Finally, Section 5 provides concluding remarks and implications of our findings for sectoral gaps in developing countries.

2. Reduced-form estimates of the wage gap

2.1. Data and descriptive statistics

Our primary sample is spread across 12 blocks (administrative units below a district) within 4 districts of the Jharkhand state in eastern India. We identified blocks that were suitable for a drought-tolerant rice seed variety that was being field tested. We selected a random sample of 200 villages amongst those with 30 to 550 households. Within each village, enumerators located a village leader and asked for names of the 25 largest rice farmers and 10 agricultural laborers. Enumerators carried out a baseline survey with the farmers and workers during the period from late April to early June 2014.

Our sample of laborers consists of people who are landless or have small amounts of land. In contrast to large landowners, these workers generate most of their income from supplying labor to the agricultural casual labor market. This population makes up the majority share of those dependent on agriculture in rural India.

Table A1 probes the representativeness of our sample of laborers. We extract from the 2011–12 National Sample Survey (NSS) the set of casual agricultural laborers from Jharkhand. Comparing averages from the two, our sample closely resembles the representative NSS sample in terms of caste, gender, education, and total area cultivated. Thus, the laborers identified by village leaders do not underrepresent people on these key characteristics. However, our sample is different on some other dimensions. We overrepresent workers who are older, Hindu, and currently cultivating land. However, we show below that none of these differences matter for our results. Specifically, weighting our analysis to make it representative of the NSS sample in all of Jharkhand does not change the findings.

Hiring and wages in casual labor markets in India are generally determined on a daily basis. Yet, most studies rely on data that aggregate labor market outcomes over a longer period. This misses short-term movement between sectors. To better measure labor-market outcomes, we collected daily data on wages and employment by conducting phone surveys. These surveys took place during the planting (including plowing) and harvesting periods across the 2014, 2015, and 2016 seasons. Wet-season rice is the dominant crop in our sample area. Planting takes place in late July to early August, and harvesting in late November. Our phone surveys took place during these times to coincide with peak agricultural periods because lack of irrigation limits cultivation and agricultural employment during other times of the year.[2]

In the first two phone surveys in August and November 2014, we collected data on whether laborers worked on another person's or their own farm, the wage they received, whether the work took place in their own village, and their activity if they did not work in agriculture. We gathered this information for the seven days preceding the phone call. We repeated the same process in the 2015 and 2016 seasons with two major differences. First, we expanded the sample to include 6 female laborers per village, whereas the original sample contained only 3 female laborers per village. We selected the 3 added laborers from a census in all villages on households with casual laborers.[3] Second, starting with the 2015 harvesting survey, we expanded the recall window. We doubled the period to 14 days to capture the entire planting or harvesting period better. The phone surveys produced a high response rate: we reached an average of 86% of the workers from the baseline.[4] While we observe 7 or 14 days for each worker, these surveys completely cover the planting and harvesting periods, each of which lasts about 3–4 weeks depending on when farmers choose to plant or harvest. It is for those weeks that we document the choice between agricultural and non-agricultural casual work.

These data let us observe daily employment outcomes for planting and harvesting for all three years. We also collected non-agricultural wages in the 2015 planting and both 2016 surveys. Non-agricultural work predominantly consists of casual wage labor — rather than self employment. We observe the daily wage for 82% of the non-agricultural work days. These data, combined with the agricultural wages, enable us to estimate the sectoral wage gap while controlling for unobserved heterogeneity across individuals. The people switching sectors across survey rounds give identification in our main specification in which we include individual and survey round fixed effects. Furthermore, a smaller share of workers also switch across sectors within a survey round, allowing us to estimate a specification with individual-by-survey round fixed effects. Unlike longer term changes, switching sectors within one to two weeks is less likely to be correlated with a time-varying change in ability or training.

Table 1 shows average differences between workers who switch between sectors and those who stay in agriculture. About 20% of the workers from the baseline survey switched sectors. Switchers are more likely to be male and are poorer in several dimensions. For example, they are less likely to have access to electricity, more likely to be using the government's rural employment guarantee (MGNREGS), have larger households, and more likely to belong to lower castes or come from households with temporary migrants. Yet, switchers have no less land. The average laborer household cultivates 0.58 acres during the rainy season, with only about 18% of households cultivating no land at all.[5]

Table 1
Baseline characteristics.

	Ag only (N=1499)	Switchers (N=387)	p-value
Individual variables:			
Female	0.388	0.101	0.000***
Years of education	3.477	3.463	0.947
Cognitive ability	2.787	2.708	0.131
Household variables:			
Household size	5.932	6.214	0.052*
Access to electricity	0.512	0.453	0.038**
House has mud walls	0.674	0.739	0.015**
Number of rooms in house	3.571	3.708	0.169
Area cultivated (acres)	0.575	0.583	0.950
Landless	0.175	0.145	0.159
Has private tubewell	0.038	0.034	0.671
Owns mobile phone	0.933	0.912	0.149
BPL card holder	0.769	0.806	0.122
NREGS job card holder	0.749	0.796	0.053*
NREGS active user	0.193	0.240	0.041**
Scheduled Caste or Tribe	0.517	0.651	0.000***
Has loan	0.167	0.119	0.019**
Has savings account	0.685	0.628	0.032**
Has permanent migrant	0.097	0.098	0.931
Has temporary migrant	0.096	0.140	0.013**

Notes: The table shows average values of baseline characteristics between workers who worked only in agriculture for all three surveys that were used to estimate the agricultural wage gap (column 1) and those who worked in both sectors (column 2). Column 3 shows p-value of the t-test for equal means. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels. Cognitive ability is the score on a reverse digit span test. Active NREGS user is household that had NREGS income during April 2014, just before the baseline started. Has loan is an indicator for having any loan during the last 12 months. Permanent migrant is individual who is away for at least 10 months of the year. A temporary migrant is defined as an individual who leaves the village during the dry season but returns home during the wet season.

Fig. 1 further describes our data by showing a breakdown of daily activities. About 30% of the sample work only on their own farms. About 25% of workers do both agricultural wage labor and own-farm work, while another 25% only do agricultural wage labor. Non-agricultural work is predominantly done by male laborers, who are thus the main source of variation in our identification of the sectoral wage gap. Moreover, around 4%–8% of workers switch sectors during the same survey round. We show later that including only these workers for identification produces the same results as does including workers who switch across rounds.

We use three additional sources of data. First, we surveyed the 10 largest farmers after harvesting each year.[6] These data help us link rainfall-induced variation in agricultural output with labor flows to the non-farm sector. Second, since the gender division of our laborer sample does not represent the labor market, we have phone surveys with farmers where we collected the gender of hired laborers. Using these data, we compute gender-specific weights for our sample of laborers. Third, to measure weather, we use daily rainfall estimates from the Climate Hazards Group InfraRed Precipitation with Station dataset (CHIRPS) (Funk et al., 2015). CHIRPS incorporates 0.05° resolution satellite imagery with station-level data to create a gridded daily time series, which we use to create daily village-level precipitation. Figure A1 helps visualize these data. It shows that 2014 and 2015 – the first two years of our data collection – were dry years. The 2014 season had little rain past mid September. During 2015, almost no rain fell past the end of August. In contrast, 2016 was the wettest year since 2000. The importance of timely rainfall is highlighted by the productivity data

[2] Our baseline survey asked respondents for the number of days the main agricultural worker in the household worked during the previous season. The mean number of days is 33.6. Of these, 11.3 were reported to be during plowing, 9.5 during transplanting, and 7.7 during harvesting. We concentrated our surveys on the planting and harvesting periods because they are when most of the agricultural labor happens.

[3] We discovered after looking at our first year of data that our sample of laborers was under-representative of females based on their importance as agricultural workers. In addition to including more females to the sample, we make use of data on hiring from farmers to weight our worker data by gender. We do this to make our labor-market outcomes representative of an average agricultural worker. Section 2.2 provides details on the gender weights.

[4] The response rate ranged from 79% in the third-year planting survey to 91% in the second-year planting survey.

[5] The average cultivated area of the laborer households amounts to about 20% of the average cultivated area of the sample of large farmers.

[6] These farmers were selected amongst the 25 farmers listed at the beginning of the study.

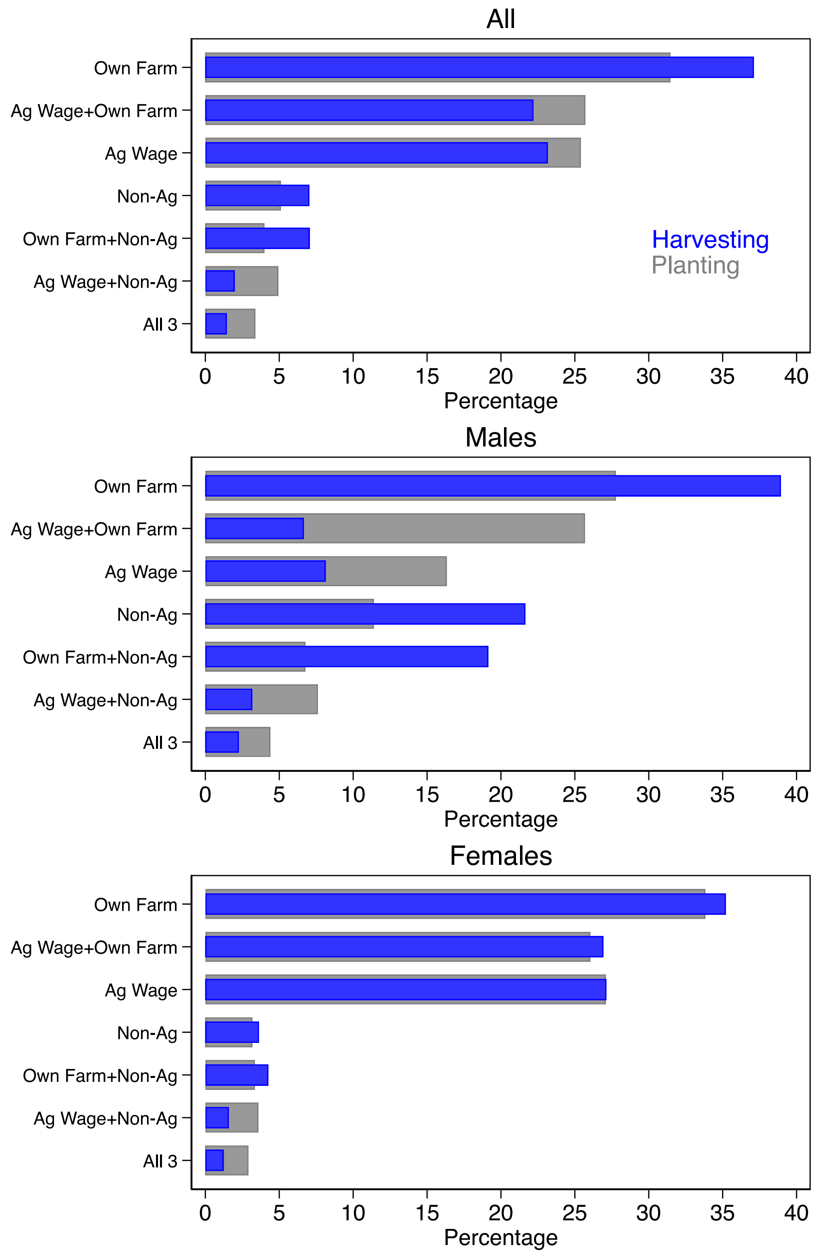

Fig. 1. Activities of workers during 7–14 day survey period.
Notes: The figure shows a classification of workers into seven groups, depending on which activities they did during the 7 or 14 day survey period. The top panel is for all respondents and is weighted by gender to represent the sex ratios of the population of agricultural workers hired by large farmers. The bottom two panels are separate for males and females. "Own farm" indicates working on their own farm, "ag wage" indicates working for a wage in agriculture, and "non-ag" indicates non-agricultural work. The gray bars denote percentages of respondents across the three planting surveys, while the blue bars denote the same values for the harvesting surveys. As an example, around 39 percent of the male respondents work only on their own fields during harvesting (top bar in the middle panel).

from farmers : relative to 2016, yields were lower by 25% in 2014 and 56% in 2015.

2.2. Empirical approach

The daily data permit us to estimate the wage gap between agricultural and non-agricultural work. To do so, we estimate,

$$\log(wage_{ivtd}) = \alpha_i + \gamma_t + \beta NonAg_{ivtd} + \varepsilon_{ivtd}, \tag{1}$$

where $wage_{ivtd}$ is the wage for worker i in village v during survey round t on day d, $NonAg_{ivtd}$ is an indicator for wage labor in the non-agricultural sector, α_i is an individual fixed effect, γ_t is a survey-round fixed effect, and ε_{ivtd} is an error term that we cluster at the village level. We limit the data for this estimation to the three survey rounds where we collected wages in both sectors. The parameter β measures the wage difference between sectors. The individual fixed effect eliminates time-invariant individual attributes. We also estimate the stricter specification with individual-by-survey round fixed effects. Doing so reduces the worry that time-varying unobservables, such as changes in skills or physical health, drive the result. Previous work on rural–urban migration has estimated sectoral wage gaps using people who switch sectors over longer time periods (Herrendorf and Schoellman, 2018; Pulido and Swiecki, 2019; Alvarez, 2020; Hamory et al., 2021). By contrast, our specification with the shorter time window allows us to estimate the wage gap within rural areas for jobs that can be taken within a period of just one to two weeks.

We weight the data in our main specifications to correct for the representativeness of female workers. This is because the phone surveys

Table 2
The agricultural wage gap amongst agricultural laborers.

	Individ, Survey (1)	Individ by Survey (2)	Survey (3)	Village, Survey (4)	Village by Survey (5)
Non-ag work	0.205***	0.211**	0.305***	0.325***	0.325***
	(0.042)	(0.083)	(0.040)	(0.036)	(0.035)
Mean ag wages (Rs per day)	169	169	169	169	169
No. of workers	2285	2285	2285	2285	2285
No. of observations	28 598	28 598	28 598	28 598	28 598
R squared	0.785	0.940	0.315	0.538	0.748

Notes: This table shows regression results from estimating equation (1). The data are from three surveys where non-agricultural wages were collected: the planting survey of 2015, and the planting and harvesting surveys of 2016. The dependent variable in all columns is the log of daily wages. Column 1 includes individual and survey fixed effects, column 2 includes individual-by-survey fixed effects, column 3 includes only survey fixed effects, column 4 includes village and survey fixed effects, and column 5 includes village-by-survey fixed effects. Columns 1 and 3–5 also include surveyor fixed effects. Observations are weighted by the respondent gender shares in the farmers survey. 472 respondents contribute to the identification in column 1 by working in both the agricultural and non-agricultural sectors across the three surveys. This number of switchers is larger than the number in Table 1 (387) because there are 85 switchers who were not part of the baseline. These 85 laborers were part of the additional sample added after year 1. 230 workers contribute to the identification in column 2, *i.e.* they work in both sectors in the same survey round. Standard errors (in parenthesis) are clustered at the village level in all specifications. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

with farmers show that 82% of the workers hired are females, a proportion that is larger than that we selected in our sample.[7] Specifically, we calculate the weight for female observations as the share of the hired workers who are female – across all our phone surveys with farmers – divided by the share of respondents from that survey round who were female. We define the weights in the same way for males. This weighting scheme ensures that our sample is representative of the average casual agricultural worker — although it does not affect our estimates.

2.3. Reduced-form results

Table 2 shows our estimates of the agricultural wage gap based on regression Eq. (1). Column 1 includes individual and survey-round fixed effects. We find that agricultural workers increase their daily wages by 23% when moving to non-agricultural work.[8] This estimate is partly identified off of people switching sectors across survey rounds.

One may worry that unobservables that vary across survey waves may affect our result. To address this concern, we also exploit variation from workers who switch sectors within a span of one to two weeks. This feature of the data allows us to include individual-by-survey round fixed effects, confining the identification to fewer individuals. For example, individuals could accumulate more skill over a period of months, but they are less likely to gain these skills in a short time period. As shown in column 2, we estimate the same wage gap of 23%. Therefore, time-varying unobservables across survey rounds do not appear to drive our estimate.

Columns 3–5 show the unadjusted agricultural wage gaps where we do not include individual fixed effects. Non-agricultural wages are higher by about 36% compared to agricultural wages — regardless of whether we use variation within or across villages. In our case, individual attributes explain only about a third of the wage gap. Unlike the literature on rural–urban migrants that finds that this type of selection accounts for most of the rural–urban wage gap, we find that much of the wage gap remains even when conditioning on individual fixed effects.

Moreover, we show that our results are robust to alternative weighting methods. First, none of the estimates in Table 2 change meaningfully if we omit the gender weights (Table A2). Second, we consider another approach where we weight to correct for observable differences between our sample and the NSS sample. Our point estimates are again unaffected when applying these weights (Table A3).

To put our estimate in context, Herrendorf and Schoellman (2018) use census data from 13 countries to show that non-agricultural wages are 80% higher than agricultural wages. Their estimate decreases to 33% when adjusting for education, gender, and spatial location. Our estimate focuses on the rural non-agricultural sector and eliminates the most cited cause of unobserved ability. The non-agricultural gap in our setting is about 23%, which is close to the rural–urban wage gap in India of 25% after adjusting for differences in costs of living (Munshi and Rosenzweig, 2016).

2.4. Understanding the sectoral wage gap

We draw on survey evidence to help understand sources of the sectoral wage gap. Our last survey posed a simple question to laborers: what is the top reason why you would continue to work in agriculture if non-agricultural wages are higher? The answers to this question provide suggestive evidence on what drives the worker's choice to work in agriculture, despite earning lower wages.

Fig. 2 shows that two explanations stand out. Over 32% say that non-agricultural jobs either are unavailable or require going too far from home. This suggests that mobility constraints in accessing non-agricultural jobs represent one set of factors. Consistent with this, recent evidence shows that rural roads facilitate the transition to non-agricultural work in rural India (Asher and Novosad, 2020). Thus, the difficulty of accessing jobs outside of one's own village offers an explanation for a persistent wage gap.

The figure points to another common explanation. Around 23% of workers select non-agricultural jobs being "too hard" as the reason for not taking them.[9] Although this finding does not pinpoint what makes these jobs harder, it provides suggestive evidence that preferences for job types constitute another reason that workers tend to choose agricultural work. Such preferences could arise because non-agricultural jobs are more physically demanding, require longer hours, are riskier, or involve tasks that are less familiar than agricultural activities.[10] Indeed, non-agricultural work in rural areas often requires laborious tasks. During this same survey we asked workers what they do when working in the non-farm sector. These jobs involve some form of construction around 68% of the time. Other popular activities include working in local coal mines or brick kilns. In addition, we find that marginalized groups report lower disutility for doing non-agricultural work: 28% of

[7] Part of the reason for this is that our phone surveys collected information during planting and harvesting — two activities more likely to be done by females. Males are more active during land preparation (plowing) and post-harvest activities like crop threshing.

[8] The precise calculation from the log wage regression is $e^{0.205} - 1 = 0.23$.

[9] Results in the online appendix (Figure A2) show that the share of the sample responding that non-agricultural work is too difficult is slightly higher among the group of switchers. This is inconsistent with an explanation where people not taking non-agricultural work misperceive its difficulty.

[10] The preference for agricultural work remains puzzling even if non-agricultural employers require longer days. It indicates that workers would prefer to earn less in a day in exchange for continuing to work in agriculture — even when they spend many other days without wage employment, *i.e.* working on their own very small farms or doing household chores. Our 2014 follow-up survey includes information on the length of the agricultural work day. Farmers report an average agricultural work day of 7.7 h for males and 7.5 h for females. Using variation in daily hours, Table A4 shows that daily wages are not positively correlated with the length of the working day. These data suggest that the relevant unit for wage determination is the day, rather than the hour.

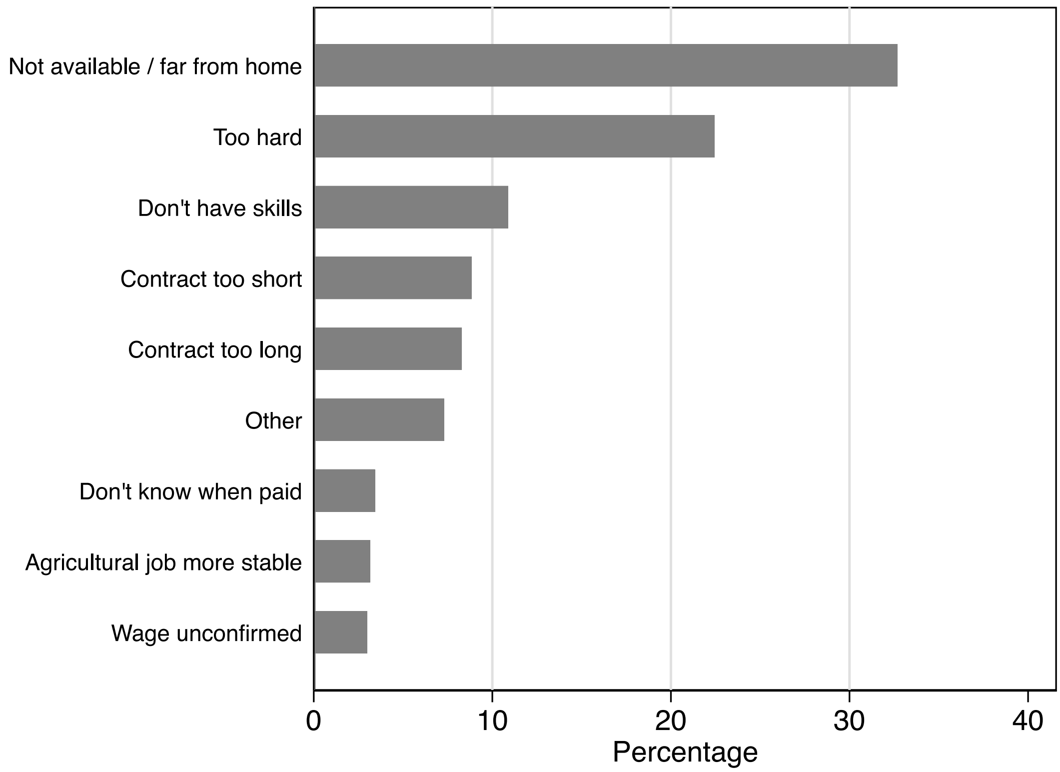

Fig. 2. Stated reasons why laborers still do not work in the non-agricultural sector even when wages are higher.
Notes: The figure shows responses from the third follow-up survey with agricultural laborers. The exact question posed to laborers was "Suppose wages are a bit lower for agricultural jobs than for non-agricultural jobs, what is the top reason why you may still work in agricultural jobs?". The bars represent the percentage of workers who state a particular reason shown on the vertical axis.

non-ST or SC workers cite the difficulty of non-farm work as the top reason in contrast to only 18% of ST or SC workers. These findings are also consistent with the possibility that the difficulty and negative features of non-farm work affect its social acceptability for certain groups (Oh, 2023). Relatedly, while female laborers are no more likely to report non-farm work as difficult than males, they are less likely to engage in it.

Our estimated wage gap reflect both mobility constraints to switching sectors and disamenities associated with job characteristics. We observe one type of mobility constraint directly: the spatial cost of leaving one's own village. Table 3 separates the effect of non-farm employment from that of working outside the village. Column 1 shows that, relative to agricultural jobs in one's own village, non-agricultural jobs in that same village yield 18% higher wages; changing sectors and working outside the village together increase wages by 45%.[11] The larger wage gains from leaving the village may capture the costs associated with transport or being unable to work flexibly on one's own farm. However, transitioning to non-agricultural work in the same village leads to higher wages, but does not require large search- or mobility-related costs.[12] These results suggest that the wage gap we estimate reflects more than merely mobility constraints.

In combination, multiple factors may drive the wage gap between sectors. But mobility constraints and the difficulty of non-farm work are two common explanations. While the literature emphasizes the important role that mobility constraints play in maintaining wage gaps across sectors and space (Heise and Porzio, 2022; Bryan and Morten, 2019), we provide evidence that sectoral wage gaps also exist within very narrow geographic areas and that job-specific attributes constitute an additional factor in explaining the sectoral wage gap.

Table 3
Breaking down the wage gap by location of the work.

	Individ, Survey (1)	Individ by Survey (2)	Survey (3)	Village, Survey (4)	Village by Survey (5)
Non-ag work own village	0.167***	0.190**	0.244***	0.265***	0.282***
	(0.046)	(0.086)	(0.043)	(0.038)	(0.039)
Non-ag work other village	0.372***	0.372***	0.520***	0.516***	0.452***
	(0.049)	(0.077)	(0.037)	(0.039)	(0.039)
Mean ag wages (Rs per day)	169	169	169	169	169
No. of workers	2285	2285	2285	2285	2285
No. of observations	28 598	28 598	28 598	28 598	28 598
R squared	0.787	0.941	0.325	0.546	0.751

Notes: This table shows wage gap estimates by the location of work. The data are from three surveys where non-agricultural wages were collected: the planting survey of 2015, and the planting and harvesting surveys of 2016. The dependent variable in all columns is the log of daily wages. Column 1 includes individual and survey fixed effects, column 2 includes individual-by-survey fixed effects, column 3 includes only survey fixed effects, column 4 includes village and survey fixed effects, and column 5 includes village-by-survey fixed effects. Columns 1 and 3-5 also include surveyor fixed effects. Observations are weighted by the respondent gender shares in the farmers survey. 472 respondents contribute to the identification in column 1 by working in both the agricultural and non-agricultural sectors across the three surveys. 230 workers contribute to the identification in column 2, *i.e.* they work in both sectors in the same survey round. Standard errors (in parenthesis) are clustered at the village level in all specifications. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

[11] The calculation is as follows: $e^{0.167} - 1 = 0.18$ and $e^{0.372} - 1 = 0.45$.

[12] We do not find much evidence of wage dispersion in our data. Jeong (2021) finds a causal link between search costs and higher levels of wage dispersion. Figure A3 shows that within villages, wage dispersion is small and similar for the two sectors. Once controlling for gender, the 90/10 ratio is 1.34 for agriculture and 1.42 for non-agriculture.

3. A model of rural labor allocation

Our analysis up to this point suggests there are two key constraints to doing non-agricultural work. First, it can require transportation outside the village. Second, there are job disamenities — besides perhaps being located outside the village — that make it less desirable. In other words, workers have a disutility for characteristics of *rural* non-agricultural work, which we refer to as non-agricultural disamenities. Our survey evidence suggests that the physically demanding nature of jobs such as construction and brick laying is one of these disamenities. In this section, we formulate a model that quantifies mobility-related disamenities, which are commonly explored sources of both the spatial and sectoral earnings gaps in the development literature, but within each sector. We use these disamenities as a benchmark for understanding the magnitude of the lesser explored non-agricultural disamenities, which we quantify within villages.

The two types of disamenities have different implications for candidate policy responses in increasing earnings for the rural poor. For instance, some policies could reduce mobility-related constraints for accessing jobs outside the village. These policies will do little, however, if non-agricultural disamenities explain why workers stay in agriculture.

To assess the relative importance of the two types of disamenities, we estimate a discrete choice model of rural labor allocation. In the model, workers choose between the two sectors, and within each sector, they choose whether to leave the village. We empirically separate disamenities associated with non-agricultural work and those associated with working outside the village.

3.1. A discrete choice model of rural labor allocation

Our model has a full set of occupational choices, which consist of agriculture inside the village, agriculture outside the village, non-agriculture inside the village, non-agriculture outside the village, working in one's own field, and not working (being unemployed). This breakdown enables us to estimate disamenities associated with non-agricultural work and with working outside the village separately. Non-agricultural disamenities limit movement between job types. We allow these disamenities to be a function of worker characteristics as well as randomly distributed across workers. Spatial disamenities limit movement from inside to outside the village.

An ideal dataset would contain random wage offers for both sectors. In practice, we only observe wages for the chosen options, but not the (counterfactual) wage offers for unchosen ones. To address this issue, we start from the premise that both job opportunities and wages in rural areas depend on timing in the agricultural season (planting vs. harvesting) and the weather.[13] In addition to these variables related to labor opportunity, we allow occupational choices to depend on past choices to account for potential path dependency or job switching costs. We also allow idiosyncratic preferences for alternative options, which depend on workers' characteristics and include an unobserved random element.

Let i denote the worker and j the employment choice that consists of agriculture inside the village, agriculture outside the village, non-agriculture inside the village, non-agriculture outside the village, working in one's own field, and unemployment. Further, let b denote block, d day, and t survey round. Worker i's utility, U_{ijdt}, is represented as follows:

$$\begin{aligned} U_{ijtd} &= V_{ijtd} + \epsilon_{ijtd} \\ &= \alpha_{j0} Harvest_t + \alpha_{j1} W_{bt} Harvest_t + \alpha_{j2} W_{bt} Plant_t + I_j \delta_i \\ &\quad + X_i \beta_j + \gamma y_{ijtd-1} + \epsilon_{ijtd}, \\ & j \in \{\text{ag inside, ag outside, nonag inside, nonag outside,} \\ & \quad \text{own field, unemployment}\}, \end{aligned}$$

where W_{bt} is cumulative rainfall during the growing season (at the block-survey level), and $Harvest_t$ and $Planting_t$ are indicators for whether the survey round occurs in the harvesting or planting season; these variables are meant to control for important determinants of job availability.[14] X_i is a vector of farmer characteristics such as gender and whether the worker belongs to a marginalized caste group to characterize how individuals differently perceive those amenities based on gender and caste. y_{ijtd-1} is an indicator variable for whether option j is chosen on previous day to account for possible switching costs. I_j is an indicator vector for each employment choice, and δ_i is a vector of random utility terms, taking on a multivariate normal distribution $\delta_i \sim N(\mu, \Sigma)$, with mean μ and variance–covariance Σ, which captures idiosyncratic heterogeneity of individual preferences for the different jobs (coming from either differential perception of its amenities or differential comparative advantages and wages). We will discuss in Section 4 our interpretation of the term δ_i as measuring perception of amenities in light of the empirical results. Finally, ϵ_{ijtd} is a random component that is assumed to follow a Type-I Extreme Value distribution. Note that the random utility terms can be potentially correlated across employment options, allowing for flexible patterns of substitution.

3.2. Data for model estimation

Our data describe the daily activities of casual workers from 12 blocks over 3 years. There are two distinct seasons per year, planting and harvesting. The data contain two panel dimensions, as we observe workers 7–14 days in each season, and we have 6 different seasons. For the estimation, we consider each block as a separate labor market in each of the 6 seasons. We define rainfall during the planting season as total precipitation for the months of June through July, and we use the months of June through October for the harvesting season. This definition reflects how the quality of the harvest depends on the total rainfall during the growing period. These variables are measured at the block level and standardized in the analysis to ease interpretation (so a one-unit change represents a one standard-deviation change from the mean rainfall). We take unemployment as the reference option.

Source of variation and identification. In the model, the random component in the utility function affects the daily employment status of a worker and can be interpreted as random variation in the offered wage. We capture variation across seasons with the seasonal and rainfall variables, and we capture preference variation with individual characteristics and random preference shocks. The model also incorporates variation across years in the same season with the rainfall variable and variation in past choices.

The identification strategy relies on cumulative rainfall being a good proxy of unobserved wages. We present evidence that there is a strong relationship between rainfall and occupational choice in our setting. Our data show that workers turn to non-agricultural jobs in years when agricultural work is less available. Because agricultural labor demand at harvesting depends on rainfall earlier during the growing season, rainfall provides a quasi-random source of variation in agricultural labor demand. To illustrate, we focus on the three harvesting surveys and estimate

$$y_{ivtd} = \alpha_v + \gamma_t + \beta Rainfall_{vt} + \epsilon_{ivtd}, \tag{2}$$

[13] An advantage of this approach is that although observed wage offers are endogenous, conditioning on weather enables us to extract an exogenous component of the wage variation.

[14] Harvesting takes place in November or early December. Therefore, we calculate cumulative rainfall from June through October as proxy for shocks to agricultural labor demand. We consider each block as a separate labor market in the analysis, so we use block-level rainfall to capture these shocks.

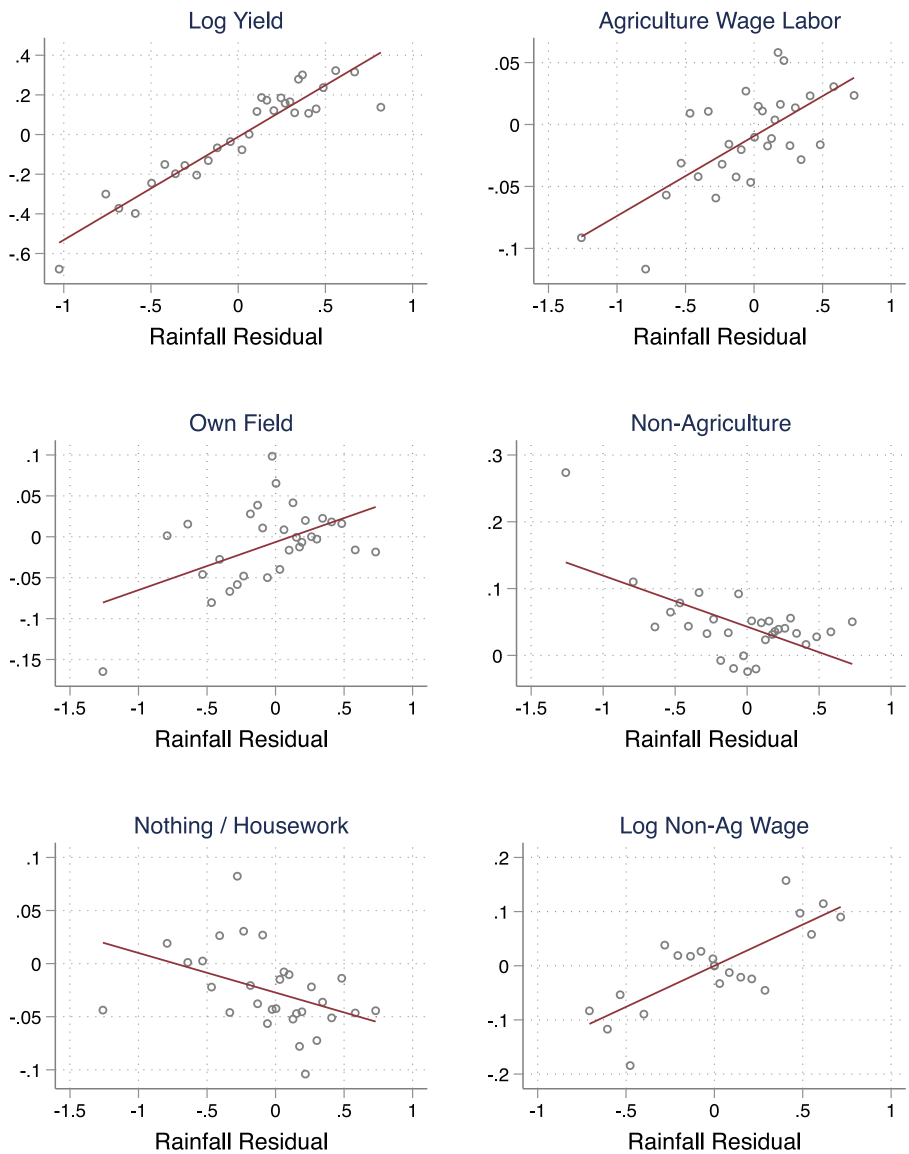

Fig. 3. The relationships between rainfall realizations, agricultural productivity, and labor allocation.
Notes: The figure shows binned scatter plots of various outcomes against rainfall realizations. The data are first residualized by regressing the outcomes and June-October rainfall on surveyor, survey-round, and village fixed effects. Each graph then shows the partial relationship between the outcome and rainfall. The dots are for 30 bins of the rainfall residuals, with equal numbers of observations per bin. The regression line is shown in red. The upper left graph uses the 3-year panel survey with farmers to plot the relationship between rainfall and log rice yield. With the exception of non-agricultural wages (lower right), the remaining outcome variables are from the labor allocation survey with agricultural workers. The outcomes are an indicator for working in agriculture as a wage laborer (upper right), an indicator for doing own-farm work (middle left), an indicator for non-agricultural work (middle right), and an indicator for staying at home or doing housework (lower left), all measured at time of harvesting. The log of non-agricultural wages (lower right) comes from the year 1 follow-up survey and the year 3 phone survey, the only two periods where we observe non-agricultural wages during harvesting.

where the dependent variable is one of four indicator variables for worker i in village v to work as an agricultural wage laborer, on own field, in the non-agricultural sector, or not work/do housework on day d of survey round t; α_v and γ_t denote village and survey-round fixed effects. We run a similar specification for estimating rice yield for farmer i in village v in survey round t. We use total precipitation during the agricultural season for the rainfall variable.[15]

One might be concerned that rainfall have direct effects on non-farm labor supply if work can only be done on dry days. For one, it might be uncomfortable to work in the rain. Or high rainfall might limit mobility. However, these direct effects of current rainfall on labor allocation are unlikely to drive the estimate of β in Eq. (2). Our harvesting surveys took place well after monsoon rains had stopped. Moreover, Eq. (2) uses cumulative rainfall variation that happened weeks before our surveys. Figure A1 makes this point clear. It shows that our surveys, which took place in late November to early December, happened during periods of no rainfall.

Fig. 3 visualizes the results from estimating Eq. (2). The figure shows binned scatter plots of different outcome variables against rainfall — after residualizing the data to remove fixed effects. The upper-left panel of the figure shows a tight positive association between total precipitation and rice yield. Going from the driest to the wettest observations causes yield to more than double. Thus, agricultural productivity increases with rainfall. The next four panels show how the

[15] While village-level rainfall is used in the regression Eq. (2), there is little difference in the variation from that across blocks (the R-squared of regressing rainfall on block-survey dummies is 0.98). We obtain similar results when using block-level rainfall instead (Table A5).

Table 4
Effects of rainfall realizations on agricultural productivity and employment choices.

		Daily activity				
	(1) Log yield	(2) Ag	(3) Own field	(4) Non-Ag	(5) Nothing/House	(6) Log non-Ag wage
Rainfall	0.520***	0.071***	0.036*	−0.051***	−0.045***	0.152**
June–October	(0.050)	(0.016)	(0.021)	(0.013)	(0.016)	(0.068)
Village fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Survey fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Mean outcome	0.36	0.22	0.37	0.16	0.24	237.84
No. of laborers		2645	2645	2645	2645	
No. of observations	5898	78 449	78 449	78 449	78 449	294
R squared	0.463	0.241	0.140	0.170	0.162	0.698

Notes: The estimates in column 1 are based on a 3-year panel survey with 2,000 large farmers (10 per village). The dependent variable in column 1 is the log of overall rice yield (across all plots). Columns 2–5 are estimated for the harvesting surveys with agricultural laborers of 2014, 2015, and 2016. The dependent variables are an indicator for working in agriculture as a wage laborer (column 2), an indicator for working on one's own field (column 3), an indicator for working in the non-agricultural sector (column 4), an indicator for not working or doing housework (column 5), and log non-agricultural wage (column 6). The estimates in column 6 come from village level data on non-agricultural wages during harvesting time. The year 2 data are from our followup survey, and the year 3 data are from the phone survey with workers. The rainfall variable is total rainfall (measured in 100's of mm from June–October). Observations in columns 2–5 are weighted by the respondent gender shares in the farmers survey. These regressions also include surveyor fixed effects. Standard errors (in parenthesis) are clustered at the village level in all specifications. Asterisks indicate a coefficient that is statistically significant at the 1% ***, 5% **, and 10% * levels.

time allocation of casual laborers *at the time of harvesting* responds to these rainfall shocks. Dry years lead to decreases in agricultural work and increases in non-agricultural work. But the increase in non-agricultural work offsets only part of the fall in agricultural labor: workers are more likely to report "doing nothing" or carrying out housework with low rainfall, as shown in the bottom left panel of the graph.

The bottom right panel of the figure combines different data to look at non-agricultural wages.[16] Non-agricultural wages increase with rainfall, which is consistent with workers supplying less labor to the non-agricultural sector when agricultural labor demand is high. Conversely, workers shift into non-agriculture during drier years, and this pushes wages down.

Table 4 provides the parameter estimates. Notably, a *decrease* in rainfall by 100 mm lowers agricultural work by 10.7 percentage points, with most of this effect coming from wage labor (columns 2 and 3). Non-agricultural work, on the other hand, increases by 5.1 percentage points (32%). Column 5 shows that the remainder of the displaced workers do not find employment or end up doing household work.

Rain may cause people to delay non-agricultural labor demand to the future. This provides an alternative explanation where growing season rain has an impact through labor demand and not supply. The online appendix tests this alternative. We show effects of many lags of daily rainfall on labor allocation during planting (Figure A4). We find no evidence that non-agricultural labor increases with previous lags of rainfall. Instead, non-agricultural labor *decreases* around 3–5 weeks after heavy rainfall. Agricultural labor increases during this period. This is consistent with how farmers plant rice. They prepare seedbeds when rains arrive and transplant the seedlings around a month later.

4. Model estimation and quantification of disamenities

4.1. Model estimation results

Table 5 reports the estimated parameters of the discrete choice model. The columns report coefficients for agricultural work inside the village, agricultural work outside the village, non-agricultural work inside the village, non-agricultural work outside the village, and own farm work. Unemployment (not working or doing housework) is the reference option. The model estimates the means, standard deviations, and correlation matrix of the random utilities associated with different job options. We allow the mean utility of each job type to depend on gender and caste.

[16] We use the follow-up survey after the first year and the harvesting phone survey after the third year to form a village-level panel.

The results provide evidence for disamenities associated with non-agricultural work and location. For example, within villages, workers have a much higher disutility for non-agricultural work than for agricultural work. This disutility for non-agricultural work outside the village is somewhat smaller for workers from lower castes (ST or SC). This is consistent with the self-reported survey evidence on why workers choose agricultural work over non-agricultural work even if earnings are lower: the proportion of non-ST or SC workers who cite the difficulty of non-farm work as the top reason (28%) is 10 percentage point higher than that of ST or SC workers (18%). Workers also have a higher disutility for leaving the village, particularly in agriculture.[17] This is consistent with the strong correlation between preferences for working on their own field and doing agricultural work in the village, as shown in column 5. The correlation suggests that there are complementarities between these two choices. Finally, compared with males, females have a greater disutility for non-agricultural work, particularly when done outside the village. This suggests that non-agricultural disamenities may partly explain the gender earnings gap.

As mentioned above, the term δ_i is the sum of comparative advantage and amenities for working in a particular sector relative to unemployment. Since those numbers are negative (except for working in one's own field), it has to be the case that they are dominated by disamenities. And if the wage premium that workers obtain in the non-agricultural sector were partly due to higher productivity in that sector, then their disamenities would have to be even higher than what we estimate.

To support our interpretation of absence of comparative advantage, we report the wage distributions in the two sectors, separately for the specialized workers and the switchers, in Figure A5. It shows that switchers and specialized workers have similar wage distributions in the non-agricultural sector, suggesting no differential productivity across these two groups. In the agricultural sector, while switchers have a wider distribution of wages with both some lower and some higher wages than the specialized workers, as a group they do not exhibit any differential average productivity, either.

[17] This argument requires workers to be taking similar types of jobs when leaving their village. For agriculture, the standard tasks are plowing, transplanting, harvesting, and threshing, all for rice. These tasks vary little across space. For non-agriculture, we only have detailed data on tasks in our last in-person survey. Those data show that the proportion of working days for different tasks is similar for work inside and outside the village. For example, 53% of non-agricultural days inside the village are for just three tasks: plastering, fencing, and brick work. This figure is similar at 50.5% for working outside the village. The similarity of tasks by location helps rule out that the location disamenity is compensation for different types of work.

Table 5
Model estimation of preference parameters.

	Agriculture		Non-agriculture		Own field
	(1) Inside	(2) Outside	(3) Inside	(4) Outside	(5)
Utility (μ)	−0.323***	−3.878***	−2.543***	−3.466***	0.487***
	(0.044)	(0.153)	(0.095)	(0.115)	(0.035)
SD of utility (Σ)	1.042***	2.530***	1.835***	2.261***	0.814***
	(0.025)	(0.086)	(0.051)	(0.068)	(0.019)
Corr with Ag inside (Σ)		0.124***	0.148***	0.085***	0.322***
		(0.042)	(0.039)	(0.031)	(0.028)
Corr with Ag outside(Σ)			0.077**	−0.050	0.019
			(0.038)	(0.036)	(0.034)
Corr with non-Ag inside(Σ)				0.209***	0.153***
				(0.030)	(0.034)
Corr with non-Ag outside(Σ)					0.090***
					(0.032)
Harvest (α_0)	−0.425***	−0.544***	−0.476***	−0.459***	0.132***
	(0.025)	(0.052)	(0.051)	(0.061)	(0.021)
Rainfall harvest (α_1)	1.097***	1.863***	−1.190***	−1.569***	0.634***
	(0.021)	(0.046)	(0.044)	(0.052)	(0.017)
Rainfall planting(α_2)	0.099***	−0.396***	−0.435***	−0.693***	0.097***
	(0.016)	(0.037)	(0.033)	(0.038)	(0.015)
Female (β_1)	0.892***	0.733***	−1.421***	−2.541***	−0.131***
	(0.056)	(0.180)	(0.127)	(0.187)	(0.047)
ST or SC (β_2)	0.015	−0.048	0.064	0.350***	−0.111***
	(0.051)	(0.156)	(0.107)	(0.128)	(0.042)
Same last choice (γ)	1.074***				
	(0.011)				
Share of work days in data	0.209	0.036	0.054	0.049	0.400
Shares of work days in model	0.211	0.038	0.055	0.049	0.395

Notes: The table shows coefficients results from the estimation of the discrete choice model. Columns 1–5 report estimated coefficients corresponding to each employment options: agriculture inside the village, agriculture outside the village, non-agriculture inside the village, non-agriculture outside the village, and working on own field. The unemployment option (not working or doing housework) is used as the reference category. The last two rows show the shares of work days in each employment category in the data and as predicted by the model. Standard errors are in parentheses. Asterisks indicate a coefficient that is statistically significant at the 1% ***, 5% **, and 10% * levels.

4.2. Quantifying disamenities

We next quantify the disamenities associated with non-agricultural work and location based on the model estimation. However, a challenge is that we do not have data for random wage offers to estimate a direct measure of the marginal utility of money, and we thus cannot directly convert estimated disamenities into monetary terms. Instead, we quantify them using two different approaches. First, we use quasi-random variation in wages created by rainfall. We then convert the parameter estimates to monetary terms in a way that resembles computing equivalent variation. We refer to this method as a revealed preference approach. Second, we use a stated preference approach. One of our worker surveys included hypothetical wage offers to trace out the labor supply curve. Using these data, we compute the equivalent increase in daily wages that would have the same effect on labor supply as those of the disamenities.

Revealed preference approach. We measure the average relative preference for choice j over choice k (conditional on weather and past choice) by $\mu_j - \mu_k$. We start in agriculture inside the village and ask what would be the change in rainfall that would have the same welfare effect as moving to sector j?. The rainfall equivalent is then

$$\Delta W_j = \left| \frac{\mu_j - \mu_{\text{ag inside}}}{\alpha_{\text{ag inside}}} \right|.$$

This computes the equivalent change in rainfall if the worker had stayed working in agriculture inside the village.[18] To compute the rainfall equivalent for the role of disamenities associated with leaving the village, we let j be agriculture outside the village, and for disamenities associated with non-agricultural work, we let j be non-agriculture inside the village. To then convert rainfall equivalents into monetary terms, we use the agricultural wage regression:

Table 6
Regression of wages on rainfall.

	(1) Male workers	(2) Male workers	(3) Female workers	(4) Female workers
Rainfall	23.72***	23.05***	12.93***	13.24***
	(3.53)	(3.53)	(3.93)	(3.89)
Block fixed effects	Yes	Yes	Yes	Yes
Survey fixed effects	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
No. of observations	16 584	16 584	10 048	10 048
R squared	0.187	0.188	0.244	0.246

Notes: The table shows results by regressing wages on standardized rainfall in the harvest season. The coefficients correspond to the change in wage if rainfall is increased by a one SD of rainfall. Columns 1–2 show regression results for male workers, and columns 3–4 are for female workers. The regressions in columns 2 and 4 also control for an indicator variable of ST or SC castes. Robust standard errors are in parentheses. Asterisks indicate a coefficient that is statistically significant at the 1% ***, 5% **, and 10% * levels.

$$wage_{ibtd} = \lambda_b + \lambda_t + \theta W_{bt} + v_{itd}, \tag{3}$$

where λ_b and λ_t denote block and survey-round fixed effects. Table 6 shows regression results from estimating Eq. (3) separately for male and female workers. For male workers, a one standard-deviation increase in rainfall raises agricultural wage by 23 rupees; for female workers, the corresponding increase is 13 rupees (using columns 2 and 4 as preferred specifications).

[18] We use the coefficients of the harvest season in the conversion to rainfall equivalents because rainfall variation is a stronger predictor of the wage in the wage-weather relationship in the harvest season than in the planting season.

Non-agricultural disamenities are equivalent to $\theta \Delta W_j$, with j being non-agricultural work inside the village. Similarly, disamenities associated with leaving the village are measured by choosing j as agriculture outside the village. We tale from Table 5 estimates for average utility $\hat{\mu}_{\text{ag inside}} = -0.323$, $\hat{\mu}_{\text{nonag inside}} = -2.543$, $\hat{\mu}_{\text{ag outside}} = -3.878$, and $\hat{\alpha}_{\text{ag inside}} = 1.097$. Hence for male, non-SC/ST workers:

$$\begin{aligned}\text{Non-agricultural disamenity} &= \hat{\theta}\Delta\hat{W}_{\text{nonag inside}}\\ &= 23 \times \left|\frac{-2.543-(-0.323)}{1.097}\right| = 46.6 \text{ rupees.}\\ \text{Outside-village disamenity} &= \hat{\theta}\Delta\hat{W}_{\text{ag outside}}\\ &= 23 \times \left|\frac{-3.878-(-0.323)}{1.097}\right| = 74.5 \text{ rupees.}\end{aligned}$$

Similar calculations yield corresponding non-agricultural and outside-village disamenities for female, non-SC/ST workers to be 53.7 and 44.0 rupees, respectively. Male agricultural workers in our survey earned an average of 205 rupees per day, while female wages average 140 per day. Therefore, non-agricultural disamenities amount to 23% of the average daily wage for male, non-SC/ST workers and 38% for female counterparts, while outside-village disamenities are 36% of the daily wage for males and 31% for females.

The geographical disamenity reflects the cost of leaving one's village. It relates to the valuation of the commuting cost estimated in the labor literature. For example, Le Barbanchon et al. (2021) estimate the commute time to be valued at 80% of the hourly wage for men and 98% for women among job seekers in France. They cite a large literature that finds that the average value of commuting time varies between 20% and 100% of the gross wage rate in industrialized countries. To our knowledge, similar studies for rural areas in developing countries do not exist. In our data, we estimate the disamenity of working "outside your own village" to be 36% of the daily wage, or 2.9 times the hourly wage for an 8-hour workday.

Further, the quantitatively and statistically significant estimates of the standard deviations of the disutility terms in the discrete choice model suggest substantial heterogeneity in disamenities across workers, as in Mas and Pallais (2017). Using the standard deviation and correlation estimates, we obtain the distributions of estimated sectoral and geographic disamenities. For male, non-SC/ST workers, the 25th–75th percentile range of the non-agricultural disamenity is 23.7–73.9 rupees, and that of the outside-village disamenity is 40.3–111.8 rupees. This indicates that, while their average non-agricultural disamenity is of the same order as the wage gap or 23% of their daily wage, there is still 25% of the male workers that would not go into non-agricultural work even for a compensating wage differential of 74 rupees.

Stated preference approach. As an alternative approach, we calculate disamenities associated with non-agricultural work and leaving the village using the estimated labor supply curve based on a worker survey we conducted. First, in order to assess the size of the non-agricultural disamenities, we predict the change in labor allocation that would occur if the distribution of non-agricultural disamenities were the same as that of agricultural disamenities, keeping disamenities for leaving the village constant.[19] Denote the change in choice probability by $\Delta P_{\text{nonag inside}}$.

We quantify disamenities for working outside the village in a similar way. Specifically, we decrease the disutility of agricultural work outside the village to that of agricultural work inside the village. Doing so eliminates the spatial friction, but keeps non-agricultural disamenities constant. Denote this change in choice probability by $\Delta P_{\text{ag outside}}$.

Table 7 shows these changes in the choice probabilities. Eliminating non-agricultural disamenities increases non-agricultural work by $\Delta P_{\text{nonag inside}} = 9.4$ percentage point (row 1, column 3). And eliminating disamenities associated with working outside the village increases agricultural labor outside the village by $\Delta P_{\text{ag outside}} = 13.1$ percentage point (row 2, column 2). These increases in choice probabilities are drawn from the other occupational options. In particular, the share of unemployment drops in both cases, suggesting that disamenities associated with non-agricultural work or working outside the village have implications for rural unemployment.

Table 7
Quantification of disamenities for non-agricultural work and leaving the village.

	Agriculture		Non-agriculture		Own field	Nothing/
	Inside (1)	Outside (2)	Inside (3)	Outside (4)	(5)	House (6)
Non-agricultural disamenity	−0.021	−0.003	0.094	−0.002	−0.040	−0.028
Outside-village disamenity	−0.029	0.131	−0.005	−0.003	−0.053	−0.041

Notes: The table shows changes in the choice probabilities under different scenarios relative to the baseline. Rows 1–2 report results for the scenarios where disamenities associated with non-agricultural work or leaving the village are eliminated, *i.e.*, when the distribution (mean and SD) of the utility of working in agriculture outside the village or non-agriculture inside the village, respectively, is the same as that of working in agriculture inside the village.

We convert these changes into monetary terms using an estimated hypothetical labor supply curve. During the follow-up survey from year 2, we asked workers their willingness to work in agriculture at a random wage. For this survey, we drew a random wage from the uniform distribution and asked the workers the number of days in a month they would be willing to work at that wage. Figure A6 displays a binned scatter plot of the data. We use these data to compute a wage-equivalent change for any change in the probability to work in agriculture on a given day. The corresponding regression results for the figure show that an increase of 12.91 rupees in the daily wage corresponds to one additional day of agricultural work over the 30-day period.[20] In other words, each percentage point of work maps to an increase of $\beta_{\text{wtw}} = 3.87$ rupees in the daily wage.[21] We then measure non-agricultural and outside-village disamenities for male, non-SC/ST workers as follows:

$$\begin{aligned}\text{Non-agricultural disamenity} &= \beta_{\text{wtw}}\Delta P_{\text{nonag inside}} = 3.87 \times 9.4\\ &= 36.4 \text{ rupees.}\\ \text{Outside-village disamenity} &= \beta_{\text{wtw}}\Delta P_{\text{ag outside}} = 3.87 \times 13.1\\ &= 50.7 \text{ rupees.}\end{aligned}$$

Although this approach is based on stated preferences, the advantage is that wages were randomized, offering us a validity check for the revealed preference approach. Importantly, the estimated disamenities are of the same order of magnitude for both the revealed and stated preference approaches.

This approach again indicates large heterogeneity of compensating differentials across workers. Using the standard deviation and correlation estimates from the model, we obtain the distributions of the estimates for sectoral and geographic disamenities based on the stated preference approach: The 25th–75th percentile range of the non-agricultural disamenity is 7.8–61.6 rupees, and that of the outside-village disamenity is 15.4–82.0 rupees.

We also estimate the choice model separately for the switchers, with results shown in Table A6. Carrying out the calculations based on the revealed preference approach, we obtain the non-agricultural

[19] This step involves setting both the mean and standard deviation of the utility of doing non-agricultural labor inside the village to be equal to that of doing agricultural labor there.

[20] The estimated labor supply curve does not seem to vary by gender: if we add an interaction term between female and the hypothetical wage in the labor supply regression, the coefficient on the interaction term is insignificant.

[21] The calculation is $\frac{12.91}{1/30\times100} = 3.87$.

disamenity to be 22.7 rupees for switchers (male, non-SC/ST). Given that there are about 4 times as many specialized agricultural workers as there are switchers, this number is, not surprisingly, smaller than the 46.6 rupees estimated for all workers. The switchers' disamenity for working outside the village is 63.6 rupees, also smaller than the estimate for all workers.

5. Concluding remarks

Models of labor (mis)allocation in developing countries tend to focus on reallocation across space from rural to urban areas. Reallocation across sectors within rural areas has received less attention. We have presented evidence that laborers in rural Indian villages can increase daily earnings by about 23% from moving out of agriculture and working in the nearby non-agricultural sector. Our worker surveys reveal that the type and location of work available in the rural non-agricultural sector might be less desirable than the familiar work in agriculture. Building on this observation, we estimate a model of labor allocation across sectors to quantify these disutilities. The model estimation shows that disamenities associated with non-farm work even within the same village amount to about 23% of the daily wage for males and 38% for females.

A multitude of factors underlie the phenomenon that workers remain engaged in agriculture in rural areas. Most explanations from the literature center around barriers to rural–urban migration. But rural–urban migration is not the only source of structural transformation, particularly in places like India where the rural non-agricultural sector has grown in recent years. As such, there is a need to understand what keeps people from moving to that sector. Our findings show that while workers can earn higher wages in rural non-agricultural work, there may be characteristics of these jobs that cause workers to demand more compensation. We see value in future work that continues to explore the rural non-farm sector and its role in structural transformation.

CRediT authorship contribution statement

Ceren Baysan: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Manzoor H. Dar:** Conceptualization, Investigation, Supervision. **Kyle Emerick:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Zhimin Li:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Elisabeth Sadoulet:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jdeveco.2024.103270.

References

Alvarez, J.A., 2020. The agricultural wage gap: Evidence from Brazilian micro-data. Am. Econ. J.: Macroecon. 12 (1), 153–173.

Asher, S., Novosad, P., 2020. Rural roads and local economic development. Amer. Econ. Rev. 110 (3), 797–823.

Baseler, T., 2023. Hidden income and the perceived returns to migration. Am. Econ. J. Appl. Econ. 15 (4), 321–354.

Binswanger-Mkhize, H.P., 2013. The stunted structural transformation of the Indian economy. Econ. Political Weekly 48 (26–27), 5–13.

Bryan, G., Morten, M., 2019. The aggregate productivity effects of internal migration: Evidence from Indonesia. J. Polit. Econ. 127 (5), 2229–2268.

Clark, C., 1957. The Conditions of Economic Progress. Macmillan.

Duncan, G.J., Holmlund, B., 1983. Was adam smith right after all? Another test of the theory of compensating wage differentials. J. Labor Econ. 1 (4), 366–379.

Fisher, A.G.B., 1939. Production, primary, secondary, and tertiary. Econ. Record 15 (1), 24–38.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., et al., 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. Sci. Data 2 (1), 1–21.

Government of India, 2021. Periodic Labour Force Survey (2019–2020). National Statistical Office.

Hamory, J., Kleemans, M., Li, N.Y., Miguel, E., 2021. Reevaluating agricultural productivity gaps with longitudinal microdata. J. Eur. Econom. Assoc. 19 (3), 1522–1555.

Heise, S., Porzio, T., 2022. Labor misallocation across firms and regions. National Bureau of Economic Research w30298.

Herrendorf, B., Schoellman, T., 2018. Wages, human capital, and barriers to structural transformation. Am. Econ. J.: Macroecon. 10 (2), 1–23.

Imbert, C., Papp, J., 2020. Costs and benefits of rural–urban migration: Evidence from India. J. Dev. Econ. 146, 102473.

Jeong, D., 2021. Creating (digital) labor markets in rural tanzania. Available at SSRN 4043833, Working Paper.

Johnston, B.F., 1970. Agriculture and structural transformation in developing countries: A survey of research. J. Econ. Lit. 8 (2), 369–404.

Kuznets, S., 1957. Quantitative aspects of the economic growth of nations: II. industrial distribution of national product and labor force. Econom. Dev. Cult. Chang. 5 (4), 1–111.

Lagakos, D., 2020. Urban-rural gaps in the developing world: Does internal migration offer opportunities? J. Econ. Perspect. 34 (3), 174–192.

Lagakos, D., Marshall, S., Mobarak, A.M., Vernot, C., Waugh, M.E., 2020. Migration costs and observational returns to migration in the developing world. J. Monetary Econ. 113, 138–154.

Lagakos, D., Mobarak, A.M., Waugh, M.E., 2023. The welfare effects of encouraging rural–urban migration. Econometrica 91 (3), 803–837.

Le Barbanchon, T., Rathelot, R., Roulet, A., 2021. Gender differences in job search: Trading off commute against wage. Q. J. Econ. 136 (1), 381–426.

Lewis, W.A., 1954. Economic development with unlimited supplies of labour. Manch. Sch. 22 (2), 139–191.

Mas, A., Pallais, A., 2017. Valuing alternative work arrangements. Amer. Econ. Rev. 107 (12), 3722–3759.

Morten, M., 2019. Temporary migration and endogenous risk sharing in village India. J. Polit. Econ. 127 (1), 1–46.

Munshi, K., Rosenzweig, M., 2016. Networks and misallocation: Insurance, migration, and the rural–urban wage gap. Amer. Econ. Rev. 106 (1), 46–98.

Oh, S., 2023. Does identity affect labor supply? Amer. Econ. Rev. 113 (8), 2055–2083.

Pulido, J., Swiecki, T., 2019. Barriers to mobility or sorting? Sources and aggregate implications of income gaps across sectors in Indonesia. Working Paper. University of British Columbia.

Reddy, D.N., Reddy, A.A., Nagaraj, N., Bantilan, C., 2014. Rural non-farm employment and rural transformation in India. ICRISAT Working Paper Series, 57, International Crops Research Institute for the Semi-Arid Tropics.

Smith, R.S., 1979. Compensating wage differentials and public policy: A review. Indu. Labor Relat. Rev. 32 (3), 339–352.

World Bank, 2017. Growing the rural nonfarm economy to alleviate poverty. Independent Evaluation Group.

Young, A., 2013. Inequality, the urban-rural gap, and migration. Q. J. Econ. 128 (4), 1727–1785.