The Agricultural Wage Gap Within Rural Villages

Ceren Baysan, Manzoor H. Dar, Kyle Emerick, Zhimin Li, and Elisabeth Sadoulet*

September 4, 2023

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, those associated with location and those associated with job characteristics, such as the difficulty of the work. We find that while costs of working outside the village are significant even within rural areas, job characteristics of non-agricultural work are nearly the same in magnitude.

*Baysan: University of Toronto, ceren.baysan@utoronto.ca; Dar: International Crops Research Institute for the Semi-Arid Tropics, m.dar@cgiar.org; Emerick: Tufts University and CEPR, kyle.emerick@tufts.edu; Li: Peking University HSBC Business School, zhimin.li@pku.edu.cn; Sadoulet: University of California at Berkeley, esadoulet@berkeley.edu. We acknowledge financial support from the J-PAL & CEGA's Agricultural Technology Adoption Initiative and from 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 gratefully acknowledges financial support from the NSF Graduate Research Fellowship program while part of this research was carried out.

1 Introduction

Most of the poor live in rural areas and work in agriculture, earning lower wages than 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 more productive activity, non-agriculture (Fisher, 1939; Lewis, 1954; Clark, 1957; Kuznets, 1957; Johnston, 1970). Most literature on structural transformation focuses on the rural-urban productivity divide (Lagakos, 2020). But in many low-income or lower middle-income countries, the non-farm sector in rural areas has become an important source of employment (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 a detailed panel 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 23% wage gap is based on our main regression in which we control for unobservable workers attributes with individual and survey fixed effects. However, using the variation from a smaller sample of workers switching sectors within one to two weeks for identification, we can also include individual-by-survey fixed effects in our regression. We find that the estimated wage gap remains at 23%. Therefore, our estimate controls not only for individual time-invariant confounders, but also components of unobserved ability that may vary over time.

[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.

For example, a worker may gain additional skills in between agricultural seasons. Or they may migrate in response to a time-varying change in human capital. These changes may occur over several months, but are unlikely to happen over a period of just a few days. Our estimate also represents a specific phenomenon: switching sectors without migrating. This makes it more striking that an individual can earn much higher wages from casual non-farm labor without physically moving.

To shed light on alternative 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 not being able 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 part of the sectoral wage gap reflects mobility constraints. But 78% of the wage gap remains even when focusing on within-village transitions.

We then consider the role of non-agricultual 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. One interpretation of these findings is that attributes of jobs make rural non-agricultural work less preferred, but workers transition to non-farm jobs when agricultural work is hard to find. 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 is consistent with our observation 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. This suggests that there may be complementarities between working in agriculture and self-employment in agriculture. Altogether, we refer to these attributes of non-agricultural work as factors that decrease worker utility and constitute non-agricultural job disamenities.

Finally, we estimate a discrete choice model of labor supply. 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 importance of the lesser explored non-agricultural disamenities within village. We are able to separate these two disamenities by observing people working in different sectors — both inside and outside their villages. The monetary value of mobility related disamenities equals about 65 rupees, or around 32% of the male agricultural wage. 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 a small geographic space. At the same time, we find that other disamenities of non-agricultural work amount to around 20% of the male wage. This 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 male workers.

In the development economics literature, previous research has found urbanization to be an important source of the sectoral earnings gap (e.g., Lagakos, Mobarak, and Waugh, forthcoming; Bryan and Morten, 2019; Morten, 2019; Imbert and Papp, 2020; Baseler, 2021). 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. We find that the raw wage gap decreases from 1.38 to 1.23 after controlling for ability. Therefore, while there is some sorting that explain a raw sectoral wage gap, it is less significant than in the urbanization literature, and the wage gap remains nearly as high as the rural-urban wage gap in India. Second, we find that mobility-related constraints explain part of the remaining sectoral wage gap, but only 22%. Instead, a lesser explored hypothesis in developing countries, despite 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 evidence showing that workers earn higher wages in non-agricultural jobs. It also shows survey evidence to explain sources of the wage gap and evidence on how workers move between sectors in response to shocks. Section 3 outlines a model of daily labor allocation choices in the presence of both types of disamenities. Section 4 estimates the parameters of the model. 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 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 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 the people dependent on agriculture in rural India.

Hiring and wages in casual labor markets in India are generally determined on a daily basis. Yet, most studies rely on data that aggregates 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. We did this by conducting phone surveys. These surveys took place during the transplanting 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. Lack of irrigation limits cultivation and agricultural employment during other times of the year.

In the first two phone surveys in August and November 2014, we collected data on whether laborers worked on another person's farm 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 collected this information for the seven days preceding the phone call. We repeated this same process in the 2015 and 2016 seasons with two important differences. First, we expanded the sample to include 6 female laborers per village. The original sample contained only 3 female laborers per village. We selected the three added laborers from a census in all villages on households with casual laborers.[2] Second, starting with the 2015

[2]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 adding 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.

harvesting survey, we expanded the recall window. We doubled the period to 14 days to better capture the entire planting or harvesting period. The phone surveys produced a high response rate: we reached an average of 86% of the workers from the baseline.[3]

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 allow 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. 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 workers who stay in agriculture. About 20% of the workers from the baseline survey switched sectors. Switchers are more likely to be male. Switchers 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. Switchers are more likely to come from households with temporary migrants. Yet, switchers have no less land. The average laborer household cultivates 0.57 acres during the rainy season. Only about 16% of households cultivate no land at all.[4]

Figure 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. The figure illustrates that mostly male laborers are engaged in non-agricultural work and are the main source of our identification strategy. Around 4 to 8 percent of workers switch sectors during the same survey round. Including only these workers for identification produces the same results as including workers that switch across rounds.

We use three additional sources of data. First, we surveyed the 10 largest farmers after harvesting each year.[5] These data help us link rainfall-induced variation in agricultural

[3]The response rate ranged from 79% in the third year planting survey to 91% in the year two planting survey.

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

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

output with labor flows to the nonfarm sector. Second, 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 productivity data from farmers highlight the importance of timely rainfall. Relative to 2016, yields were lower by 56% in 2015 and 25% in 2014.

Before presenting our empirical strategy and estimation results of the sectoral wage gap, we also describe the dual economy in our setting and worker movement between the two sectors. Our data shows that workers turn to non-agricultural jobs when agricultural work is less available. Agricultural labor demand at harvesting depends on rainfall earlier during the growing season. Rainfall therefore provides a quasi-random source of variation in agricultural labor demand. We focus on the three harvesting surveys and estimate

$$y_{ivtd} = \alpha_v + \gamma_t + \beta Rainfall_{vt} + \varepsilon_{ivtd}, \tag{1}$$

where y_{ivtd} is rice yield or one of four indicator variables for working as an agricultural wage laborer, on your own field, in the non-agricultural sector, or not working / doing housework for worker i in village v on day d of survey round t; α_v and γ_t denote village and survey-round fixed effects. We use total precipitation during the agricultural season for the rainfall variable.[6]

For one, it might be uncomfortable to work in the rain. Or high rainfall might limit mobility. Rainfall might have direct effects on non-farm labor demand if work can only be done on dry days. However, these direct effects of current rainfall on labor allocation are unlikely to drive the estimate of β in equation (1). Our harvesting surveys took place well after monsoon rains had stopped. Equation (1) uses cumulative rainfall variation that happened weeks before our surveys. Figure A1 makes this clear. It shows that our surveys, which took place in late November to early December, happened during periods of no rainfall.

Figure 3 visualizes these results. The figure shows binned scatter plots of different outcome variables against rainfall — after residualizing the data to remove fixed effects. Agri-

[6]Harvesting takes place in November or early December. Therefore, we calculate cumulative block-level rainfall from June through October to measure shocks to agricultural labor demand.

cultural productivity increases with rainfall. 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. The next four panels show how the 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.[7] 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, works shift into non-agriculture during drier years, and this pushes wages down.

Table 4 provides the parameter estimates. A *decrease* in rainfall by 100 mm decreases 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.

2.2 Empirical Approach

To estimate the wage gap, we observe $wage_{ivtd}$, which is the wage for worker i, residing in village v, during survey round t and on day d. The daily data permit us to estimate the wage gap between agricultural and non-agricultural work. To do so, we estimate,

$$\log(wage_{ivtd}) = \alpha_{iv} + \gamma_t + \beta NonAg_{ivtd} + \varepsilon_{ivtd}, \tag{2}$$

where $NonAg_{ivtd}$ is an indicator for wage labor in the non-agricultural sector, α_{iv} 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 check 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 estimate. Previous work on rural-urban migration has estimated sectoral wage gaps using people who switch sectors over longer time periods (Herrendorf and

[7] 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.

Schoellman, 2018; Pulido and Swiecki, 2019; Alvarez, 2020; Hamory et al., 2021). Our specification with the shorter time window allows us to estimate a gap within rural areas for jobs that can be taken within a period of just one to two weeks.

The phone surveys with farmers show that 82% of the workers hired are females. This is larger than the proportion of females we selected in our sample.[8] To correct for this, we weight the data. We calculate the weight for female observations as the share of the hired workers that are female — across all our phone surveys with farmers — divided by the share of respondents from that survey round that were female. We define the weights in the same way for males. This weighting scheme ensures that our estimates represent the average casual agricultural worker — although it does not affect our results.

2.3 Reduced-Form Results

Table 2 shows our estimates of the agricultural wage gap. Column 1 includes individual and survey-round fixed effects, limiting the identification to around one fifth of the sample. We find that agricultural workers increase their daily wages by 23% when moving to non-agricultural work.[9] This estimate is partly identified off of people switching sectors across survey rounds.

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. Column 2 shows that including individual-by-survey round fixed effects produces the same result. This confines the identification to fewer individuals, but we estimate the same wage gap of 23%. Unobservables that can vary across waves do not appear to drive our estimate. As one example, individuals could accumulate more skill over a period of months. But they are less likely to gain these skills in 1-2 weeks.

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, which finds that this type of selection accounts for most of the rural-urban wage gap, we find that much of the wage gap remains when conditioning on individual fixed effects. As an additional note, none of the estimates in Table 2 change meaningfully if we omit the gender weights (Table A1).

[8] 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.

[9] In line with the descriptive evidence above, only about 15% of these non-agricultural work days are from females.

To put our estimate in context, Herrendorf and Schoellman (2018) use census data from 13 countries to show that non-agricultural wages are 1.8 times higher than agricultural wages. Their estimate decreases to 1.33 when adjusting for education, gender, and spatial location. Our estimate focuses on the rural non-agricultural sector and eliminates the most likely sources of unobserved ability. The non-agricultural gap in our setting is about 1.23.[10] This is close to the rural-urban wage gap in India of 25% after adjusting for differences in cost of living (Munshi and Rosenzweig, 2016).

2.4 Understanding the Sectoral Wage Gap

Before turning to our model, we show more 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.

Figure 2 shows that two explanations stand out. Just over 32% say that non-agricultural jobs are either 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.[11] This evidence does not pinpoint what makes these jobs harder. It instead provides suggestive evidence that preferences for job type give another reason why workers choose agricultural work. This could be because non-agricultural jobs are more physically demanding, require longer hours, risky, or involve tasks that are less familiar than agricultural activities.[12] Indeed, non-agricultural work in rural areas often requires physically demanding tasks. During this same survey we

[10] The precise gap from the log wage regression is $e^{0.207} = 1.23$.

[11] Results in the online appendix (Figure A2) show that the share 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.

[12] 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 hours for males and 7.5 hours for females. Using variation in daily hours, Table A2 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.

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, as we will discuss in Section 4.1, we find that working on one's own farm and working on someone else's farm are positively correlated, suggesting that there are important complementarities between these two activities but not non-agricultural work.

Our estimated wage gap reflects both mobility constraints to switching sectors, as well as disamenities associated with characteristics of jobs available in the same village. 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. Focusing on column 1, relative to agricultural jobs in one's own village, non-agricultural jobs in that same village yield 17% higher wages. Both changing sectors and working outside the village increases wages by 37%. The larger wage gains from leaving the village can capture the costs associated with transport or not being able to work flexibly on one's own farm. At the same time, transitioning to non-agricultural work in the same village leads to higher wages, but does not require large search- or mobility-related costs.[13] These results suggest that the wage gap captures more than just mobility constraints.

In combination, multiple factors drive the wage gap between sectors. But mobility constraints and the difficulty of non-farm work are two common explanations. 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 are an additional factor in explaining the sectoral wage gap.

3 A Model of Rural Labor Allocation

Our analyses 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 being located outside the village — that make them 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 suggested that difficulty of jobs like 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

[13]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 A4 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.

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 for within village.

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 if non-agricultural disamenities explain why workers stay in agriculture.

To quantify 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 use this feature of the data to 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 allows 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. They are also 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, and we do not observe the (counterfactual) wage offers for unchosen options. Instead, 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.[14] 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 denote 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. The term b

[14] 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.

denotes block, d day, and t survey round. Worker i's utility, U_{ijdt}, is represented as follows:

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

where W_{bt} is cumulative rainfall during the growing season (collected 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. 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 disutility 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. Finally, ϵ_{ijtd} a random component that is assumed to take on Type-I Extreme Value distribution. Note that the random disutilities δ_i can be potentially correlated across employment options, allowing for flexible substitutional patterns across disutility terms.

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 at planting and harvesting times. The data contain two panel dimensions, as we observe workers 7-14 days (d) in each season, and we have 6 different seasons (t). For the estimation, we consider each block as a separate labor market in each of the 6 seasons. We define rainfall during planting season as total precipitation for the months of June through July. We use the months of June through October for the harvesting season. This 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 for 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, random variations in utility (ϵ_{ijtd}) affect the daily employment status of a worker, and can be interpreted as random variation in offered wage. We capture variation across seasons with the seasonal and weather variables.

We capture individual variation with preference variation arising from the characteristics X_i and random preference shocks δ_i. The model incorporates variation across years in the same season with the weather variations W_{vtd} and past choices y_{ijtd-1}.

4 Model Estimation and Quantification of Disamenities

4.1 Model Estimation Results

Table 5 reports the estimated parameters and their standard errors 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 is the reference option. The model estimates the means (μ), standard deviations (σ), and correlation matrix of the random disutilities δ associated with different job options. We allow the mean disutility of each job type to depend on gender and caste.

The results provide evidence for disamenities associated with non-agricultural work and with space. For example, within villages, workers have a much higher disutility for non-agricultural work than for agricultural work. This disutility is lower for workers from lower castes (ST or SC), especially for non-agricultural work outside the village. This is consistent with the self-reported survey results on why workers choose agricultural work over non-agricultural work even if earnings are lower. 28% of non-ST or SC workers cite the difficulty of nonfarm work as the top reason in contrast to 18% of ST or SC workers. Workers also have a higher disutility for leaving the village, particularly in agriculture. 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 Table 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 outside the village. This suggests that non-agricultural disamenities may partly explain the gender earnings gap.

4.2 Quantifying Disamenities

We do not have data for random wage offers to estimate a direct measure of the marginal utility of money. Thus, we cannot directly convert estimated disamenities into monetary terms. Instead, we quantify them using two different approaches. First, we use quasirandom 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 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 increase in daily wages that would have the same effect on labor supply as those from 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.[15] 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:

$$wage_{ivtd} = \lambda_v + \lambda_t + \theta W_{bt} + \upsilon_{itd}, \tag{3}$$

where λ_v and λ_t are village and survey-round fixed effects. Regression results from estimating Equation 3 are shown in Table 6. A one standard deviation increase in rainfall raises agricultural wage by 20 rupees.

The 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.

Table 5 shows estimates for average disutility $\hat{\mu}_{\text{ag inside}} = -0.323$, $\hat{\mu}_{\text{nonag inside}} = -2.543$,

[15] 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.

$\hat{\mu}_{\text{ag outside}} = -3.878$, and $\hat{\alpha}_{\text{ag inside}} = 1.097$. Hence for male, non SC/ST workers:

$$\text{Nonagricultural disamenities} = \hat{\theta}\Delta\hat{W}_{\text{nonag inside}} = 20 \times \left|\frac{-2.543 - (-0.323)}{1.097}\right| = 40.5 \text{ rupees.}$$

$$\text{Outside village disamenities} = \hat{\theta}\Delta\hat{W}_{\text{ag outside}} = 20 \times \left|\frac{-3.878 - (-0.323)}{1.097}\right| = 64.8 \text{ rupees.}$$

Similar calculations yield corresponding non-agricultural and outside village disamenities for female, non SC/ST workers to be 66.4 and 51.4 rupees, respectively. Male agricultural workers in our survey earned an average of 205 rupees per day, while female wages average 140 per day. The non-agricultural disamenities amount to 20% of this average daily wage for male, non SC/ST workers, while it is 47% for females.

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. First, in order to assess the size of the non-agricultural disamenities, we predict the change in labor allocation that would obtain if the distribution of non-agricultural disamenities were the same as that of agricultural disamenities, keeping disamenities for leaving the village constant.[16] 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. This 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 nonagricultural disamenities increases non-agricultural work by $\Delta P_{\text{nonag inside}} = 9.4$ pp (row 1, column 3). Eliminating disamenities associated working outside the village increases agricultural labor outside the village by $\Delta P_{\text{ag outside}} = 13.1$ pp (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 nonagricultural work or working outside the village have implications for unemployment.

We convert these changes to monetary terms using an estimated hypothetical labor supply curve. We asked workers their willingness to work in agriculture at a random wage during the follow-up survey from year 2. For this, we drew a random wage from the uniform distribution and asked the worker how many days in a month they would be willing to work at that wage. Figure A5 displays a binned scatter plot of the data. We use this to

[16] This involves setting both the mean and standard deviation of δ for non-agricultural labor inside the village to be equal to that of doing agricultural labor inside the village.

compute a wage-equivalent change for any change in the probability to work in agriculture on a given day. The regression results for Figure A5 show that a 12.91 Rs increase in daily wage corresponds to one additional day of agricultural work over the 30 day period. In other words, each percentage point of work maps to $\beta_{\text{wtw}} = 3.877$ Rs daily wage increase. We then measure non-agricultural and outside-village disamenities for male, non SC/ST workers as follows:

$$\text{Nonagricultural disamenities} = \beta_{\text{wtw}} \Delta P_{\text{nonag inside}} = 3.877 \times 9.4 = 36.4 \text{ rupees.}$$
$$\text{Outside village disamenities} = \beta_{\text{wtw}} \Delta P_{\text{ag outside}} = 3.877 \times 13.1 = 50.8 \text{ rupees.}$$

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

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. Surveys with workers revealed that the location and the type of work available in the rural non-agricultural sector might be less desirable than the familiar jobs in agriculture. Building on this, we estimated 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 20% of the daily wage for males and 47% for females.

There are many reasons why 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 need more compensation. We see value in future work that continues to explore the rural non-farm sector and its role in structural transformation.

References

Alvarez, Jorge A. 2020. "The agricultural wage gap: Evidence from Brazilian micro-data." *American Economic Journal: Macroeconomics* 12 (1):153–73.

Asher, Sam and Paul Novosad. 2020. "Rural roads and local economic development." *American Economic Review* 110 (3):797–823.

Baseler, Travis. 2021. "Hidden income and the perceived returns to migration." Working Paper.

Binswanger-Mkhize, Hans P. 2013. "The stunted structural transformation of the Indian economy." *Economic and Political Weekly* 48 (26-27):5–13.

Bryan, Gharad and Melanie Morten. 2019. "The aggregate productivity effects of internal migration: Evidence from Indonesia." *Journal of Political Economy* 127 (5):2229–2268.

Clark, Colin. 1957. *The Conditions of Economic Progress*. Macmillan.

Duncan, Greg J and Bertil Holmlund. 1983. "Was Adam Smith right after all? Another test of the theory of compensating wage differentials." *Journal of Labor Economics* 1 (4):366–379.

Fisher, Allan G. B. 1939. "Production, primary, secondary, and tertiary." *The Economic Record* 15 (1):24–38.

Funk, Chris, Pete Peterson, Martin Landsfeld, Diego Pedreros, James Verdin, Shraddhanand Shukla, Gregory Husak, James Rowland, Laura Harrison, Andrew Hoell et al. 2015. "The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes." *Scientific Data* 2 (1):1–21.

Government of India. 2021. "Periodic Labour Force Survey (2019-2020)." *National Statistical Office* .

Hamory, Joan, Marieke Kleemans, Nicholas Y Li, and Edward Miguel. 2021. "Reevaluating agricultural productivity gaps with longitudinal microdata." *Journal of the European Economic Association* 19 (3):1522–1555.

Heise, Sebastian and Tommaso Porzio. 2022. "Labor misallocation across firms and regions." National Bureau of Economic Research w30298.

Herrendorf, Berthold and Todd Schoellman. 2018. "Wages, human capital, and barriers to structural transformation." *American Economic Journal: Macroeconomics* 10 (2):1–23.

Imbert, Clément and John Papp. 2020. "Costs and benefits of rural-urban migration: Evidence from India." *Journal of Development Economics* 146:102473.

Jeong, Dahyeon. 2021. "Creating (digital) labor markets in rural Tanzania." Working Paper.

Johnston, Bruce F. 1970. "Agriculture and structural transformation in developing countries: A survey of research." *Journal of Economic Literature* 8 (2):369–404.

Kuznets, Simon. 1957. "Quantitative aspects of the economic growth of nations: II. Industrial distribution of national product and labor force." *Economic Development and Cultural Change* 5 (4):1–111.

Lagakos, David. 2020. "Urban-rural gaps in the developing world: Does internal migration offer opportunities?" *Journal of Economic Perspectives* 34 (3):174–92.

Lagakos, David, Samuel Marshall, Ahmed Mushfiq Mobarak, Corey Vernot, and Michael E Waugh. 2020. "Migration costs and observational returns to migration in the developing world." *Journal of Monetary Economics* 113:138–154.

Lagakos, David, Ahmed Mushfiq Mobarak, and Michael E Waugh. forthcoming. "The welfare effects of encouraging rural-urban migration." *Econometrica* .

Lewis, W Arthur. 1954. "Economic development with unlimited supplies of labour." *The Manchester School* 22 (2):139–191.

Mas, Alexandre and Amanda Pallais. 2017. "Valuing alternative work arrangements." *American Economic Review* 107 (12):3722–3759.

Morten, Melanie. 2019. "Temporary migration and endogenous risk sharing in village India." *Journal of Political Economy* 127 (1):1–46.

Munshi, Kaivan and Mark Rosenzweig. 2016. "Networks and misallocation: Insurance, migration, and the rural-urban wage gap." *American Economic Review* 106 (1):46–98.

Pulido, José and Tomasz Swiecki. 2019. "Barriers to mobility or sorting? Sources and aggregate implications of income gaps across sectors in Indonesia." Working Paper.

Reddy, D Narasimha, A Amarender Reddy, N Nagaraj, and Cynthia Bantilan. 2014. "Rural non-farm employment and rural transformation in India." *ICRISAT Working Paper Series* (57).

Smith, Robert S. 1979. "Compensating wage differentials and public policy: A review." *Industrial and Labor Relations Review* 32 (3):339–352.

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

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

Tables

Table 1: Baseline characteristics

	Ag Only (N=1499)	Switchers (N=387)	p-value
<u>*Individual Variables:*</u>			
Female	0.388	0.101	0.000***
Years of education	3.477	3.463	0.947
Cognitive ability	2.787	2.708	0.131
<u>*Household Variables:*</u>			
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**

The table shows average values of baseline characteristics between workers that worked only in agriculture for all three surveys that were used to estimate the agricultural wage gap (column 1) and those that 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. 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 that is away for at least 10 months of the year. A temporary migrant is defined as an individual that leaves the village during the dry season but returns home during the wet season. Cognitive ability is the score on a reverse digit span test.

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
Number workers	2285	2285	2285	2285	2285
Number of Observations	28598	28598	28598	28598	28598
R squared	0.785	0.940	0.315	0.538	0.748

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 are clustered at the village level in all specifications. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

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
Number workers	2285	2285	2285	2285	2285
Number of observations	28598	28598	28598	28598	28598
R squared	0.787	0.941	0.325	0.546	0.751

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 are clustered at the village level in all specifications. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

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
Rainfall	0.520***	0.071***	0.036*	-0.051***	-0.045***
June-October	(0.050)	(0.016)	(0.021)	(0.013)	(0.016)
Village fixed effects	Yes	Yes	Yes	Yes	Yes
Survey fixed effects	Yes	Yes	Yes	Yes	Yes
Mean outcome	0.36	0.22	0.37	0.16	0.24
Number laborers		2645	2645	2645	2645
Number of Observations	5898	78449	78449	78449	78449
R squared	0.463	0.241	0.140	0.170	0.162

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), and an indicator for not working or doing housework (column 5). 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 are clustered at the village level in all specifications. Asterisks indicate a coefficient that is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table 5: Mixed logit estimation of preference parameters

	Agriculture		Non-Agriculture		Own Field
	(1)	(2)	(3)	(4)	(5)
	Inside	Outside	Inside	Outside	Own Field
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)		
Shares in data	0.209	0.036	0.054	0.049	0.400
Predicted shares	0.211	0.038	0.055	0.049	0.395

The table shows coefficients results from the mixed logit estimation of the 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 is used as the reference category. The last two rows show the share of workers 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.

Table 6: Regression of wages on rainfall

	(1) Wage	(2) Wage	(3) Wage	(4) Wage
Rainfall	20.45***	20.03***	20.09***	19.49***
	(2.60)	(2.59)	(2.69)	(2.69)
Village fixed effects	Yes	Yes	No	No
Block fixed effects	No	No	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Number of observations	26632	26632	26632	26632
R squared	0.314	0.322	0.205	0.209

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 controlling for village fixed effects, and 3-4 controlling for block fixed effects. Additional control variables include gender and cast of farmers. Standard errors are in parentheses. Asterisks indicate a coefficient that is statistically significant at the 1% ***, 5% **, and 10% * levels.

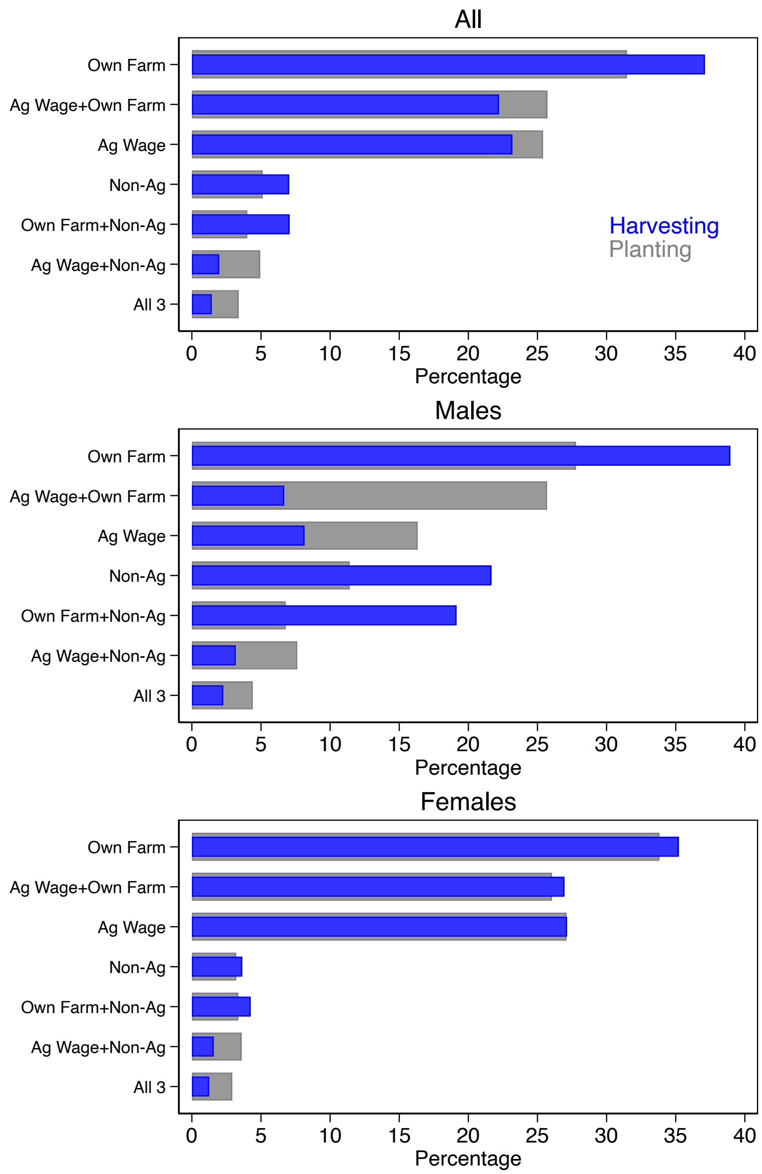
Table 7: Quantification of disamenities for non-agricultural work and leaving the village

	Agriculture		Non-Agriculture		Work in	Unemployed
	Inside	Outside	Inside	Outside	Own Field	
	(1)	(2)	(3)	(4)	(5)	(6)
Nongricultural disamenities	-0.021	-0.003	0.094	-0.002	-0.040	-0.028
Outside village disamenities	-0.029	0.131	-0.005	-0.003	-0.053	-0.041

The table shows changes in the choice probabilities relative to the baseline under different scenarios. 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 disutility of agriculture outside or non-agriculture inside labor, respectively, is the same as that of agriculture inside.

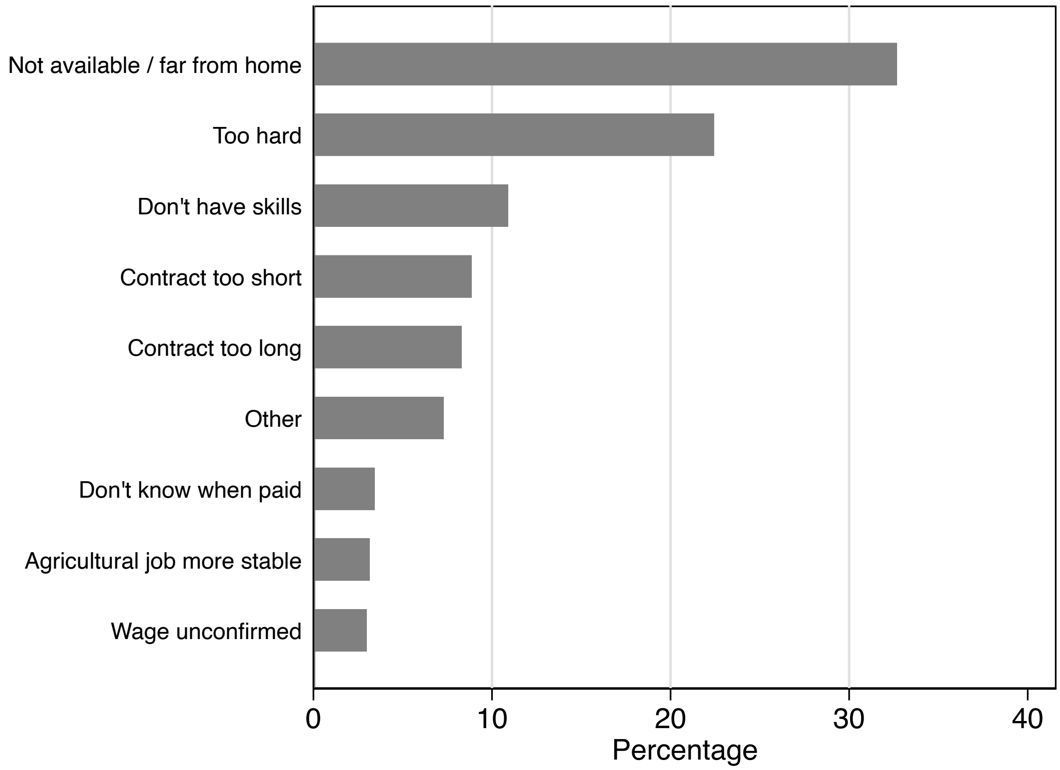
Figures

Figure 1: Activities of workers during 7-14 day survey period



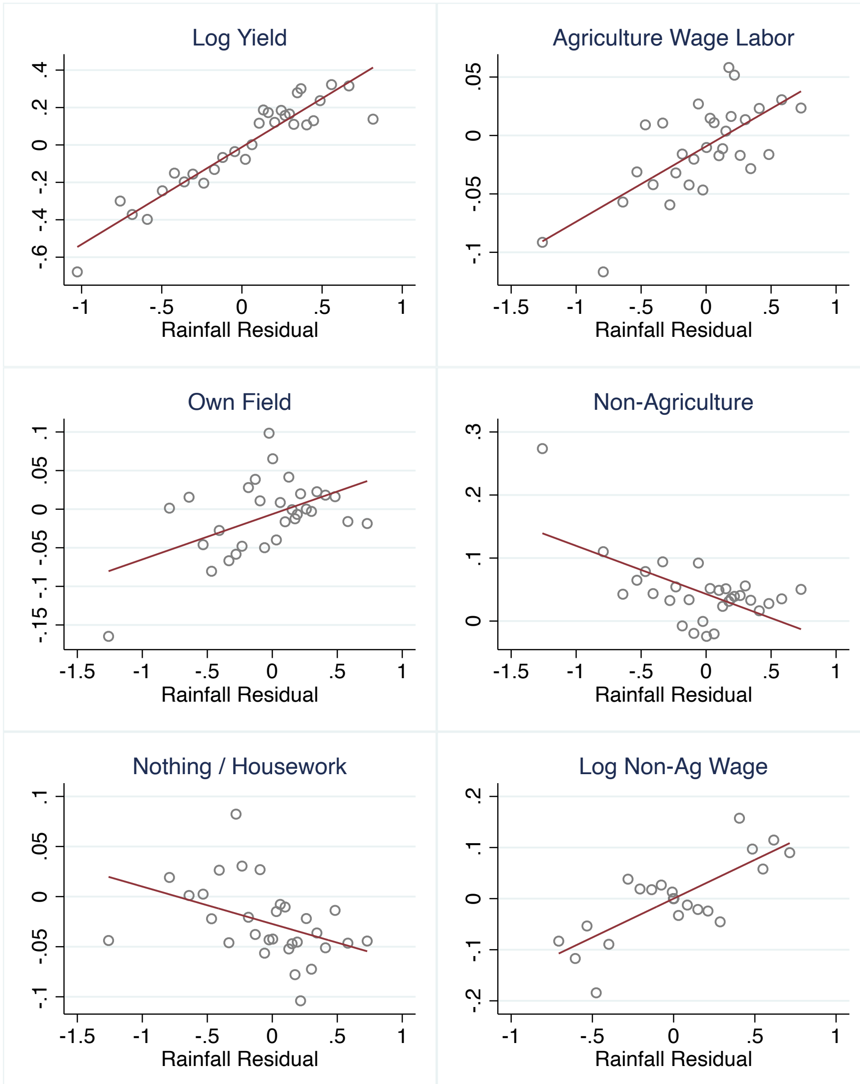
The figure shows a classification of workers into seven groups, depending on which activities they did during the 7 or 14 day survey period. The top panel is for all respondents and is weighted by gender to represent the sex ratios of the population of agricultural workers hired by large farmers. The 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-agricultural" indicates non-agricultural work. The grey 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).

Figure 2: Stated reasons why laborers still do not work in the non-agricultural sector even when wages are higher



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."

Figure 3: The relationships between rainfall realizations, agricultural productivity, and labor allocation



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.

Appendix: Additional Tables and Figures for Online Publication

Table A1: Unweighted estimates of the agricultural wage gap

	Individ, Survey (1)	Individ by Survey (2)	Survey (3)	Village, Survey (4)	Village by Survey (5)
Non-ag work	0.217***	0.175***	0.312***	0.322***	0.325***
	(0.024)	(0.043)	(0.026)	(0.024)	(0.024)
Mean ag wages	169	169	169	169	169
Number workers	2288	2288	2288	2288	2288
Number of Observations	28610	28610	28610	28610	28610
R squared	0.854	0.960	0.485	0.632	0.765

The table presents the same regressions as Table 2 but without weighting observations by gender. The specifications are otherwise the same. The data are from three surveys where non-agricultural wages were collected: the planting survey of 2015, and 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 gender of the respondent, based on the 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 are clustered at the village level in all specifications. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table A2: Correlation between agricultural daily wages and the length of the work day

	Male Log Wages		Female Log Wages	
	(1)	(2)	(3)	(4)
Hours	-0.072***	-0.040*	-0.036***	-0.019*
	(0.019)	(0.021)	(0.013)	(0.011)
Planting	-0.066***	-0.039**	-0.036	-0.083**
	(0.017)	(0.016)	(0.093)	(0.036)
Weeding	-0.094**	-0.036	0.005	-0.064*
	(0.040)	(0.040)	(0.094)	(0.035)
Threshing	-0.014	-0.032***	-0.025	-0.060*
	(0.012)	(0.009)	(0.091)	(0.036)
Harvesting	-0.069**	-0.059***	-0.032	-0.079**
	(0.028)	(0.022)	(0.092)	(0.036)
Village fixed effects	No	Yes	No	Yes
Mean wages (level)	186	186	117	117
Number of Observations	1835	1835	2520	2520
R squared	0.044	0.513	0.013	0.605

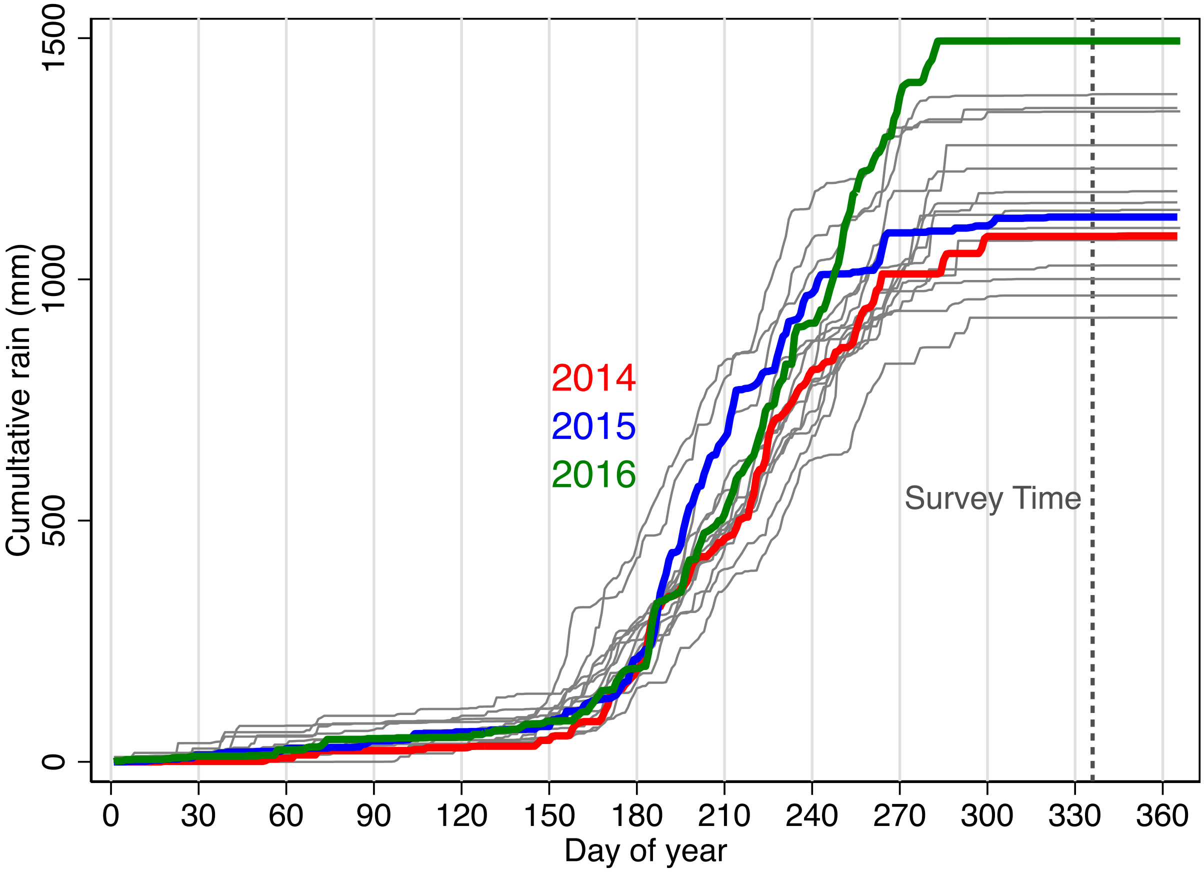
The data are from the survey with farmers after the 2014 season. Farmers were asked for male and female wages, separately by task and gender. Farmers were also asked for the length of a typical work day by gender and task. The dependent variables are the log of male wages (columns 1 and 2) and the log of female wages (columns 3 and 4). Standard errors are clustered at the village level in all specifications. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table A3: Robustness to dropping non-agricultural work outside of the worker's own village

	Individual	Individual by Survey
	(1)	(2)
Non-ag work	0.166***	0.192**
	(0.048)	(0.087)
Mean ag wages	169	169
Number workers	2242	2242
Number of Observations	27236	27236
R squared	0.774	0.936

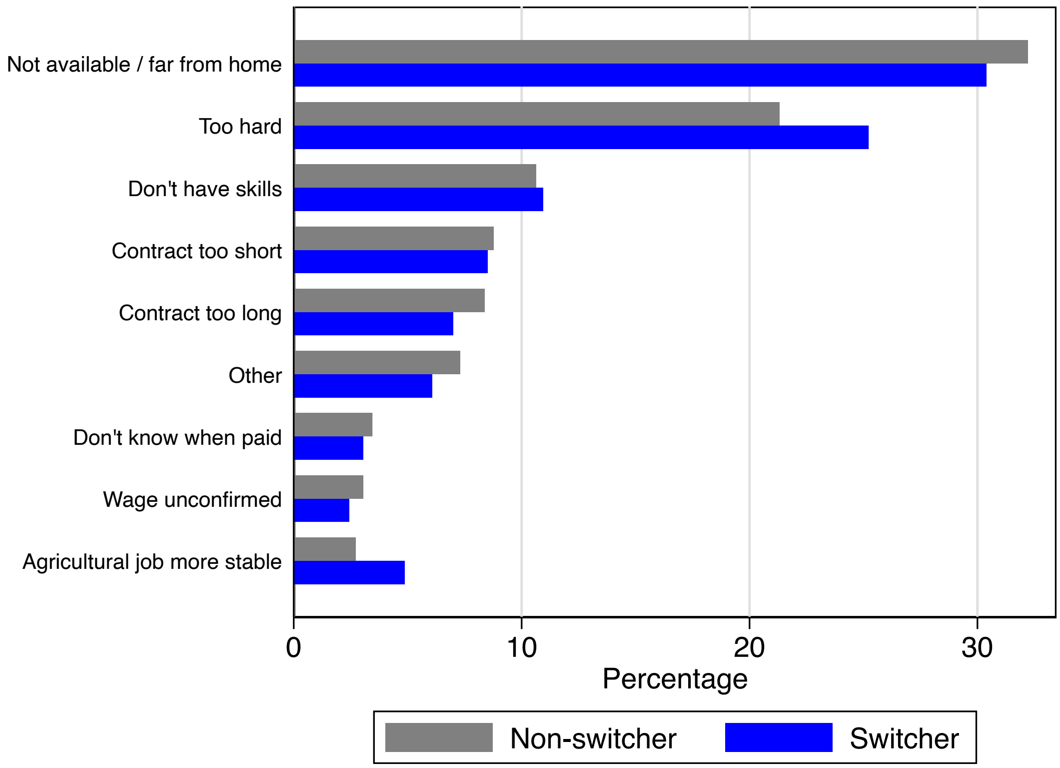
The data are from three surveys where non-agricultural wages were collected: planting time of 2015, and the planting and harvesting surveys of 2016. This table drops days of non-agricultural work which were classified as outside the village (either migrant labor or when the work was outside the village). The dependent variable in both columns is the log of daily wages. Column 1 includes individual, survey, and surveyor fixed effects. Column 2 includes individual-by-survey fixed effects. Observations are weighted by the gender of the respondent, based on the gender shares in the survey with farmers. 384 respondents contribute to the identification in column 1 by working in both the agricultural and non-agricultural sectors across the three surveys. 205 workers contribute to the identification in column 2, i.e. they work in both sectors in the same survey round. Standard errors are clustered at the village level in all specifications. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Figure A1: Cumulative rainfall in study area, 2000-2016



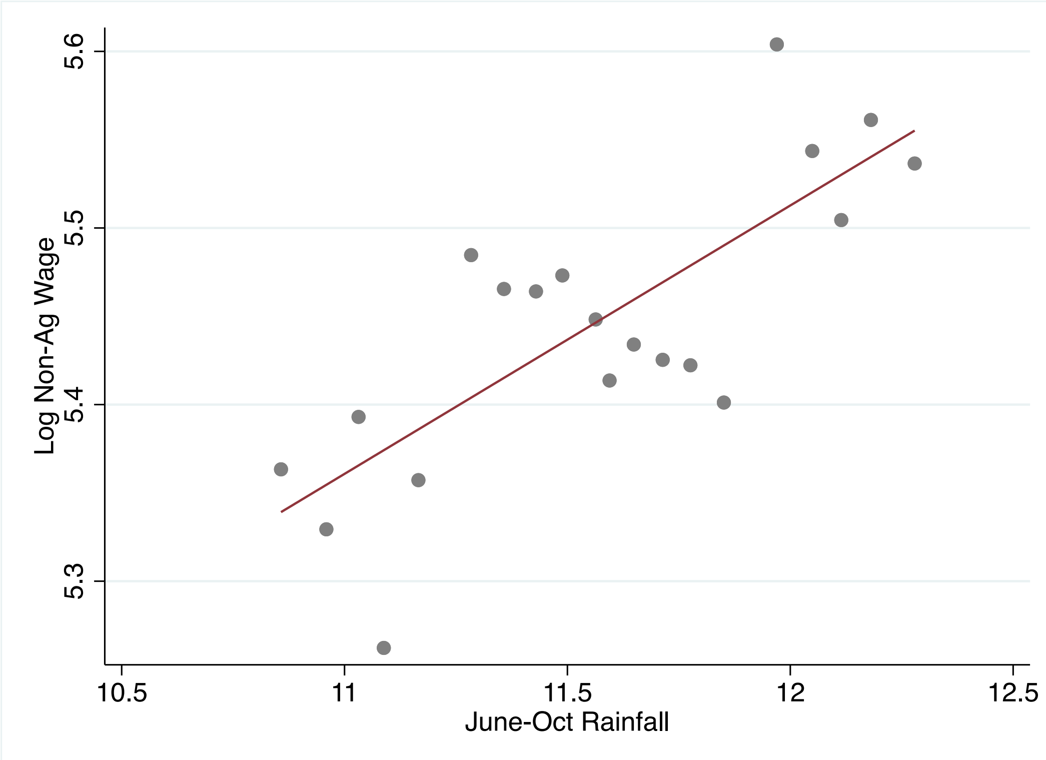
The figure shows cumulative rainfall plotted against the day of the year. Each line is for a separate year. Daily rainfall was first averaged across the 200 sample villages to generate a daily average precipitation for the sample area. The daily rainfall values are satellite observations taken from CHIRPS. The vertical dashed line is the median survey date for the 3 harvesting surveys, measured in days after January 1st of that year.

Figure A2: Stated reasons for not wanting non-agricultural jobs, separately for switchers and non-switchers



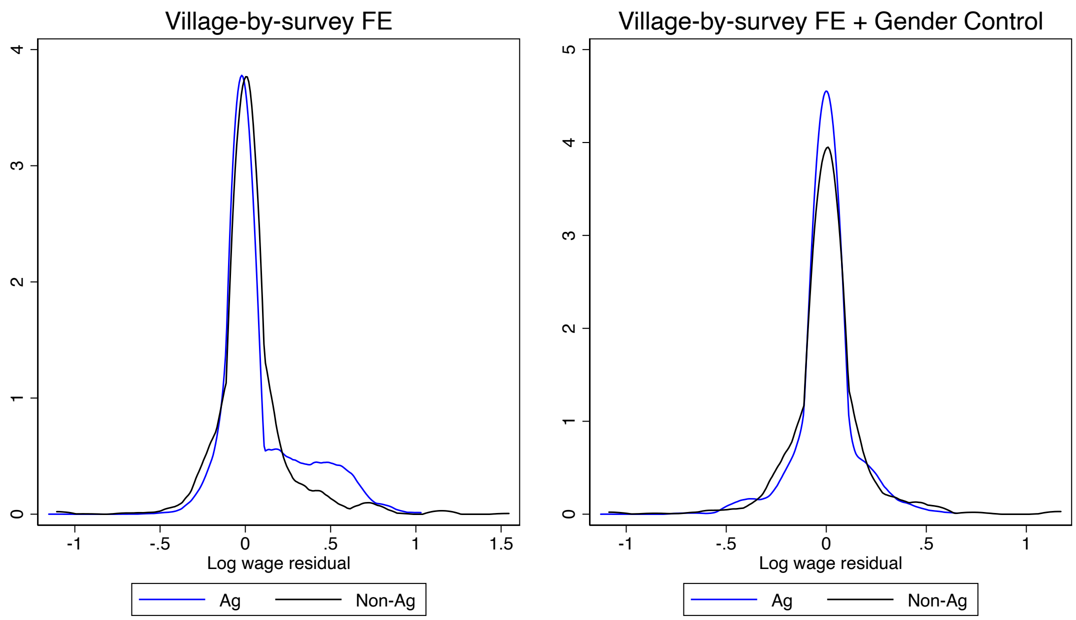
The figure shows responses from the third follow-up survey with agricultural laborers. The exact question posed to laborers was "Suppose wages are a bit lower for agricultural jobs than for non-agricultural jobs, what is the top reason why you may still work in agricultural jobs". The grey bars are for workers that always worked in agriculture, while the blue bars are for people that worked in non-agriculture for at least one day during the sample period.

Figure A3: Relationship between non-agricultural wages and rainfall



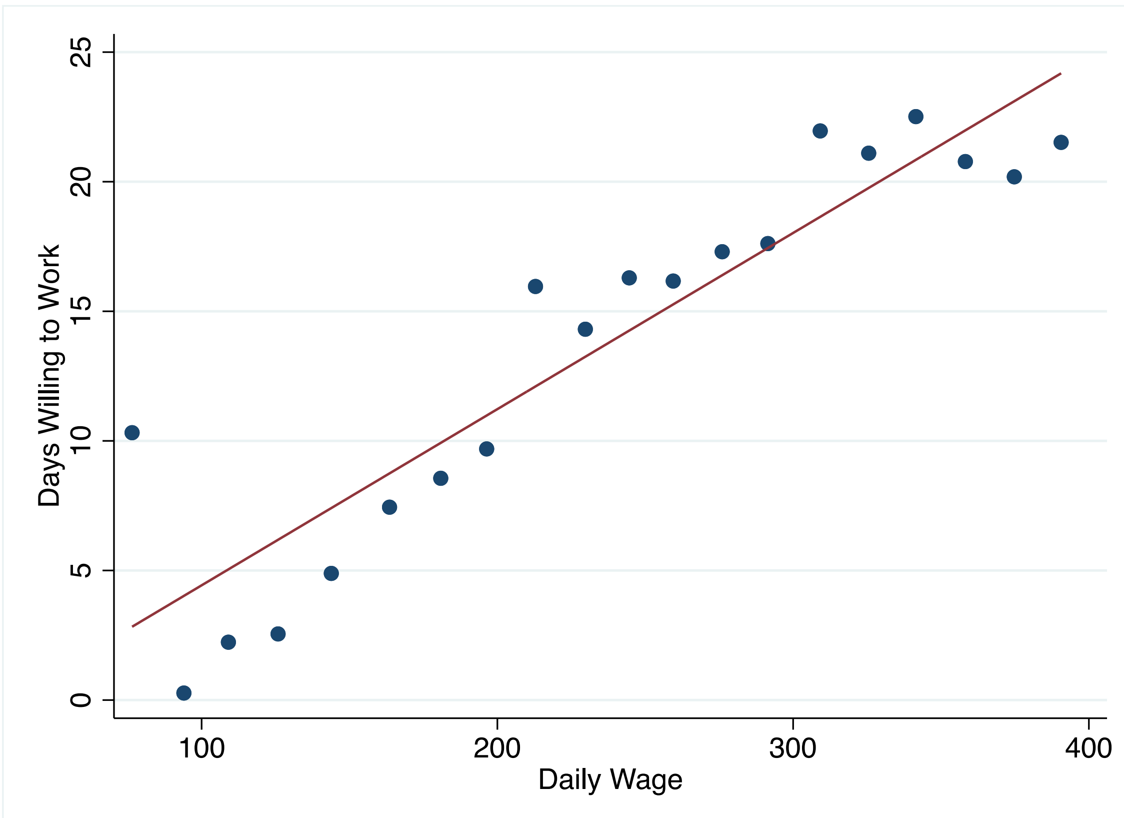
The figure shows the relationship between log non-agricultural wages and monsoon rainfall, at the village level and net of village and year fixed effects. The information for year 1 comes from the follow-up survey, in which a question on non-agricultural wages during harvesting of that year was asked for each household member. The information for year 3 comes from the harvesting phone survey with the sample of laborers. We observe wages for all 200 villages during the year 1 follow-up survey because we asked about each household member, but we only observed non-agricultural work in 94 unique villages for the year 3 harvesting survey. The regression thus has 294 observations. The coefficient from the regression is 0.15 and the t statistic is 2.21.

Figure A4: Wage dispersion within villages



The figure shows kernel densities of residuals from regressions of log wages on village-by-survey fixed effects (left panel) and village-by-survey fixed effects plus gender (right panel).

Figure A5: Estimated labor supply curve from hypothetical choice experiment



The figure shows willingness to work in agriculture at random wage offers in a survey. We drew a random wage from the uniform distribution and asked how many days farmers would be willing to work at that wage over a month's period.