The Agricultural Wage Gap Within Rural Villages

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

September 12, 2022

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. The evidence suggests that compensating differentials rather than sorting on ability are important in explaining the wage gap. We then estimate a discrete choice model of daily labor supply that decomposes preferences for jobs into spatial frictions associated with location and sectoral frictions associated with attributes of jobs. We find that while spatial frictions are significant even within rural areas, sectoral frictions are nearly the same in magnitude and have implications for smoothing shocks. Counterfactual analysis shows that the unemployment effect of a one standard deviation decrease in rainfall would be halved in the absence of sectoral frictions.

*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, 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 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. As of 2016, 35 to 50 percent of rural income was generated in non-farm activities in developing 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 within rural areas. We do so by observing workers who move between sectors within rural areas, and often within the same village. Our analysis proceeds in four steps. First, using a detailed panel of daily labor market outcomes from Jharkhand, India, 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 then investigate whether or not there are barriers that obstruct workers from switching sectors and increasing their wages. In the development economics literature, barriers to urbanization are considered an important source of the sectoral earnings gap (Lagakos, Mobarak, and Waugh, 2018; 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 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, there are no potential gains from switching sectors. A lesser explored hypothesis in developing countries, despite the difficult and precarious nature of casual labor in both rural

[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 followup phone surveys during the planting and harvesting times for that season and the following two seasons.

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.

A novelty of our dataset is that it allows 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 uses these workers for identification by including individual-by-survey fixed effects into the regression. Therefore, 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 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 different 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.

Second, to shed light on other explanations, we posed a simple question to workers. We asked them the top reason for working in agriculture if wages are a bit lower than in non-agricultural jobs. A frequent response stands out: the difficulty of non-farm work. 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 comprise a sectoral friction.

Third, we estimate a discrete choice model of labor supply. We incorporate the possibility of not only sectoral frictions, but also spatial frictions, allowing us to quantify their relative importance. Although migration rates are low, workers sometimes engage in daily casual work outside their village. We account for these two potential sources of frictions by observing people working in different sectors — both inside and outside their villages. The monetary value of the spatial (geographic) friction 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 spatial frictions also matter for the choice of whether to leave the village for casual labor. But more interestingly, we find that sectoral frictions 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, and that they face larger sectoral frictions than spatial frictions.

Fourth, we highlight the implications of sectoral frictions using our estimated parameters in counterfactual simulations. Empirical evidence from India shows that rural populations turn to non-farm employment when agriculture faces negative shocks (Blakeslee, Fishman, and Srinivasan, 2020; Colmer, 2021; Lanjouw and Lanjouw, 2001). In the reduced-form analysis, the same result appears in our data. A one standard deviation decrease in precipitation causes agricultural work during harvesting to fall by 11.3 percentage points, while it increases non-agricultural work by 8.6 percentage points. Meanwhile, this decrease in rainfall increases unemployment by 8.0 percentage points. Yet, the use of non-farm work is an imperfect instrument for employment stability because of sectoral frictions. We use our estimated parameters to predict effects of a rainfall shock on labor flows in the absence of sectoral frictions. Counterfactual analysis shows that without sectoral frictions, the rainfall shock would have less than half the effect on unemployment. In other words, eliminating or lowering sectoral frictions would make non-farm work a more effective instrument for offsetting job losses in agriculture.

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 suggesting that the wage gap can be partly explained by utility costs of doing non-agricultural work. Section 3 outlines a model of daily labor allocation choices in the presence of sectoral frictions. Section 4 estimates the parameters of the model and quantifies sectoral and spatial frictions. 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 casual labor market. This population makes up a non-trivial 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 occupations. 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. Farmers harvest 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.

During the first year (August and November 2014) surveyors tried to contact each laborer. Surveyors asked the laborers whether they 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 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]

[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.

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

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 usually 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 measure the agricultural wage gap while controlling for unobserved heterogeneity across individuals.[4] The people switching sectors give identification. This makes it useful to compare them to the individuals that work only in agriculture. About 20% of the workers from the baseline survey switched sectors. Table 1 shows the differences between these two groups. 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, 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.[5]

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. Around 4 to 8 percent of workers switch sectors during the same survey wave. Using only these workers for identification produces the same results as using people that switch across waves.

We use three more 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, 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 vilage-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

[4] Our main specification includes individual and survey round fixed effects. These estimates could be affected by unobserved time-varying attributes, such as changes in ability or training. We also include 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.

[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.

2014. Third, 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.

2.2 Empirical Approach

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,

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

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 a stricter specification with individual-by-survey round fixed effects. 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). 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.[7] 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 wave 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 fixed effects, limiting the identification to around one fifth of the sample. We find

[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.

that agricultural workers increase their daily wages by 21% when moving to non-agricultural work.[8] This estimate is partly identified off of people switching sectors across survey rounds.

Some workers only switch sectors across survey waves. Others switch sectors within a span of one to two weeks. This allows us to include 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. 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 21%. 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 30% 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).

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 this setting is about 1.23.[9] Moreover, Munshi and Rosenzweig (2016) find that the rural-urban wage gap in India is 25% after adjusting for differences in cost of living.

2.4 Worker Explanations for Avoiding Non-agricultural Work

Before turning to our model, we use survey data to assess a possible explanation for this wage gap. Workers might dislike some attributes of non-agricultural jobs. We posed a simple question to our sample of laborers during the 2016 followup survey. We asked why they would continue to work in agriculture if non-agricultural wages are higher? The answers to this question are not based on revealed behavior. But the responses help give credibility to a model where compensating differentials contribute to the wage gap beween sectors.

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

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

Figure 2 shows that several of the top explanations can be characterized as job-specific disutilities. The top answer is that non-agricultural jobs are "too hard." 23% of workers point to this as a reason for not wanting to close the wage gap between sectors.[10] This evidence does not pinpoint what exactly makes non-agricultural jobs harder. It instead provides suggestive evidence that workers prefer a day of agricultural work over local non-agricultural employment. This could be because non-agricultural jobs are more physically demanding, require longer hours, or involve tasks that are less familiar than agricultural activities.[11] Indeed, non-agricultural work in rural areas often requires physically demanding tasks. During this same survey we asked workers what they do when working in non-agricultural jobs. These jobs involve some form of construction around 68% of the time. Other popular activities include working in local coal mines or brick kilns.

2.5 Other Explanations

Amenities and Transportation Costs. Lost amenities and transportation costs are two common explanations for a nonagricultural wage gap. For example, leaving the farm may cause one to lose valued amenities such as quality housing (Lagakos, Mobarak, and Waugh, 2018) or access to traditional risk-sharing arrangements (Munshi and Rosenzweig, 2016). On the one hand, these are characteristics of the job that workers dislike. On the other hand, they are distinct from attributes of the actual work, such as its difficulty. But we do not believe that these play a key role in the sectoral wage gap in our context.

First, disamenities are most often associated with rural-urban migration. We analyze non-agricultural jobs that workers take within a few days of working in agriculture. This makes it unlikely that lost amenities can account for the gap between wages in the two sectors. 31% of the non-agricultural work days in Table 2 took place in another village. Dropping these observations leads to similar results (Table A3). Using only within-village transitions eliminates the possibility that we can explain the gap by features of locations, rather than jobs.

Second, Figure 2 shows that some workers cite distance from home as a reason for not

[10] 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.

[11] 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.

taking non-agricultural jobs. About 54% of workers (124 out of 230) that switch sectors do so on back-to-back days. Switching sectors on successive days likely does not involve large transportation costs. Yet, we find a similar wage gap in these daily transitions. Compared to back-to-back agricultural days, switching to non-agricultural work increases wages by 17.6%. A transition in the reverse direction decreases wages by 28.6%.[12] These results suggest that transportation costs might not drive the sectoral wage gap.

Non-availability. 16% of workers give 'non-availability' as a reason for not taking non-agricultural work. Some of these answers could come from laborers for whom non-agricultural jobs are never available. For these laborers, the type of jobs they want or qualify for might not be available in their area. However, we estimate the wage gap among switchers, who qualify and do find an adequate job at times. The question is whether there is shortage of jobs for the workers able and willing to work in non-agriculture. The wage gap would in that case reflect a disequilibrium wage, fixed by the demand side (based on a wage efficiency theory, for example), rather than an equilibrium wage where the wage gap reflects a compensating differential. We look at this possibility as one alternative explanation.

We start by looking at how shocks to agricultural labor demand affect non-farm employment. Agricultural labor demand at harvesting depends on local rainfall. Rainfall therefore provides a quasi-random source of variation in agricultural labor demand. We focus on the three harvesting surveys and estimate

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

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 fixed effects. We use total precipitation during the agricultural season for the rainfall variable.[13]

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. Agricultural 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 remaining 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

[12] These estimates come from a regression where the change in log daily wages ($log(wage_{ivtd}) - log(wage_{ivtd-1})$) is regressed on two dummy variables: one for a transition from agriculture to non-agriculture and another for a transition from non-agriculture to agriculture.

[13] 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.

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.

Table 3 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.

This combined with other evidence suggests that there is not a binding constraint on non-agricultural labor demand. In dry years, demand declines for both agricultural labor and own-farm work, suggesting a corresponding increase in labor supplied to non-agriculture. A binding capacity constraint on non-agricultural labor demand would prevent an increase in non-agricultural employment. If such a constraint binds, then a shifting labor supply would not change employment. Two additional findings suggest that changes in labor demand do not drive the increase in non-agricultural employment during dry years. First, large farmers may increase labor demand by shifting to non-farm production during dry years.[14] But very little of the non-agricultural labor demand comes from large farmers. Second, Figure A3 shows that non-agricultural wages decrease during dry years. This aligns with supply shifting.

Search costs. One may be concerned that the 'non-availability' reason could be interpreted as the presence of large search costs for non-agricultural jobs. If search costs are greater for non-agricultural jobs, then they could be part of a compensating differential. For example, Jeong (2021) finds a causal link between search costs and higher levels of wage dispersion. We do not find evidence of high search costs and bargaining in our data. 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.

Overall, although these findings cannot eliminate all possible competing explanations for the within-individual gap in wages between sectors, the other explanations above do not match the wage gap, employment flows, and limited wage dispersion that we see in the data. Building on this, the model that follows separates spatial frictions (within sector) from sectoral frictions. This allows us to compare the magnitude of spatial frictions, a commonly cited source of an earnings gap in the development literature, with job-specific amenities, a lesser explored source.

[14]Our year 3 followup survey asked the large farmers how much they hire for *non-agricultural tasks*. The average amount hired was 0.3 person days a month. Laborers report working an average of over 6 non-farm person days during the same months.

3 A Model of Rural Labor Supply

Our analysis up to this point suggests there is disutility in doing non-farm work. This explanation implies that the wage gap reflects a compensating differential. Workers have a disutility for characteristics of *rural* non-agricultural work. We refer to the non-agricultural disutility as a sectoral friction.

We first build a model showing that after controlling for selection on unobserved productivity, the gap in average wages between sectors is an increasing function of the sectoral frictions. Building on this, we estimate a discrete choice model of labor supply. To benchmark the size of the sectoral friction to a common source of wage variation we also model spatial frictions, but within sectors. In modeling spatial frictions, we assume that workers have disutility for working outside the village. The two mechanisms have different implications for candidate policy responses in reducing frictions. For instance, some policies could make searching for work outside the village easier. These policies will do little if sectoral frictions explain why workers stay in agriculture.

The model quantifies the importance of the two types of frictions. In the model, workers choose between the two sectors. Within each sector, they choose whether to leave the village. We use this feature of the data to separate the sectoral and spatial frictions.

3.1 A Simple Model Linking Wage Gap to Disutility for Non-agricultural Work

To illustrate the key idea of sectoral frictions in our setting, we first present a simple model of occupational choice. Workers decide jobs based on wage offers and individual preferences for sectors. To highlight the key insight, we ignore spatial frictions for now.[15] The model shows that for the people switching sectors, there is a positive association between the sectoral wage gap and the disutility for non-agricultural jobs. This shows that part of the wage cap can be a sectoral friction.

Consider a simple labor market with two competitive sectors: agriculture (a) and non-agriculture (n). Suppose that all workers are equally qualified for jobs in either sector. This assumption matches our estimated wage differential on switchers and with individual fixed effects. For simplicity, assume that workers (indexed by i) receive wage offers from the two sectors that are independently distributed with $w_{ij} = \overline{w} + \epsilon_{ij}$, where $\epsilon_{ij} \sim N(0, \sigma_j^2), j \in \{a, n\}$, and that they have a (negative) preference for working in the non-agricultural sector δ_i. Utility for worker i in sector j is: $U_{ij} = \overline{w} + \delta_i I_n + \epsilon_{ij}$, where I_n is an indicator for choosing

[15] We allow for spatial friction in the structural model we estimate in the next subsection.

sector n. Upon receiving offers from both sectors, the worker selects to work in the sector that gives the maximum utility. Then worker i will choose sector n if $\nu_i \equiv \epsilon_{in} - \epsilon_{ia} > -\delta_i$. Note that $\sigma_\nu^2 = \sigma_a^2 + \sigma_n^2$.

Hence the probability that worker i works in sector n is:

$$\text{Prob}(i \text{ works in } n) = \text{Prob}(\nu_i > -\delta_i) = \Phi\left(\frac{\delta_i}{\sigma_\nu}\right),$$

and its expected wage if he chooses to work in sector n is

$$\begin{aligned} E(w_i \mid i \text{ works in } n) &= \overline{w} + E(\epsilon_{in} \mid \nu_i > -\delta_i) \\ &= \overline{w} + \frac{\sigma_n^2}{\sigma_\nu} \frac{\phi(\frac{\delta_i}{\sigma_\nu})}{\Phi(\frac{\delta_i}{\sigma_\nu})}, \end{aligned}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the density and cumulative distribution functions of a standard normal.[16] Worker i's expected wage is an increasing function of disutility for non-agricultural work $-\delta_i$. In other words, workers with lower preference (higher distaste) for non-agricultural work need to be compensated with a higher expected wage to work there. The average wage among all workers who choose to work in sector n is

$$\overline{w}_n = \overline{w} + \frac{\sigma_n^2}{\sigma_\nu} \int \frac{\phi(\frac{\delta_i}{\sigma_\nu})}{\Phi(\frac{\delta_i}{\sigma_\nu})} f(\delta_i) d\delta_i,$$

where $f(\delta_i)$ is the distribution of disutility for non-agricultural work, so the average (observed) non-agricultural wage is an increasing function of a monotonic shift in disutility. There is hence a positive correlation between the wage gap and the disutility of working in non-agriculture.

3.2 A Model of Rural Labor Supply with a Full set of Occupational Options

Based on this simple framework, we estimate a discrete choice model of rural labor supply. The model has the full set of occupational choices. These 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 separate sectoral from spatial frictions. Sectoral frictions limit move-

[16] The last step uses a property of conditional normal distribution that $E(\epsilon_{in}|\nu) = \frac{\sigma_n^2}{\sigma_\nu^2}\nu$. The term $\frac{\phi(\cdot)}{\Phi(\cdot)}$ is known as the Inverse Mills Ratio, which is a decreasing function of its argument.

ment between job types. We allow the sectoral frictions term to be a function of worker characteristics. It is also randomly distributed across workers. Spatial frictions limit movement from inside to outside the village.

An ideal dataset would have random wage offers for both sectors. In practice, we only observe wages for the chosen options. 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.[17] In addition to these labor opportunity variables, occupational choices depend on past choices. This accounts for potential path dependency and/or transactions costs in changing jobs. We also allow worker idiosyncratic preferences for alternative options. These preferences 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 denotes block, d denotes day, and t denotes 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 rainfall (collected at the block-survey level), $Harvest_t$ and $Planting_t$ are indicators for whether the survey round occurs in the harvesting or planting season, I_j is an indicator vector for each employment choice, δ_i is a vector of random disutility terms, with $\delta_i \sim N(\mu, \Sigma)$, X_i is a vector of farmer characteristics such as gender and whether the worker belongs to a marginalized caste group, y_{ijtd-1} is an indicator variable for whether option j is chosen on previous day, and ϵ_{ijtd} a random component that takes on Type-I Extreme Value distribution. Note that the disutilities δ_{ij} can be potentially correlated across employment options, allowing for flexible substitutional patterns across disutility terms.

Estimating the distribution of work disutility. Conditional on the random disutility vector δ_i, the probability of choosing job options is the same as the conditional logit model:

$$P_{ijtd}(\delta_i) = \frac{\exp(V_{ijtd})}{\sum_{j'} \exp(V_{ij'td})}.$$

[17] 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.

Since worker disutility δ_i is uncertain, we integrate the above probability over all possible values of δ_i in order to obtain the employment probability. Since the integral has no closed-form solution and cannot be computed analytically, we approximate the probability through simulation. The simulated likelihood for farmer i over all choice occasions on day d in survey t is then

$$L_i = \prod_t \prod_d \sum_{j=1}^{J} y_{ijtd} \widehat{P}_{ijtd},$$

where $\widehat{P}_{ijtd}$ is the simulated conditional probability and y_{ijtd} is an indicator that takes on the value 1 for the chosen alternative and 0 otherwise. The overall simulated log-likelihood function is $\sum_{i=1}^{N} \log L_i$, which is maximized for estimation.

3.3 Data for Model Estimation

Our data describe the daily activities of casual workers from 12 blocks over 3 seasons. There are two distinct periods per season at planting and harvesting times. The data has 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 season. 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. Random variations in utility can be interpreted as random variation in wage offered. We capture variation across seasons with the seasonal and weather variables. We capture individual variation with variation in preferences, from the characteristics X_i, and the random shock in preference 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 Results

Table 4 reports the estimated parameters and their standard errors. The columns report coefficients for agricultural work inside the village, agricultural work outside the village, nonagricultural work inside the village, nonagricultural 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 spatial and sectoral frictions. For example, within villages, workers have a much higher disutility for non-agricultural work than for agricultural work. This disutility is lower for ST or SC workers, 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 report that it's because non-farm work is difficult in contrast to 18% of ST or SC workers. Male 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 your own field and doing agricultural work in the village shown in Table 4. The correlation suggests that there are complementarities between these two choices. Finally, compared to males, females have a greater disutility for non-agricultural work and leaving the village. But in contrast to males, this spatial friction is higher for non-agricultural work. This suggests that sectoral frictions may explain more of the gender earnings gap.

4.1 Quantifying Frictions

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 frictions into monetary terms. Instead, we quantify the frictions 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 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 this data, we compute the increase in daily wages that would have the same effect on labor supply as eliminating the frictions.

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 spatial friction, we let j be agriculture outside the village, and for sectoral friction, 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 time fixed effects. Regression results from estimating Equation 3 are shown in Table 5. A one standard deviation increase in rainfall raises agricultural wage by 20 rupees.

The sectoral friction is equivalent to $\theta \Delta W_j$, with j being non-agricultural work inside the village. Similarly, spatial friction is measured by choosing j as agriculture outside the village.

Table 4 shows estimates for average disutility $\hat{\mu}_{\text{ag inside}} = -.323$, $\hat{\mu}_{\text{nonag inside}} = -2.543$, $\hat{\mu}_{\text{ag outside}} = -3.878$, and $\hat{\alpha}_{\text{ag inside}} = 1.097$. Hence for male, non SC/ST workers:

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

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

Similar calculations yield sectoral and spatial frictions for female, non SC/ST workers to be 66.4 and 51.4 Rs, respectively. Male agricultural workers in our survey earned an average of 205 Rs per day, while female wages average 140 per day. The sectoral friction amounts 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, we calculate frictions using the estimated labor supply curve. We first eliminate the sectoral friction. We do this by letting the non-agricultural disutility have the same distribution as that of agricultural work.[19] This

[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.

[19] 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.

holds spatial frictions constant, but removes sectoral frictions. Denote the change in choice probability by $\Delta P_{\text{nonag inside}}$.

To quantify the spatial frictions, we remove them from the model in a similar way. Specifically, we decrease the disutility of agricultural work outside the village to that of agricultural work. This eliminates the spatial friction, but keeps sectoral frictions constant. Denote this change in choice probability by $\Delta P_{\text{ag outside}}$.

Table 6 shows these changes in the choice probabilities. Eliminating the sectoral friction inside the village increases non-agricultural work by $\Delta P_{\text{nonag inside}} = 9.4$ pp (row 1, column 3). Eliminating the spatial friction for agricultural work 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.

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 compute a wage equivalent change for any changes 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 spatial and sectoral frictions for male, non SC/ST workers as follows:

$$\text{Sectoral friction} = \beta_{\text{wtw}} \Delta P_{\text{nonag inside}} = 3.877 \times 9.4 = 36.4 \text{ rupees.}$$
$$\text{Spatial friction} = \beta_{\text{wtw}} \Delta P_{\text{ag outside}} = 3.877 \times 13.1 = 50.8 \text{ rupees.}$$

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 sectoral and spatial frictions are of the same order of magnitude for the revealed and stated preference approaches.

4.2 Counterfactual Analysis

Our counterfactual analysis looks at whether sectoral frictions amplify effects of agricultural shocks. Workers use non-farm work to smooth productivity shocks in the agricultural sector. Thus, a disutility of non-farm work exacerbates the effect of productivity shocks on unemployment.

We use our model estimates to simulate labor market adjustments to a 1 standard deviation decrease in rainfall. The model estimates in Table 4 permit us to compare adjustments under two scenarios: i) a baseline scenario where the sectoral friction parameter is set at its estimated value, and ii) a counterfactual scenario where sectoral frictions are eliminated. The changes in probabilities of choosing various occupations for the baseline scenario are shown in row 3 of Table 6. If rainfall decreases by 1 standard deviation in the harvest season, the shares of farmers working in agriculture inside, agriculture outside, and their own fields will decrease by 11.3, 2.6, and 10.3 percentage points. The share of people working in their own fields drops by a similar magnitude. This is partly due to the high correlation between agricultural work inside the village and working in own fields (the correlation coefficient is 0.322 according to estimates in Table 4).

Some workers will move to non-agricultural work in response to the shock. In particular, the share employed in non-agriculture inside and outside the village increases by 8.6 and 7.7 percentage points, respectively. In line with the regression evidence above, the unemployment share increases by 8.0 percentage points. Overall, the baseline case shows that unfavorable weather conditions lead to more non-agricultural work, but also increase unemployment.[20]

How much more responsive would non-agricultural employment or unemployment shares be to rainfall shocks if the non-agricultural work disutility disappeared? We consider the case where sectoral frictions are eliminated, i.e. when the sector-specific utilities are equal. First, the second row of the table shows that under normal rainfall the share working in non-agriculture inside the village would increase by 9.4 percentage points. Add a rainfall shock, and non-agricultural employment increases by 24 pp (row 4). Hence, the rainfall shock has a much larger effect on non-agricultural employment in the absence of sectoral frictions: it increases the share of non-farm workers inside the village by $24 - 9.4 = 14.6$ percentage points. As a result, the unemployment effect of the shock is less severe. Eliminating sectoral frictions causes the unemployment effect of the rainfall shock to decrease from 8 percentage points (row 3) to 3.6 percentage points (row 4 minus row 1).

These results show that the non-agricultural disutility exacerbates the effect of agricultural shocks on local employment. While workers do move to non-agricultural work when agricultural demand falls, which the model predicts and is observed in the reduced-form regressions, this response is partially muted by the disutility associated with doing these jobs. A reduction in this disutility would mitigate the impact of agricultural shocks on local employment.

[20] As a comparison, the rainfall shock from third year of good rainfall to first two years of drought is -1.67 standard deviations. In the data, this shock induces a decrease of 13.7 percentage points in agriculture inside share and 14.9 percentage points in working own fields, and an increase of 10.0 percentage points in non-agriculture inside, 9.3 percentage points in non-agriculture outside, and 15.5 percentage points in employment shares. So the simulated effect is reasonable in terms of magnitude.

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 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 structural model of labor allocation across sectors to quantify this disutility. The model estimation shows that frictions associated with non-farm work even within the same village amount to about 20% of the daily wage for males and 47% for females. Moreover, reducing or eliminating these frictions would make non-agricultural work a more effective instrument for stabilizing employment after agricultural shocks.

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.

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.

Blakeslee, David, Ram Fishman, and Veena Srinivasan. 2020. "Way down in the hole: Adaptation to long-term water loss in rural India." *American Economic Review* 110 (1):200–224.

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.

Colmer, Jonathan. 2021. "Temperature, labor reallocation, and industrial production: Evidence from India." *American Economic Journal: Applied Economics* 13 (4):101–24.

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.” Working Paper.

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. 2018. “The welfare effects of encouraging rural-urban migration.” National Bureau of Economic Research w24193.

Lanjouw, Jean O and Peter Lanjouw. 2001. “The rural non-farm sector: issues and evidence from developing countries.” *Agricultural Economics* 26 (1):1–23.

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
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**

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	0.207***	0.211**	0.305***	0.325***	0.325***
	(0.041)	(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 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 3: 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 gender of the respondent, based on the 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 4: 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 of Ag Inside (Σ)		0.124***	0.148***	0.085***	0.322***
		(0.042)	(0.039)	(0.031)	(0.028)
Corr of Ag Outside(Σ)			0.077**	-0.050	0.019
			(0.038)	(0.036)	(0.034)
Corr of Non-Ag Inside(Σ)				0.209***	0.153***
				(0.030)	(0.034)
Corr of Non-Ag Outside(Σ)					0.090***
					(0.032)
Harvest (α)	-0.425***	-0.544***	-0.476***	-0.459***	0.132***
	(0.025)	(0.052)	(0.051)	(0.061)	(0.021)
Rainfall Harvest (α)	1.097***	1.863***	-1.190***	-1.569***	0.634***
	(0.021)	(0.046)	(0.044)	(0.052)	(0.017)
Rainfall Planting(α)	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)
Last Choice (γ)	1.074***	1.074***	1.074***	1.074***	1.074***
	(0.011)	(0.011)	(0.011)	(0.011)	(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, nonagriculture 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 5: 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 FE	Yes	Yes	No	No
Block FE	No	No	Yes	Yes
Year FE	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.

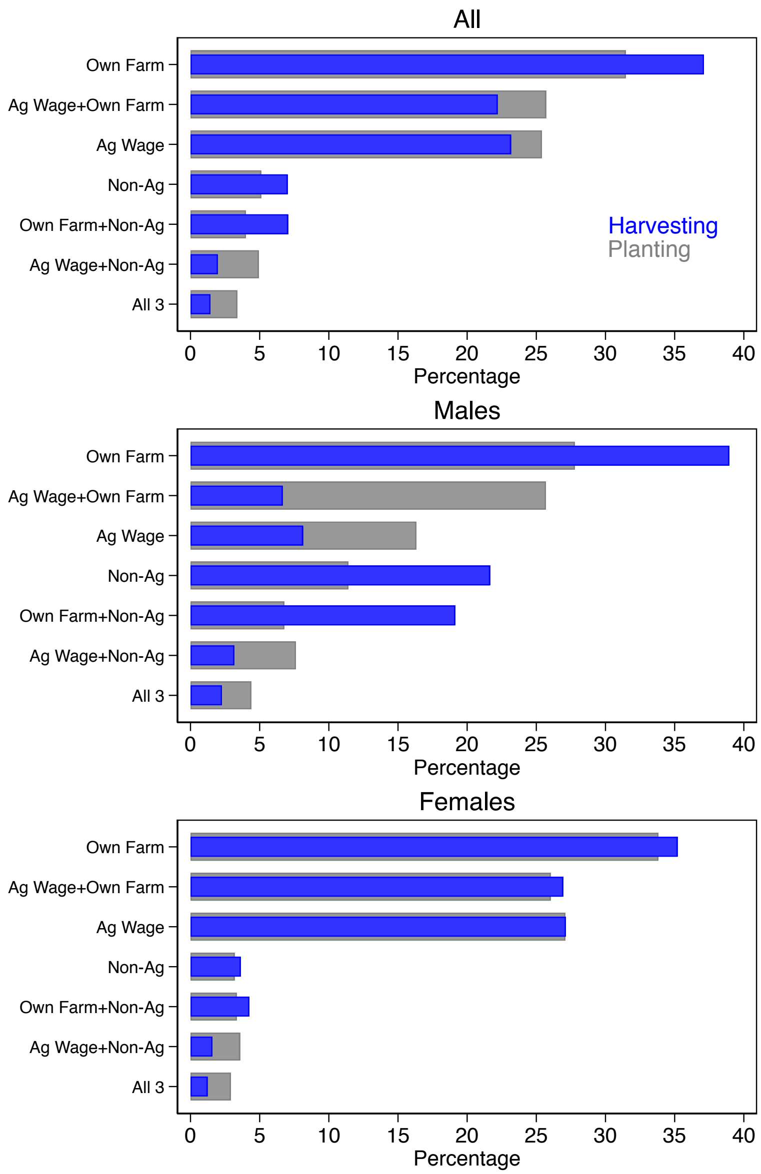
Table 6: Changes in choice probabilities in different scenarios

	Agriculture		Non-Agriculture		Work in	Unemployed
	Inside	Outside	Inside	Outside	Own Field	
	(1)	(2)	(3)	(4)	(5)	(6)
No sector friction	-0.021	-0.003	0.094	-0.002	-0.040	-0.028
No geographic friction	-0.029	0.131	-0.005	-0.003	-0.053	-0.041
Rain shock in baseline case	-0.113	-0.026	0.086	0.077	-0.103	0.080
Rain shock without sectoral friction	-0.124	-0.026	0.240	0.054	-0.151	0.008

The table shows changes in the choice probabilities relative to the base line under different scenarios. Rows 1-2 report results for the scenarios where sectoral or geographic frictions are eliminated, i.e., when the distribution (mean and SD) of the disutility of agriculture outside or nonagriculture inside labor, respectively, is the same as that of agriculture inside. Rows 3-4 report results for a 1 SD drop in rainfall in the harvest season, in scenarios with and without sectoral friction.

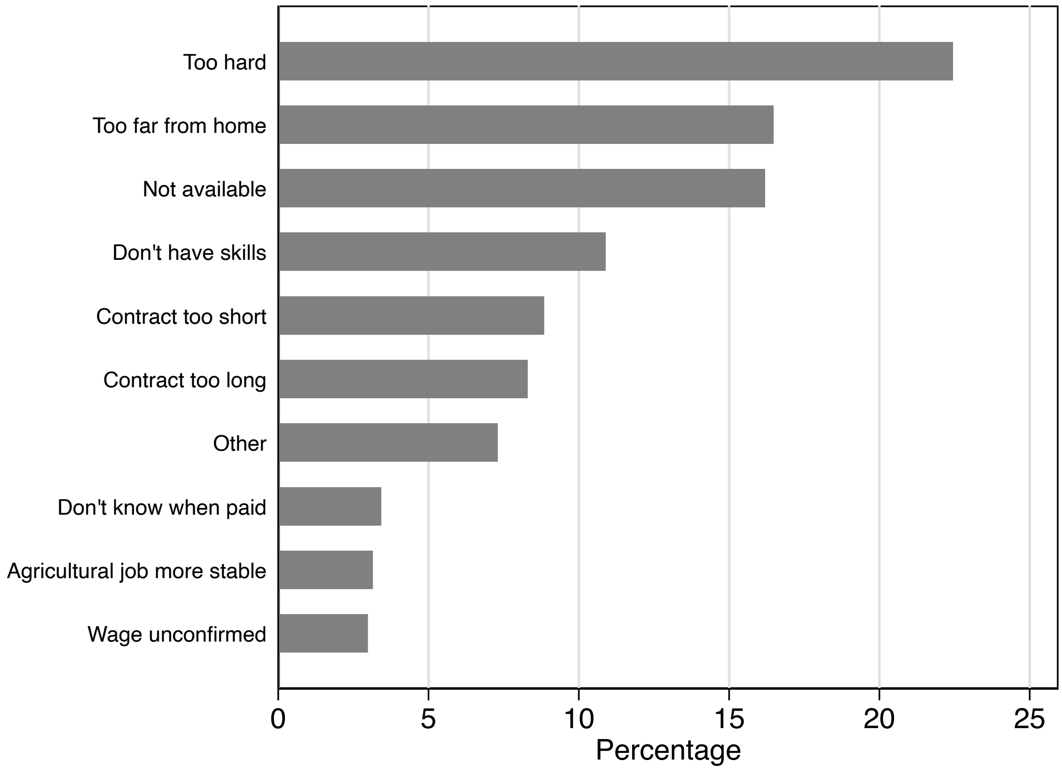
Figures

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



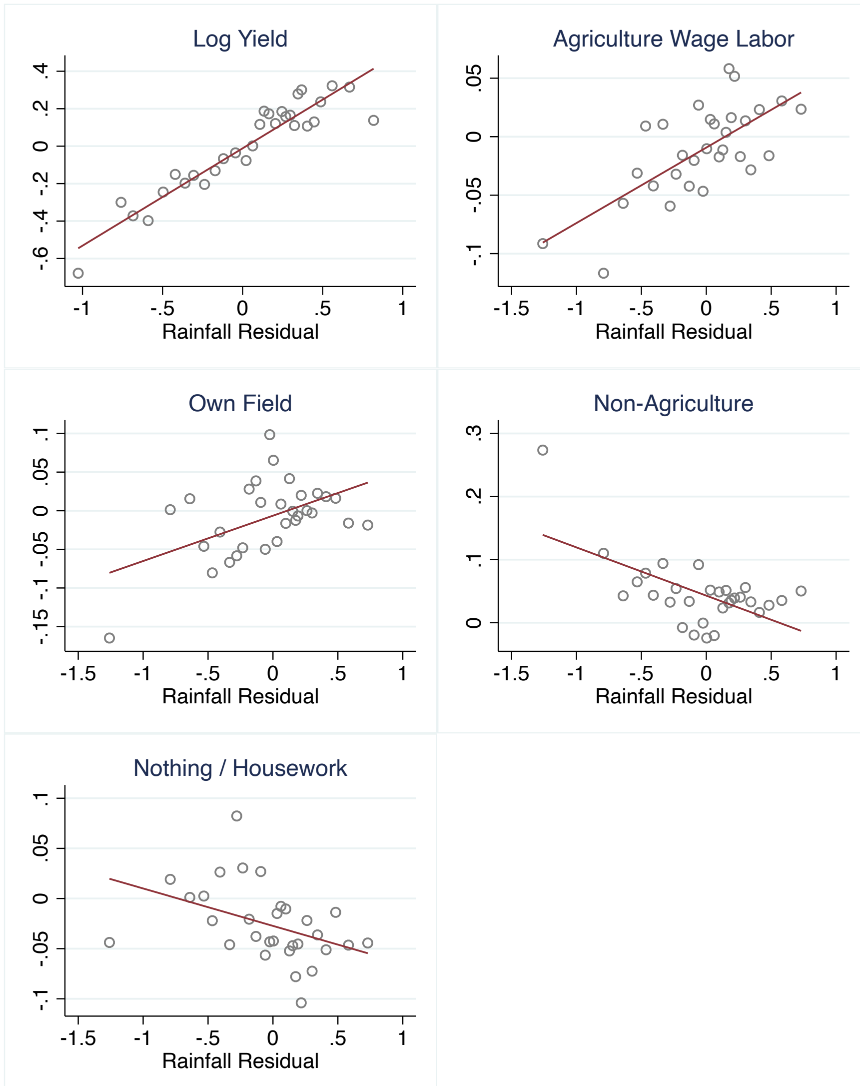
The figure shows a classification of workers into seven groups, depending on which activities they did during the 7 or 14 day survey period. The top panel is for all respondents and is weighted by gender to represent the sex ratios of the population of agricultural workers hired by large farmers. The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

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

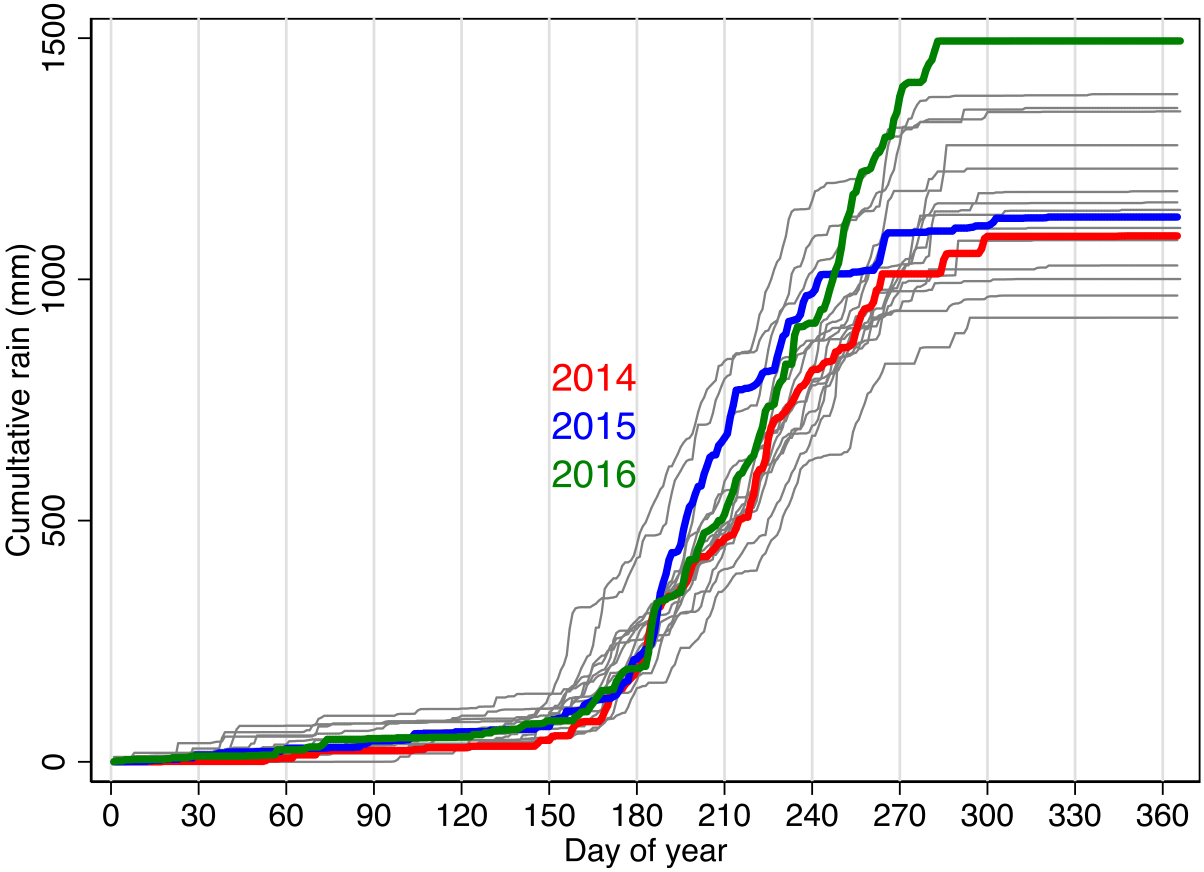
The data are from the survey with farmers after the 2014 season. Farmers were asked for male and female wages, separately by task and gender. Farmers were also asked for the length of a typical work day by gender and task. The dependent variables are the log of male wages (columns 1 and 2) and the log 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 (1)	Individual by Survey (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

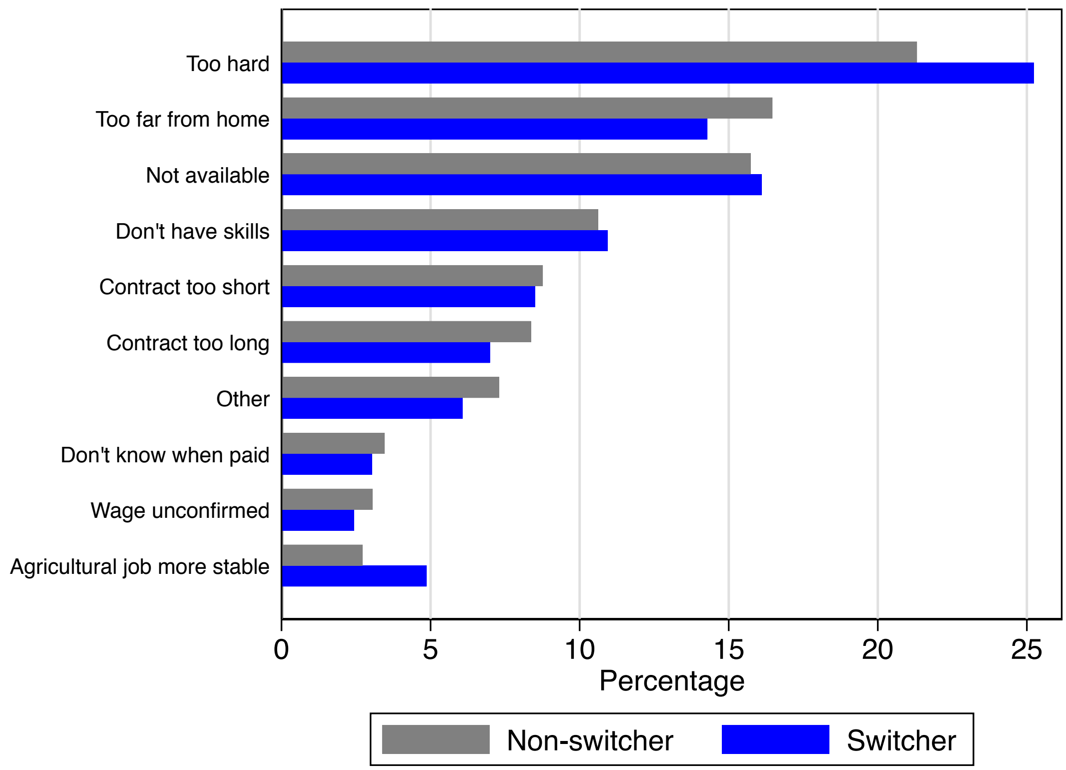
The data are from three surveys where non-agricultural wages were collected: planting time of 2015, and the planting and harvesting surveys of 2016. This table drops days of non-agricultural work which were classified as outside the village (either migrant labor or when the work was outside the village). 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



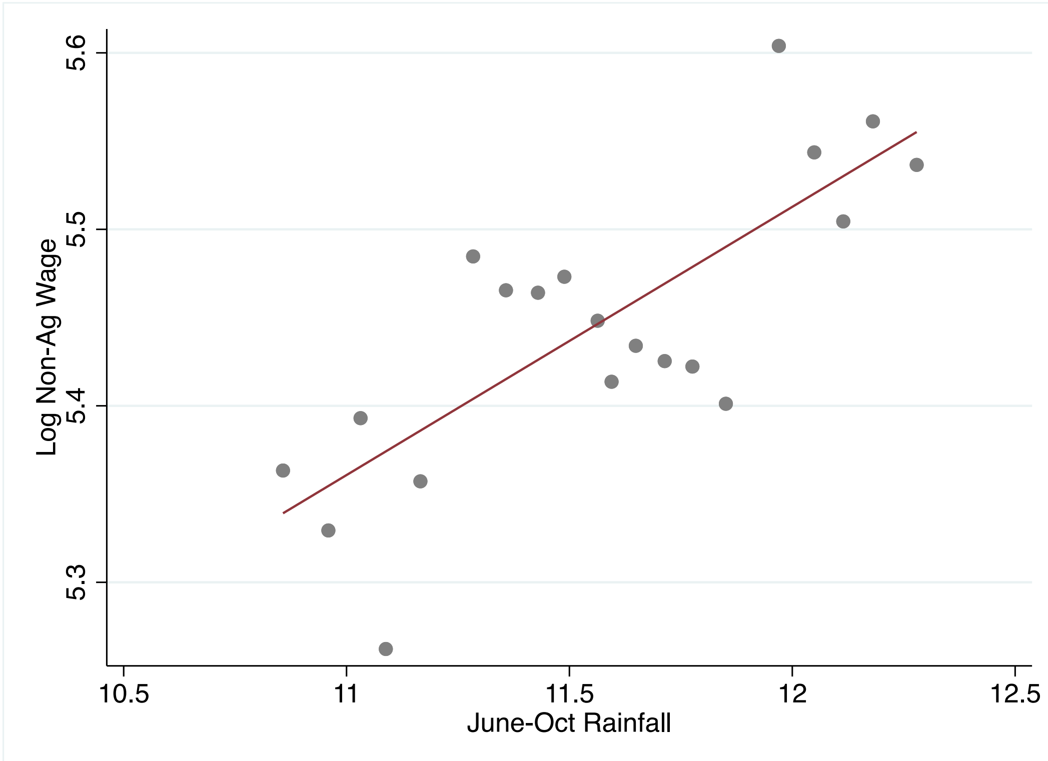
The figure shows cumulative rainfall plotted against the day of the year. Each line is for a separate year. Daily rainfall was first averaged across the 200 sample villages to generate a daily average precipitation for the sample area. The daily rainfall values are satellite observations taken from CHIRPS.

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



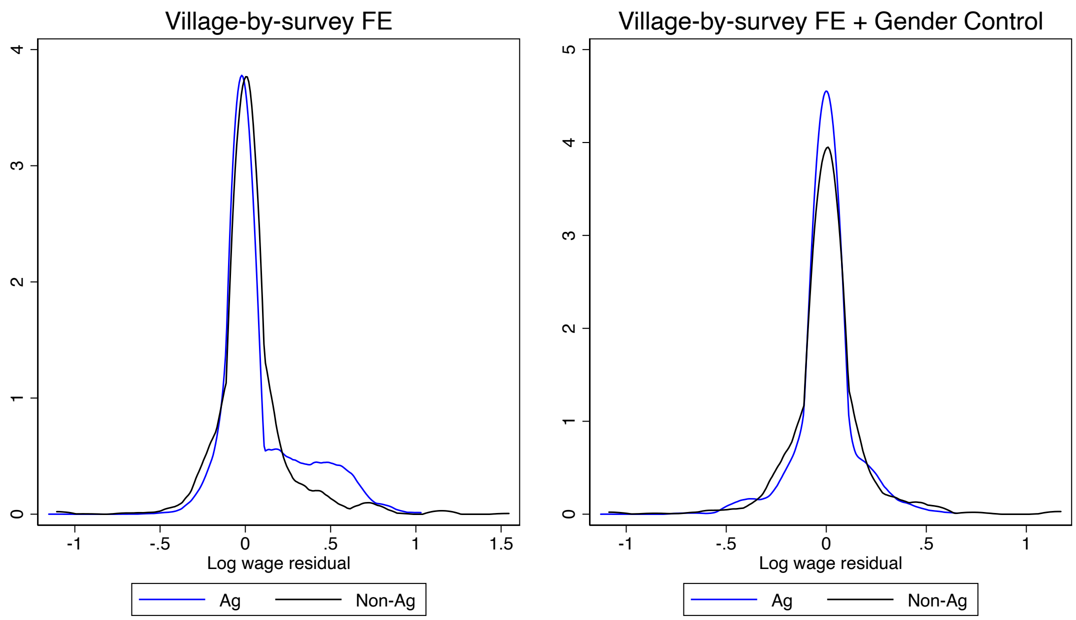
The figure shows responses from the third followup survey with agricultural laborers. The exact question posed to laborers was "Suppose wages are a bit lower for agricultural jobs than for non-agricultural jobs, what is the top reason why you may still work in agricultural jobs". The grey bars are for 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



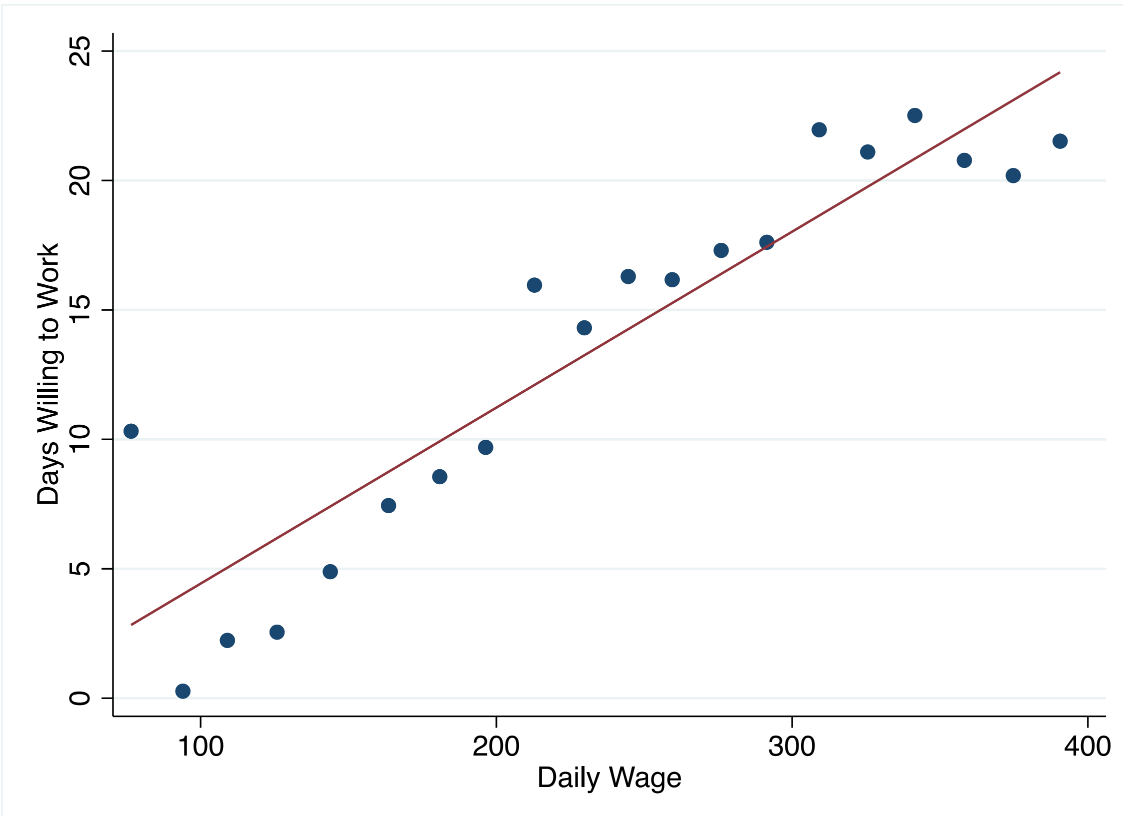
The figure shows the relationship between log non-agricultural wages and monsoon rainfall, at the village level and net of village and year fixed effects. The information for year 1 comes from the followup survey, in which a question on non-agricultural wages during harvesting of that year was asked 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 followup 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.