Sharing a Workforce: the Effect of Agricultural Productivity Shocks on Industrial Performance

(Job Market Paper Draft)

Giulia Zane*
London School of Economics & STICERD
September 16, 2015

Abstract

In countries undergoing structural transformation, manufacturing and agricultural sectors compete over the same labour force. The large supply of workers during the lean agricultural seasons allows manufacturing firms to pay relatively low wages but, during the peak season, workers may find casual jobs in agriculture attractive and leave the firm temporarily. Using firm level data from India, I analyse to what extent the competition from the agricultural sector acts as a constraint for manufacturing sector firms. I find that workers’ attendance is lower during the months of harvest and more so in states that have a more “pro-worker” labour protection regulation. I exploit exogenous shocks to labour demand in agriculture to estimate the elasticity of workers’ attendance with respect to the local agricultural wage. Moreover, I estimate the cost of this phenomenon in terms of output and productivity loss, finding large effects.

*Email: g.zane@lse.ac.uk.
1 Introduction

The classical growth literature characterises the process of development as a transition from an economy dominated by agriculture to one dominated by manufacturing and services (Kuznets, 1957). The first step of this structural transformation is the movement of workers away from the primary sector and into industry (Lewis, 1954). In this phase, manufacturing firms benefit from the agricultural “labour surplus” that provides them with a large labour supply at a low cost. However, the labour requirements in agriculture are not constant over the year, on the contrary, they are highly seasonal. A labour surplus during the lean agricultural months is counterbalanced by a labour shortage during the peak season. Consequently, higher wages are offered for occasional jobs during these periods. In these circumstances agricultural wages provide an attractive outside option for manufacturing workers that may temporarily leave their jobs and return to the fields.

The purpose of this paper is to analyse to what extent the competition from the agricultural sector acts as a constraint for manufacturing sector firms. After verifying that workers attendance is lower during the months of harvest, I exploit exogenous shocks to labour demand in agriculture to identify the response of workers’ attendance to changes in their outside option. Moreover, I estimate the cost of this phenomenon in terms of output and productivity loss.

This study is based on firm level data representative of the entire Indian registered manufacturing sector, collected through 8 rounds the Annual Survey of Industry (ASI), covering the period from 2000-01 to 2007-08. They contain a rich amount of information on firms characteristics and productive activity, which has been widely used by researchers in economics1. Moreover, they provide a unique source of data on workers’ attendance behaviour. Indeed, firms report their attendance data on a monthly frequency2, which allows me to test whether attendance patterns are related to the agricultural production cycle.

India is an ideal setting for this research: its economy is still dominated by the agricultural sector, according to the Population Census, in 2001 it was employing 53% of the labour force, but its future growth crucially depends on the development of its manufacturing sector. Moreover, the size of the country makes it possible to exploit local shocks to agricultural productivity, affecting local labour market outcome (Jayachandran, 2006), while controlling for aggregate

---

1These include Besley and Burgess (2004) and Hsieh and Klenow (2009).
2Although only for 4 months of the year: March, June, September and December.
shocks that affect the whole economy and in particular the product market. Finally, the existence of differences in labour regulation across states makes it possible to shed some light on the role of labour protection in determining workers’ attendance behaviour.

The problem of workers’ absenteeism in developing countries has attracted a lot of researchers’ attention in recent year (see Chaudhury et al. (2006) and Banerjee and Duflo (2006) for a review). While this strand of literature focuses on public sector workers, whose absence rates have been found to be extremely high, absenteeism seems to be an important issue also in the private sector. Indeed, in the Indian manufacturing sector 8.79%\(^3\) of working days are lost every year because of workers absence, a very high rate if compared to 1.2% in the US\(^4\).

Absenteeism of public sector workers is usually explained by the lack of monitoring and punishment, issues that should concern the private sector to a much lesser extent. However, the evidence suggest that workers’ outside options may play an important role also in this context\(^5\). Moreover, the difficulty to punish absentee workers is an issue also for private sector firms: employers can reduce the workers’ wages up to 100% for the days in which they are absent, but firing is virtually impossible, especially in larger firms\(^6\).

A large body of research has explored the role of employment protection regulation on firms’ outcomes (Besley and Burgess (2004), Ahsan and Pagés (2009) and many more), most of which has focused on the Indian manufacturing sector. The main finding of these studies are that having more “pro-worker” labour regulation has a negative effect on firms’ output, employment, investment and productivity. This paper contributes to this literature by exploring one channel in which labour protection may affect firms’ productivity: firms’ located in “pro-worker” states have a lower possibility to punish workers, this may reduce the cost of absenteeism for workers and encourage them to take advantage of outside opportunities more often. Indeed, I find that absence rates in “pro-worker” states are higher and more responsive to changes in the outside option.

Finally, this paper contributes to the literature on the effect of workers absence on productivity.

---

\(^3\)Estimate based on ASI 2009 data.

\(^4\)CPS 2012 data.

\(^5\)According to Chaudhury et al. (2006) differences in outside option may explain why health workers are more absent than teachers and why men are more absent than women.

\(^6\)Firms employing more than 100 workers are subject to section V-B of the 1947 Industrial Disputes Act which sets the requirement of government permission even for firing a single worker.
Although this effect is usually believed to be negative, as it creates an extra cost for the firm in terms of reassignment of tasks to workers and substitutes (Allen, 1983), it is not necessarily so: if the present workers work harder to cover for the absent ones there may be an increase in productivity (measured in output per hour actually worked). Moreover, the firms in this context do not have to pay absentee workers, so the savings in the wage bill may result in higher profits. It remains, therefore, an empirical question largely unexplored by the literature. The major difficulty is the presence a clear endogeneity problem: less productive firms are more likely to face higher level of absenteeism. This may happen for several reasons: for example poor management quality may reduce both productivity and workers motivation; alternatively, firms facing a low demand shock may encourage workers to be absent. One solution to this problem is to use workers’ rather than firms’ productivity. However, this is difficult to measure and often unavailable so researchers have focused on teachers’ productivity, measured in terms of students outcomes (Miller et al. (2008) and Herrmann and Rockoff (2012)). Another possibility is to use exogenous sources of workers’ absence, as in the study of Krueger and Mas (2004), who estimate the effect of strikes on the quality of output.

In this paper I exploit exogenous shocks to workers outside option to estimate the effect of absence on productivity. Since workers are more likely to be absent when there are more work opportunities in the agricultural sector, I can use weather shocks that affect agricultural productivity as an instrument for workers attendance. This identification strategy allows me to estimate the effect of a particular type of absence: the one caused by a temporary movement of workers back into the agricultural sector.

The rest of the paper proceeds as follows: Section 2 proposes a theoretical framework; Section 3 describes the data; Section 4 estimates the response of agricultural wages to shocks to agricultural productivity; Section 5 estimates the effects of changes in workers outside option on workers’ attendance; Section 6 estimates the impact of absenteeism on productivity; Section 7 concludes.

2 **Theoretical Framework**

I propose a simple two sector model in which workers employed in the manufacturing sector have the option to be absent and work in agriculture. The manufacturing firm’s production function is assumed to be Cobb-Douglas, modified to take into account absenteeism. The firm’s profits
can be written in the as follows:

$$\pi_M = A(L_M(1 - a)^\gamma K^\beta - (1 - a)w_M L_M - rK)$$

(1)

where $a$ represents the workers’ absence rate; $L$ and $K$ represent labour and capital, respectively; $w$ is wage and $r$ is rental rate of capital. I assume that workers are payed only when present and, when they are absent they earn the current agricultural wage. Capital is assumed to fixed and exogenous.

The actual amount of labour used is $L(1 - a)$: the number of workers employed multiplied by the attendance rate. The coefficient $\gamma$ represents the effect of absenteeism on output. In particular, if $\gamma = 0$ absenteeism has no effect on output, suggesting that the firm can fully adjust; whereas, if $\gamma = 1$ there is no adjustment and the effect of absenteeism is equivalent to the effect of a change in the number of workers; if $\gamma < 0$ productivity increases with absence rate, as in the case in which the present workers exert extra effort to compensate for the absent ones; finally, if $\gamma > 1$ absenteeism has a negative effect on productivity.

The value of $\gamma$ is estimated in the empirical section of this paper and appears to positive and large. For this reason, and to avoid the trivial results in which the firm has no incentive to incentivise workers’ attendance, I restrict my attention to the case in which $\gamma \geq 1$.

The agricultural sector production function requires only labour input and is characterised by constant returns to scale. Agricultural productivity is represented by a parameter $\theta$, which varies seasonally: during the lean season, which lasts for a fraction $\rho$ of the time, agricultural productivity is low ($\theta = \underline{\theta}$); whereas, during the peak season, or for the remainder $1 - \rho$ of the time, agricultural productivity is high ($\theta = \bar{\theta} > \underline{\theta}$). The agricultural sector profits can be written as:

$$\pi_A = \theta L_A - w_A L_A$$

(2)

Finally, I assume that the labour market in the agricultural sector is competitive and fully flexible. Therefore, the agricultural wage in each period is equal to the current agricultural productivity $w_A = \theta$ and all workers can find a job in agriculture for this wage. Manufacturing sector workers will decide to be absent and work in agriculture when $w_M < w_A$. If the size of the labour force is $L$ then employment in agriculture will be $L_A = L - L_M(1 - a)$. 

5
2.1 Flexible wages and labour

To compute the flexible market outcome I assume that the manufacturing firm can adjust both wages and labour in the short run. In this case the manufacturing wage will always be equal to the agricultural wage $w_M = w_A = \theta$ and and workers will never be absent. The profit maximising labour demand in the manufacturing sector is given by:

$$L^*_M(\theta) = \left( \frac{A\alpha K^\beta}{\theta} \right)^{\frac{1}{1-\alpha}}$$  \hspace{1cm} (3)

which is decreasing in $\theta$. Therefore, the firm will decrease its labour force during the peak season in agriculture and increase it during the lean season. Similarly, manufacturing output will be lower during the peak agricultural season but this is efficient because the agricultural sector, which is more productive will produce more.

2.2 Flexible wages and fixed labour

In practice, labour market regulations make it hard for manufacturing firm to adjust their labour force in the short run. In this subsection, I assume that the firm decides how many workers to hire for the entire year knowing that the market wage, determined by agricultural productivity, will fluctuate seasonally. The firm cannot adjust $L_M$ but they can change the actual amount of labour used $L_M(1-a)$ by offering different wages to different workers. The workers that are offered a wage lower then the current agricultural one will be absent.

In some circumstances it may be optimal for the firm to have a positive absence rate. For example, if $\gamma = 1$ the optimal strategy for the firm is to offer workers a wage $w_M = \theta$ and hire $L^*_M(\theta)$ workers. During the peak agricultural season they will increase the wage to $\overline{\theta}$ only for $L^*_M(\overline{\theta})$ workers and let the other be absent. In this case, absenteeism allows the economy to attain the full flexibility outcome.

The optimal level of absence will still be positive even when $\gamma > 1$, as long as the loss in productivity caused by absenteeism is lower then its benefit in terms of savings in the wage bill. On the other hand, if $\gamma > \frac{1}{\alpha}$, the cost of absenteeism will always be higher than its benefit so the firm to increase the wage of all workers to $\overline{\theta}$ and avoid absenteeism.
2.3 Rigid wages and fixed labour

In most cases manufacturing firms must offer fixed wages contract to their employees and cannot adjust wages seasonally. Since workers can always choose to be absent and get the current agricultural wage, the firm has to options: they can offer a low wage contract, with \( w_M = \theta \) and let workers be absent during the peak agricultural season \( 1 - a = \rho \); or a high wage contract, with \( w_M = \bar{\theta} \) and zero absenteeism.

The condition under which the firm will choose to set offer the high wage contract is the following:

\[
\frac{\bar{\theta}}{\theta} \leq \frac{1}{\rho^{\gamma-1}} \tag{4}
\]

It is interesting to notice that condition 4 does not depend on \( \alpha \) so it is possible to observe firms offering low wages and experiencing high level of absenteeism even if \( \gamma > \frac{1}{\alpha} \). If \( \gamma \) is low, and in particular when it is equal 1 it optimal for the firm to offer the low wage contract and experience absenteeism during the peak agricultural season.

2.4 Model implications

The model illustrates how seasonal absenteeism can be interpreted as consequence of the coexistence of two sectors sharing the same labour force, as in the case of an economy undergoing structural transformation. In particular, it shows how asymmetric labour market rigidities (i.e. rigid labour regulation in manufacturing sector and flexible in agriculture) may generate higher absence rates during the peak agricultural season, even when these cause important productivity losses.

Interestingly, the theory suggests that, if wages are flexible and the productivity loss generated is low (\( \gamma \) close to 1), absenteeism may be used by the firm to overcome the inefficiency caused by rigid labour market regulation.

Some testable prediction can be derived from this model. First, absence rates should be higher when agricultural productivity and agricultural wages are high. This means every year during the harvest season but in particular when positive productivity shocks increase agricultural yield and wages. Second, absenteeism should be higher in firms paying lower wages. Third, absenteeism should be more prevalent in firms facing a more rigid labour regulation.
Finally, the model highlights the importance of obtaining a consistent estimate of the parameter $\gamma$. Indeed, only if $\gamma$ is high, and in particular if it is higher than $\frac{1}{\alpha}$ absenteeism should be considered an undesirable phenomenon.

3 Data and Summary Statistics

This paper is based on firm level data representative of the entire Indian manufacturing sector that are collected through the Annual Survey of Industries (ASI). These data are matched to data from the agricultural sector such as agricultural wages and crop information. The data appendix describes the each data source in detail and explains how variables are computed.

Table 1 reports summary statistics the agricultural sector and table 2 reports summary statistics for the firm level data.

3.1 Pro-worker States

The notion of “pro-worker labour regulation was initially introduced by Besley and Burgess (2004), who consider all state level of to sections V-A and V-B of Industrial Disputes act of 1947 and categorised them as “pro-worker”, “pro-employer” or “neutral”, based on their content. Based on this, they obtained a measure of strictness of labour regulation that varies across states and over time. However, these amendments took place before the beginning of the period considered in this paper, leaving only the cross-sectional variation. Following Adhvaryu et al. (2013), I classify as“pro-worker” the states that passed more amendments in “pro-worker” than in “pro-employer” direction. These are Maharashtra, Orissa and West Bengal7. Figure 1 shows the distribution of firm level yearly average absence rates, separating the sample into “pro-worker” and “pro-employer and neutral” states. In both sub-samples there is a long tail of firms with very high absence rates, however, the distribution in “pro-worker” states is shifted to right, with a large amount of firms reporting very high absence rates (between 10% and 20%).

7In their original classification, Adhvaryu et al. (2013), included also Gujarat among the “pro-worker” states, however I remove it from this category for two reasons: (i) the original coding was criticised by the literature (Bhattacharjen, 2006)(ii) during the period considered some reforms in “pro-employer” direction take place in the state.
3.2 Agricultural vs Manufacturing Wages

Figure 2 plots the distribution of the difference between the yearly average wage in the firm and the yearly average agricultural wage in the district. Manufacturing wages are in general higher than agricultural wages, but there is an overlap suggesting that agricultural wages may be attractive for some manufacturing workers.

4 Agricultural Labour Market

In most developing countries the agricultural sector still employs the vast majority of the labour force. India is no exception, according to Census data, in 2001 53.4% of the people employed worked in agriculture.

An important feature of the agricultural labour market is the prevalence of short term casual labour contracts (Kaur, 2013). Their availability is highly seasonal as it depends on the phases of the agricultural cycle. As illustrated by figure 3, the Indian crop calendar is characterised by two main growing seasons: rabi and kharif. Kharif crops are sown between June and July when the first monsoon rain arrive and harvested between October and December; while rabi crops are sown between October and November and harvest in March and April. Labour demand is highest during the months of harvest and much lower in the other times of the year. As a consequence there is large seasonal unemployment and seasonal migration from rural to urban area is a common way in which households integrate their income and smooth consumption (Morten (2013) and Bryan et al. (2013)).

In the next section I will test to what extent workers’ attendance in the manufacturing sector respond to changes in agricultural wages. To validate my analysis I verify that agricultural wages reflect changes in agricultural labour demand. Hence, I test whether they are higher during the harvest season, when most agricultural jobs are available, and when there are positive shocks to crop productivity, that increase labour requirements. I use weather shocks, in particular rainfall and temperature, to identify shocks to a crop productivity.

The causal effect of weather on crop productivity and agricultural wages has already been established by the literature. Using Indian data Jayachandran (2006) shows that rainfall has a positive effect on crop yield and agricultural wages. Following the same empirical strategy Kaur (2013) highlights the fact that nominal agricultural wages are sticky and their response
to shock is asymmetric: wages increase following a positive shock but do not decrease (in nominal terms) following a negative shock. Burgess et al. (2014) find that temperature is also an important determinant of agricultural productivity and rural income. In particular, abnormally high temperature during the growing season significantly decreases agricultural yield and rural wages.

However, these studies are based on yearly level data and define the growing season as the period following the arrival of the Southwest monsoon. While this is arguably the most important meteorological event for the Indian agriculture, it is not clear whether it should affect agricultural wages in all months of the year. In order to verify the effect of weather on monthly agricultural wages I replicate the analysis using monthly level data. For this purpose I construct a monthly measure of “weather shocks” based on the crops harvested in each month and their growing season.

In order to estimate to what extent agricultural wages follow the agricultural harvest calendar I exploit the variation in percentage of area harvested across districts. This is due to the fact that different districts specialise in crops that may differ in harvest calendar, but also to the fact that crops’ harvest calendars vary across districts. I estimate the following equation:

$$\log(agr\_wage_{tmd}) = \alpha_{harvest_{tmd}} + \delta_d + v_t + e_m + u_{tmd}$$ (5)

where $\log(agr\_wage_{tmd})$ is natural logarithm of agricultural wage in district $d$ in month $m$ of year $t$; $harvest_{tmd}$ represents the percentage of the total sown area of district $d$ that is harvested in month $m$ of year $t$; $\delta_d$, $v_t$ and $e_m$ represent district, year and month fixed effects, respectively.

To estimate the effect of crop yield on agricultural wages I use the following empirical model:

$$\log(agr\_wage_{tmd}) = \alpha_{log\_yield\_index_{tmd}} + e_{dm} + v_t + u_{tmd}$$ (6)

where $log\_yield\_index_{tmd}$ is the weighted average of the log of yield of the crops harvested in month $m$, year $t$ in district $d$; $e_{dm}$ and $v_t$ represent district-month and year fixed effects, respectively.

Following Jayachandran (2006) I use weather as instrument for crop yield. This solves the problem that agricultural wages and crop yield may move simultaneously for other reasons, for example following shocks to crop demand.

The first stage regression is:
\[ \log_{yield\_index_{tmd}} = \alpha_{\text{weather\_index}_{tmd}} + e_{dm} + \nu_t + w_{tmd} \]  

where \( \text{weather\_index}_{tmd} \) is a measure of weather relevant for the crops harvested in month \( m \) of year \( t \) in district \( d \). In particular, I create a crop specific measure of growing season cumulative rainfall and average temperature and then aggregate them at the district-month level as weighted average, with weights representing the relative importance of each crops in terms of agricultural area.

In all specification I cluster standard errors at the district level to account for the presence of autocorrelation. Moreover, to make sure that the effect of weather on crop yield is relevant for the crops harvested in all months of the year I estimate equation 7 for all months separately, the resulting coefficients and confidence intervals are reported in figure 4.

Table 1 reports summary statistics of the variables\(^8\) used in this part of the analysis. Comparing the number of observations in Panel A and B it is possible to notice that a large number of observation for agricultural wages is missing. Moreover, crop yield and weather variables are available only for the months of harvest, considering only these months considerably reduces the sample size.

Table 3 reports the estimates for equation 5. Agricultural wages are higher during the harvest season and the effect is positive and statistically significant also when month fixed effects are included, thus exploiting only the differences in harvest calendar across districts. The percentage of agricultural area sown in the month has no significant effect on wages. The magnitude of the coefficient suggests that going from 0 of agricultural land harvested to 100% increases agricultural wages by 3.5%. While these results provide suggestive evidence of seasonality in the agricultural labour market, they are likely to severely understate the importance of the phenomenon. Indeed, the changes in labour demand are likely to affect the number of jobs created much more than they affect wages because the latter tend to be sticky, as pointed out by Kaur (2013).

Table 4 reports the first stage results. The first 3 columns are estimated on the whole sample, while the last 3 only on the sub-sample of observations for which agricultural wages are non-missing. The effect of rainfall on crop yield is positive and the effect of temperature is negative, as expected. The estimated coefficients are statistically different from zero in both samples and their magnitudes are comparable.

\(^8\)See appendix A.2 for a complete description of the data.
Table 5 reports the results of the estimate of the effect of crop yield on agricultural wages. The IV estimate, reported in column (2), shows that 1% increase in crop yield results in a 0.25% increase in agricultural wages.

5 Workers’ Attendance

5.1 Monthly Level Analysis

The model in section 2 predicts workers attendance to be lower when labour demand and wages in the agricultural sector are higher. To test this hypothesis I exploit the fact that Indian districts specialise in different crops and that crop calendar varies across districts. I estimate the following empirical model:

\[
\log(\text{attendance}_{idmt}) = \alpha \text{harvest}_{dm} + \beta x_{idmt} + \delta_d + e_m + v_t + u_{idmt} \tag{8}
\]

where \(\log(\text{attendance}_{idt})\) is the natural logarithm of the average attendance rate in firm \(i\), located in district \(d\), in month \(m\) and year \(t\); \(\text{harvest}_{dm}\) is a dummy equal to 1 if the main crop cultivated in district \(d\) is harvested in month \(m\); \(x_{idmt}\) is a vector of controls that includes firm and district characteristics; \(\delta_d\), \(e_m\) and \(v_t\) represent district, month and year fixed effects, respectively.

I also test whether attendance in the manufacturing sector responds to changes in the local agricultural wage by estimating the following equation.

\[
\log(\text{attendance}_{idmt}) = a \log(\text{agr}_{}\text{wage}_{dmt}) + \beta x_{idmt} + \delta_d + e_m + v_t + u_{idmt} \tag{9}
\]

where \(\log(\text{agr}_{}\text{wage}_{dt})\) is the natural logarithm of the agricultural wage in district \(d\), in month \(m\) and year \(t\), the other variables are defined as above.

The results are reported in table 3. Column (1) shows that attendance rates are significantly lower during the harvest month; while column (2) shows that the effect is significantly higher in “pro-worker” states. These results suggest there is relationship between the seasonal patter in workers’ attendance and the agricultural cycle. However, this may be driven by other events that happen in the same month, such as harvest festivals or weddings. In order to test whether absenteeism is caused by the availability of economic opportunities, I look at the response of workers’ attendance to changes in the local agricultural wage.
The estimated response of attendance rates to changes in agricultural wages is reported in column (3) and (4): the effect is negative and statistically significant, suggesting that workers are absent more often when they have a better outside option in the agricultural sector. Moreover, the effect is higher in pro-worker states, although the difference is not statistically significant. The results are robust to the introduction of district-month fixed effects, suggesting that the component of agricultural wages that captures changes in agricultural labour demand from one year to another is also important.

However, agricultural wages are likely to be endogenous so these results cannot be interpreted as causal. Moreover, the large amount of missing observations in the agricultural wages causes an important reduction in the sample size by more than a half. This substantially reduced the variation available in some districts/states and makes it impossible to detect heterogeneous effects across location (for example for pro-worker states). Finally, measurement error in agricultural wages is likely to bias the coefficient towards zero. As a solution to these problems I propose a two-sample two-stage estimation approach. In the first stage I estimate the response of agricultural wages to exogenous shocks in agricultural productivity for the sub-sample in which agricultural wages are available. I then predict agricultural wages for the whole sample, based on the estimated parameters and the available shock data. Finally I use estimated agricultural wages to estimate the response of attendance. Since the instrument for agricultural wages are available only for the months of harvest, I perform this analysis using yearly level data.

5.2 Yearly Level Analysis

The analysis reported in appendix 4 shows that weather shocks that increase crop yield have a positive effect on agricultural wages. In this section I use weather shocks as instrument to estimate the causal effect of agricultural wages on workers' attendance. The identifying assumption is that growing season weather affects attendance only through agricultural wages. The major concern is that rainfall and temperature may have a direct effect on attendance, for example high rainfall may decrease attendance if it make it impossible to reach workplace because roads are flooded. To address this concern I control for yearly rainfall and temperature and exploit only

\footnote{Agricultural wages are disproportionally missing in the state of Maharashtra, the largest of the 3 states classified of pro-worker, it is possible that the variation left is not enough to detect an effect. Indeed, when replicating the estimate in column (2) of table 3 on the sub-sample for which agricultural wages are available the coefficient of the interaction Harvest*pro-worker is also not statistically significant.}
the variation in growing season weather. Another concern is that a good outcome in agriculture can affect demand for the goods produced by the firm and therefore the attendance indirectly. To the extent to which goods are traded all over the country, year fixed effect will control for this.

The first stage regression is the following:

\[
\log(agr\_wage_{dt}) = \gamma\text{weather\_index}_{td} + \delta_d + v_t + e_{dt}
\] (10)

where \(\text{weather\_index}_{td}\) is a measure of growing season rainfall and temperature that affects crops harvested in district \(d\) in year \(t\).

The first stage is estimated on the sub-sample for which agricultural wages are available. Then, based on the estimated coefficients from the first stage, I predict \(\log(agr\_wage_{dt})\) for the whole sample and I then use these estimates to study the second stage. The second stage therefore, becomes:

\[
\log(attendance_{idt}) = \alpha \log(agr\_wage_{dt}) + \beta x_{idt} + \delta_d + v_t + u_{idt}
\] (11)

where \(\log(agr\_wage_{dt})\) is the predicted log of agricultural wage obtained from the first stage regression. Since \(\log(agr\_wage_{dt})\) is estimated, the second stage standard errors need to be adjusted following Murphy and Topel (1985).

The results are reported in table 7. Column (1) reports the results from the first stage regression, the results is similar to that obtained using monthly level data. Column (2) reports the second stage estimate: the coefficient suggests that a 1% increase in the agricultural wage reduces yearly attendance by 0.142%. Column (3) shows that the effect is stronger in pro-worker states, where a 1% increase in agricultural wages causes a reduction in attendance of 0.212%.

6 Absenteeism and Productivity

The objective of this section is to estimate the effect of workers’ absenteeism on firms productivity. While absenteeism is usually considered to be an undesirable outcome, often associated with shirking, its effect on firms productivity is not necessarily negative. Indeed, the firm may be able to compensate for the loss in working time caused by absence by hiring substitute workers of having the present ones working overtime. It is also plausible that when some workers are absent, the productivity of those present may increase if they work harder to compensate. If
the firm is unable to find equally productive substitutes, however, absenteeism may cause a loss in productivity. Moreover, it may create disruptions in the productive activity, reduce workers’ morale and take up management time that could otherwise be devoted to more productive activity. In some cases these issues may result in lower product quality and delay in deliveries, which may harm the firm’s reputation and, ultimately, its growth.

Let us assume the firm as a Cobb-Douglas production function represented by equation 12:

\[ Y = A(L(1 - a)\gamma)\alpha K^\beta \]  

(12)

where \( a = \) absence rate, \( L = \) number of worker-days scheduled and \( K = \) capital. The actual number of worker-days worked will be \( L(1 - a) \): the number of worker-days scheduled multiplied by the attendance rate. The coefficient \( \gamma \) can be used to test whether absenteeism has an effect of both output and productivity. In particular, if \( \gamma = 0 \) absenteeism has no effect on output, suggesting that the firm can fully adjust; whereas, if \( \gamma = 1 \) there is no adjustment and the effect of absenteeism is equivalent to the effect of a change in the number of worker-days scheduled; if \( \gamma < 0 \) productivity increases with absence rate, as in the case in which the present workers exert extra effort to compensate for the absent ones; finally, if \( \gamma > 1 \) absenteeism has a negative effect on productivity.

The corresponding empirical model is illustrated by equation 13:

\[ \log(Y_{idt}) = \alpha \log(L_{idt}) + \alpha \gamma \log(attendance_{idt}) + \beta \log(K_{idt}) + \epsilon_{idt} \]  

(13)

where: \( Y_{idt} \) represents the output of firm \( i \) in year \( t \) and district \( d \); \( L_{idt} \) is the number of worker-days scheduled by firm \( i \) in year \( t \) and district \( d \); \( attendance_{idt} \) is attendance rate in firm \( i \) in year \( t \) and district \( d \).

To address the problem that \( L_{idt} \) and \( K_{idt} \) may be determined after the shock to workers’ attendance takes place, and therefore be endogenous, I use number of workers and capital at the beginning of the year. Similarly, the quantity of raw materials and other inputs is likely to depend on workers’ attendance therefore, to avoid including it in the regression, I measure \( Y_{idt} \) in terms of value-added.

The major concern for the estimation of equation 13 is that absence rate is correlated with other unobservable determinants of firm productivity. For example, in years in which the firm is facing low demand for its output it may be more willing to tolerate absenteeism or it may even encourage it. Moreover, workers employed in less productive firms are often paid lower
wages and they are more likely to be absent as it easier for them to find a better paid temporary employment elsewhere. These factors would lead to an overestimate of the effect of absenteeism on productivity. On the other hand, yearly absence rates $a_{idt}$ is likely to be measure with error as it is obtained by aggregating four monthly observations provided in the dataset. This measurement error would probably bias the estimate of $\alpha \gamma$ towards zero.

To solve this endogeneity problem and obtain causal estimates of the effect of absenteeism on productivity I use weather shocks as instruments for workers attendance. By picking up the local average treatment effect (LATE) effect of absenteeism caused by weather shocks, this strategy will allow me to estimate the effect the absenteeism that is caused the competition between manufacturing and agricultural sectors. This effect is likely to be different from the effect of other sources of absence such as sick leave for various reasons: the workers engaged in this are those who have a better outside option in agriculture and, therefore, are likely to be stronger and healthier; they should also be rather confident that, once they come back they are not going to lose their jobs so they are more likely to be permanent workers with longer tenure; moreover they are likely to leave contemporaneously generating a larger disruption for the firm. On the other hand, this type of absenteeism is easier to anticipate for the firm and they may be able to adjust accordingly.

The first stage regression in this case is the reduced form estimate for the analysis proposed in section 5.2:

$$\log(attendance_{idt}) = \gamma_{\text{weather\_index}_{idt}} + \beta x_{idt} + \delta_d + v_t + e_{dt}$$

(14)

The results are reported in table 8. Column (1) reports the OLS estimate of equation 13: the attendance coefficient suggest that 1% increase in attendance increases value added by 0.887%. The IV estimate, reported in column (3) is much larger: it indicates that an increase in attendance by 1% increases value added by 6.212%. The parameter of interest, $\gamma$, is calculated as the ratio of the attendance parameter $\alpha^\ast \gamma$ and the labour parameter $\alpha$. The IV estimate is 22.98, suggesting that absenteeism not only has a negative effect firms’ output, but it also reduces the productivity of the time effectively worked ($\gamma > 1$).
7 Conclusions

This paper shows how firms in countries undergoing structural transformation face competition over their labour force during the peak agricultural seasons. The large supply of workers during the lean agricultural seasons allows manufacturing firms to pay relatively low wages but, during the peak season, workers may find casual jobs in agriculture attractive and leave the firm temporarily.

The theory suggests that an labour marker rigidities in the manufacturing sector not only make it easier for workers to be absent when they have a better outside option, but also make it harder for the firm to prevent absenteeism even when it has large negative effect on productivity.

Using firm level representative of the entire Indian manufacturing sector, I find that workers’ attendance is lower during the months of harvest and more so in states that have a more “pro-worker” labour protection regulation. Exploiting exogenous shocks to labour demand in agriculture I estimate the elasticity of workers’ attendance with respect to the local agricultural wage and find that is negative and statistically significant.

Finally, I estimate the effects of absenteeism on firms’ output and productivity finding large effects.

References


### A Data Appendix

#### A.1 Firm Level Data

The firm level data are taken from the Annual Survey of Industries (ASI)\textsuperscript{10}, which is the principal source of information on the Indian registered manufacturing sector. They are representative of all registered “factories”, defined as plants using power employing 10 or more workers and plants not using power employing 20 or more workers, operating in the manufacturing sector. Firms employing more than 100 workers are interviewed every year while the smaller ones are randomly sampled.

The survey is composed of 2 parts: Part I, collected on yearly basis, includes data on firms’ assets and liabilities, employment and labour cost, inputs and output; while Part II\textsuperscript{11} provides monthly information on absenteeism and labour turnover. This paper uses data from 8 rounds of the 2 parts of the survey, covering the period 2000-01 to 2007-08\textsuperscript{12}.

All variables expressed in nominal terms were deflated using the all India monthly CPI for Industrial workers\textsuperscript{13}.

\textsuperscript{10}See MOSPI (2014) for a complete description of these data.

\textsuperscript{11}The two parts of the survey are administered to the same firms and can be matched using identifiers. However, the reference period of Part I is the fiscal year: from April to March; while Part II is based on the calendar year: from January to December.

\textsuperscript{12}Part I data were downloaded from the LSE India Data Centre website http://idc.lse.ac.uk/ and Part II data were obtained from the Indian Labour Bureau.

\textsuperscript{13}The CPI data was obtained from the Indian Labour Bureau website http://labourbureau.nic.in/indtab.html.
Absence Data

In the survey absence is defined as “failure of a worker to report for work when he is scheduled to work” that is “when the employer has work available for him and the worker is aware of it.” This includes: absence with or without pay, with or without permission. It does not include absence due to strikes and lock outs, lay off, weekly rest and suspension (MOSPI, 2014).

The dataset contains monthly data on number of man-days worked and number of man-days lost due to absence. However, firms operating on “perennial” basis, accounting for 88% of the sample, are only required to report this information for the months of March, June, September and December.

Moreover, absence data are only collected for regular workers employed directly employed by the firm. These are permanent, probationer and temporary workers. This classification excludes casual, badli or substitute workers, workers employed through contractors and apprentices.

Other Variables

Relevant variables have been constructed following the tabulation program provided in MOSPI (2014):

- $\text{total output} =$ value all products and semi-finished products manufactured during the year, plus value of fixed assets produced by the factory for its own use, plus receipts from services sold, plus value of goods sold in the same conditions as purchased.

- $\text{total input} =$ total value of material and fuel consumed, plus cost of services purchased (repair, insurance etc.), plus operating and non-operating expenses, plus purchase value of goods sold in the same conditions as purchased.

- $\text{gross value added} =$ total output - total input

- $\text{profits} =$ gross value added minus depreciation of fixed assets during the year, minus rent and interest paid, minus total labour cost
Sample Selection

The original Part I dataset includes 354,689 firm-year observations\textsuperscript{14}. One quarter of initial the observations are dropped because the firm was closed or did not respond to the survey. Merging Part I and Part II results in the loss of 36,949 observations, almost 14\% of the sample. The lower response rate to Part II of the survey can be explained by the fact that, while Part I includes only data that the firm is legally required to keep to produce a balance sheet, Part II requires some extra effort in data collection. The firms that do not respond to Part II are on average smaller, younger and pay lower wages.

Another 35,592 observations, about 9\% of the original sample, are dropped because some important variables have missing, zero or non-plausible values\textsuperscript{15}. Moreover, 2,623 firms report negative or zero output and are excluded from the analysis\textsuperscript{16}

Finally, firms belonging to the food-processing and tobacco sectors are excluded from the sample because they are directly affected by what happens in the agricultural sector through their inputs, which would invalidate the instrumental variable approach used in this paper. The final sample includes 107,539 firm-year observations.

A.2 Agricultural Data

Agricultural Wages

Monthly agricultural wages at the district level are collected by the Directorate of Economics and Statistics of the Indian Ministry of Agriculture in a yearly publication called “Agricultural Wages in India” (AWI)\textsuperscript{17}. The wage information was originally collected at the centre level and, in 10\% of the cases, multiple observations per district are available. In such cases I simply

\textsuperscript{14}I exclude firms not belonging to the manufacturing sector, about 5\% of the sample, and firms located in Union Territories or smaller states for which agricultural data are not available (Goa, Jammu and Kashmir, Meghalaya, Manipur, Nagaland and Tripura), about 9\% of the initial sample.

\textsuperscript{15}The variables considered are: total output; total inputs; firm size; number of regular workers; number of man-days worked; total labour cost; rural; ownership type (public, private, etc.); organisation type (private limited, partnership, etc.); firm age; number of months operational; number of manufacturing days.

\textsuperscript{16}Negative observed output is possible because it includes the change in stock of semi-finished products, which may be negative.

\textsuperscript{17}The data I am using were kindly shared with me by Thiemo Fetzer who digitised and prepared them for the paper Fetzer (2013).
take the average of the observed wages. The data contain information about wages for various occupations and are collected separately for men and women. I consider only male wages related to field labour.

The Employment and Unemployment part of National Sample Survey (NSS) provides another source of information on agricultural wages. It is a household survey in which a random sample is drawn every quarter and individuals are asked about their activity during the week preceding the survey. As interviews are spread over the quarter, I use these data to construct a district level measure of monthly agricultural and manufacturing wages for male workers. 4 rounds of the survey are available in the period considered providing information on wages from January 2004 to June 2006 and from July 2007 to June 2008 \textsuperscript{18}.

**Crop Calendar**

The information about the agricultural production cycle comes from the 1967 Indian Crop Calendar published by the Directorate of Economics and Statistics. It contains information about typical sowing and harvesting months of major crops at the district level. To match 1967 districts with 2001 districts I refer to Kumar and Somanathan (2009), who provides a mapping of Indian districts over time.

In the original dataset the crop calendar was missing for some crops in some districts. Some of these crops appear to be important in terms of area harvested and should be taken into account. I fill the gaps by inputing crop calendar for the missing observations by replacing it with the closest available observation in terms of distance between district centroids. The final dataset contains the months of sowing and harvesting of 23 crops: bajra, castor seed, chillies, coriander, cotton, ginger, gram, groundnut, jowar, jute, maize (kharif), mesta, niger seed, onion, potato, ragi, rapeseed and mustard, rice, small millets, sugarcane, tobacco, turmeric and wheat.

**Crop Yield and Area Sown**

I obtained district level data on crop output and area sown for major crops from the Crop Production Statistics Information System website\textsuperscript{19}. Moreover, I restrict the sample to crops that

\textsuperscript{18}These are NSS rounds 60, 61, 62 and 64. I downloaded the data from LSE India Data Centre website http://idc.lse.ac.uk/.

\textsuperscript{19}apy.dacnet.nic.in/
cover at least 1% of the area sown in the district and to the district for which I have information about crop covering at least 50% of the total area sown. I keep only crops for which sowing and harvest dates are provided in the crop calendar described above.\(^{20}\)

Matching these data with the crop calendar I construct monthly measures of crop yield and percentage of area harvested. Crop yield is defined as crop output divided by area under crop. For comparability across crops I normalize the yield of each crops to mean one. I then aggregate the data of the various crops to obtain a district level measure of monthly crop yield, \textit{log crop yield index}, constructed as weighed average of the log of yield of the crops harvested in the month. The weights are equal to the share of the area under each crop. This is similar to the index used by Jayachandran (2006), with the difference that here weights are based on the share of area under each crop instead of the share of revenues originating from the crop.

**Rainfall and Temperature**

I use rainfall and temperature data collected by the Center of Climatic Research at the University of Delaware\(^{21}\). The datasets contain time series of average monthly rainfall and temperature from 1900 to 2010 on a 0.5 by 0.5 degree grid. Using a shapefile of India, reporting district borders as they were at the time of the 2001 Census, I compute the coordinates of each district’s centroid and I match them to the closest point on the grid to obtain monthly average temperature and rainfall for each district.

Following Donaldson (2014) I construct a crop specific measure of weather shock defined as cumulative rainfall or average temperature during the crop’s growing season.

\(^{20}\)The excluded crops are: cashewnut, guar seed, khesari, maize (rabi), moth, oilseeds and soyabean. 1,400 observation, average area under crop is 13%.

\(^{21}\)http://climate.geog.udel.edu/ climate/
Figures

Figure 1: Absence Rates’ Distribution

Notes: The figure plots the distribution of firms’ yearly absence rates. The solid line represents the distribution of absence rates for firms operating in pro-employer and neutral states, while the dashed line represents the distribution of yearly absence rates for firms operating in pro-worker states. According to the definition used in this paper, the states labeled as pro-worker are Maharashtra, Orissa and West Bengal; all other states are included in the pro-employer & neutral category.
Figure 2: Agricultural vs Manufacturing Wages

Notes: The figure plots the distribution of the difference between firms’ average yearly wage for regular workers and the districts’ average yearly agricultural wage. The firms to the left of the vertical line pay wages on average below the local agricultural wage.
Notes: the percentage of total area harvested and sown in a given month is calculated by matching yearly data on area under major crops with the crop specific crop calendar at the district level. The results obtained are then averaged over the 2000-2007 period. When a crop is harvested/sown in more than a month, its area is divided by the number of months of harvest/sowing and inputed equally to each of them.
Figure 4: Agricultural Wages

Notes: the figures plot the estimated effect of rainfall and temperature on crop yield and their 95% confidence intervals. They are obtained regressing rainfall and temperature indices on log of crop yield index one month at a time. Log crop yield, rainfall and temperature indices are computed as the weighted average of the respective crop-level measures, with weights equal to the relative importance of each crop, in terms of area sown among those harvested in the same month. In particular, log crop yield is the natural logarithm of the crop output divided by its area planted, normalised to mean 1; rainfall and temperature are measured as cumulative rainfall and average temperature over the crop’s growing season, which is defined as the period between the first month of sowing and the last month of harvest.
# Tables

## Table 1: Summary Statistics - Agricultural Sector

<table>
<thead>
<tr>
<th>Variable</th>
<th>Panel A - Whole Sample</th>
<th>Panel B - Non-missing wages &amp; yield index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Mean</strong></td>
<td><strong>Standard Deviation</strong></td>
</tr>
<tr>
<td>Log wage</td>
<td>3.920</td>
<td>0.342</td>
</tr>
<tr>
<td>Share area harvested</td>
<td>0.093</td>
<td>0.168</td>
</tr>
<tr>
<td>Share area sown</td>
<td>0.096</td>
<td>0.162</td>
</tr>
<tr>
<td>Log crop yield index</td>
<td>-0.130</td>
<td>0.541</td>
</tr>
<tr>
<td>Rainfall index (100 mm)</td>
<td>7.103</td>
<td>6.668</td>
</tr>
<tr>
<td>Temperature index (°C)</td>
<td>24.48</td>
<td>4.676</td>
</tr>
</tbody>
</table>

Notes: CPSIS stands for Crop Production Statistics Information System and ICC stands for Indian Crop Calendar. AWI stands for Agricultural Wages of India, published by the Directorate of Economics and Statistics of the Indian Ministry of Agriculture. Log yield, rainfall and temperature indices are computed as the weighted average of the respective crop-level measures, with weights equal to the relative importance of each crop, in terms of area sown, among those harvested in the same month. In particular, log yield is the natural logarithm of the crop output divided by its area planted, normalised to mean 1; rainfall and temperature are measured as cumulative rainfall and average temperature over the crop’s growing season, which is defined as the period between the first month of sowing and the last month of harvest.
## Table 2: Summary Statistics - Firm Level Data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Observations</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A - Monthly level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absence rate</td>
<td>0.097</td>
<td>0.079</td>
<td>404,222</td>
<td>ASI II</td>
</tr>
<tr>
<td>Log attendance</td>
<td>-0.107</td>
<td>0.102</td>
<td>404,222</td>
<td>ASI II</td>
</tr>
<tr>
<td>Harvest</td>
<td>0.096</td>
<td>0.362</td>
<td>404,222</td>
<td>CPSIS and ICC</td>
</tr>
<tr>
<td>Log agr wage</td>
<td>4.067</td>
<td>0.331</td>
<td>139,815</td>
<td>AWI</td>
</tr>
<tr>
<td><strong>Panel B - Yearly level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log attendance</td>
<td>-0.101</td>
<td>0.066</td>
<td>107,539</td>
<td>ASI II</td>
</tr>
<tr>
<td>Log number of workers</td>
<td>3.302</td>
<td>1.482</td>
<td>107,539</td>
<td>ASI II</td>
</tr>
<tr>
<td>Rainfall Index (100 mm)</td>
<td>6.885</td>
<td>8.907</td>
<td>107,539</td>
<td>Univ of Delaware,CPSIS and ICC</td>
</tr>
<tr>
<td>Temperature Index (°C)</td>
<td>22.66</td>
<td>3.591</td>
<td>107,539</td>
<td>Univ of Delaware,CPSIS and ICC</td>
</tr>
<tr>
<td>Log agr wage</td>
<td>4.051</td>
<td>0.328</td>
<td>62,601</td>
<td>AWI</td>
</tr>
<tr>
<td>Log capital</td>
<td>17.48</td>
<td>2.069</td>
<td>56,786</td>
<td>ASI I</td>
</tr>
<tr>
<td>Log value addedd</td>
<td>16.59</td>
<td>0.027</td>
<td>56,786</td>
<td>ASI I</td>
</tr>
</tbody>
</table>

Notes: ASI I and ASI II stand for Annual Survey of Industries-Part I and II, respectively; CPSIS stands for Crop Production Statistics Information System and ICC stands for Indian Crop Calendar; AWI stands for Agricultural Wages of India, published by the Directorate of Economics and Statistics of the Indian Ministry of Agriculture; Absence rate is computed as number of man-days lost due to absence divided by number of man-days scheduled; log attendance is the natural logarithm of 1-absence rate; harvest is dummy equal to 1 if the main crop cultivated in the district is harvested in the month; Rainfall and temperature indices are computed as the weighted average of the respective crop-level measures, with weights equal to the relative importance of each crop, in terms of area sown, among those cultivated in the district. In particular, rainfall and temperature are measured as cumulative rainfall and average temperature over the crop’s growing season, which is defined as the period between the first month of sowing and the last month of harvest. The summary statistics for capital and value added are reported only for the sample in which both variables are non missing (excluding all observations with negative value added, zero capital or missing values in one of the two variables).
### Table 3: Agricultural Wage and Harvest Season

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>log(agr wage)</td>
<td>log(agr wage)</td>
<td>log(agr wage)</td>
<td>log(agr wage)</td>
</tr>
<tr>
<td>Share area harvested</td>
<td>0.035***</td>
<td>0.035***</td>
<td>0.027***</td>
<td>0.027***</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.008)</td>
<td>(0.009)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>Share area sown</td>
<td>-0.009</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>13456</td>
<td>13456</td>
<td>13456</td>
<td>13456</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.846</td>
<td>0.846</td>
<td>0.847</td>
<td>0.847</td>
</tr>
</tbody>
</table>

Notes: The dependent variable is the natural logarithm of the monthly agricultural wage. Share of area harvested and Share of area sown represent the percentage of total agricultural area in the district that is harvested/sown in the month. All specifications are based on unbalance panel of district level monthly data and they include district fixed effects. Columns (1) and (2) include year FE. Columns (3) and (4) include year and month fixed effects. Standard errors are clustered at the district level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

### Table 4: Weather and Crop Yield (First Stage)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>log YI</td>
<td>log YI</td>
<td>log YI</td>
<td>log YI</td>
<td>log YI</td>
<td>log YI</td>
</tr>
<tr>
<td>Rainfall Index</td>
<td>0.027***</td>
<td>0.027***</td>
<td>0.017***</td>
<td>0.018***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temperature Index</td>
<td>-0.082***</td>
<td>-0.081***</td>
<td>-0.048**</td>
<td>-0.056**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.014)</td>
<td>(0.024)</td>
<td>(0.024)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>21540</td>
<td>21540</td>
<td>21540</td>
<td>7213</td>
<td>7213</td>
<td>7213</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.644</td>
<td>0.640</td>
<td>0.647</td>
<td>0.678</td>
<td>0.675</td>
<td>0.679</td>
</tr>
</tbody>
</table>

Notes: The dependent variable, log YI, is log yield index. Log yield, rainfall and temperature indices are computed as the weighted average of the respective crop-level measures, with weights equal to the relative importance of each crop, in terms of area sown, among those harvested in the same month. In particular, log yield is the natural logarithm of the crop output divided by its area planted, normalised to mean 1; rainfall and temperature are measured as cumulative rainfall and average temperature over the crop’s growing season, which is defined as the period between the first month of sowing and the last month of harvest. All specifications are based on unbalance panel of district level monthly data and they include district-month fixed effects. Standard errors are clustered at the district level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Table 5: Crop Yield and Agricultural Wages (2SLS)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>log YI</td>
<td>log(agr wage)</td>
<td>log(agr wage)</td>
<td>log(agr wage)</td>
</tr>
<tr>
<td>Rainfall Index</td>
<td>0.018***</td>
<td>0.003**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temperature Index</td>
<td>-0.056**</td>
<td>-0.004</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.008)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>log YI</td>
<td>0.153**</td>
<td>-0.012</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.071)</td>
<td>(0.010)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>7213</td>
<td>7213</td>
<td>7213</td>
<td>7213</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.679</td>
<td>0.806</td>
<td>0.830</td>
<td>0.829</td>
</tr>
</tbody>
</table>

Notes: The dependent variable in column (1), log YI, is log yield index; while the dependent variable columns (2) to (4) is the natural logarithm of the monthly agricultural wages. Column (1) and column (2) report, respectively, the first and second stage of the 2SLS estimate of the effect of crop yield on agricultural wages; column (3) reports the reduced form; column (4) reports the OLS regression. Log yield, rainfall and temperature indices are computed as the weighted average of the respective crop-level measures, with weights equal to the relative importance of each crop, in terms of area sown, among those harvested in the same month. In particular, log yield is the natural logarithm of the crop output divided by its area planted, normalised to mean 1; rainfall and temperature are measured as cumulative rainfall and average temperature over the crop’s growing season, which is defined as the period between the first month of sowing and the last month of harvest. All specifications are based on unbalance panel of district level monthly data and they include district-month fixed effects. Standard errors are clustered at the district level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Table 6: Workers’ Attendance - Monthly Level

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(att)</td>
<td>-0.002***</td>
<td>-0.001**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harvest</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harvest*pro worker</td>
<td>-0.002*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(agr wage)</td>
<td>-0.010**</td>
<td>-0.009*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(agr wage)*pro worker</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>404222</td>
<td>404222</td>
<td>139815</td>
<td>139815</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.127</td>
<td>0.127</td>
<td>0.152</td>
<td>0.152</td>
</tr>
</tbody>
</table>

Notes: The dependent variable is the natural logarithm of monthly attendance rates; Harvest is dummy equal to 1 if the districts’ main crop, in terms of cultivated area, is harvested in the month; pro worker is dummy equal to 1 for the states of Maharashtra, West Bengal and Orissa; log(agr wage) is the natural logarithm of the districts’ monthly agricultural wage. District level controls include monthly rainfall and temperature. Firm level controls include: firm size, rural, ownership type, organisation type, two digit industry. All specifications are based on unbalance panel of firm level monthly data (for the months of: March, June, September and December) and they include district, month and year fixed effects. Standard errors are clustered at the district level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. 
Table 7: Workers’ Attendance - Yearly Level

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Log(agr wage)</td>
<td>Log(attendance)</td>
<td>Log(attendance)</td>
</tr>
<tr>
<td>Rainfall Index</td>
<td>0.009**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temperature Index</td>
<td>-0.002***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\text{Log}}(\text{agrwage})$</td>
<td>$-0.142^{**}$</td>
<td>$-0.111$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.067)</td>
<td>(0.073)</td>
<td></td>
</tr>
<tr>
<td>$\hat{\text{Log}}(\text{agrwage})\times \text{ pro worker}$</td>
<td>$-0.101^*$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.060)</td>
</tr>
<tr>
<td>Observations</td>
<td>62601</td>
<td>107539</td>
<td>107539</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.84</td>
<td>0.165</td>
<td>0.166</td>
</tr>
</tbody>
</table>

Notes: Column (1) reports the first stage regression, the dependent variable is the natural logarithm of the districts’ yearly agricultural wage. Rainfall and temperature indices are computed as the weighted average growing season cumulative rainfall and average temperature of the crops cultivated in the district, with weights equal to the relative importance of each crop, in terms of area sown. Column (2) and (3) report the second stage, the dependent variable is the natural logarithm of the yearly attendance rate; $\hat{\text{Log}}(\text{agrwage})$ is predicted from the first stage regression; pro worker is dummy equal to 1 for the states of Maharashtra, West Bengal and Orissa. District level controls include yearly rainfall and temperature, they are included in all specifications. Firm level controls include: firm size, rural, ownership type, organisation type, two digit industry, they are included only in the second stage regressions. All specifications are based on yearly firm level data and include district and year fixed effects. Standard errors are clustered at the district level in all specifications, in columns (2) and (3) they are also adjusted to take into account that log(agr wage) is estimated, following Murphy and Topel (1985). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Table 8: Workers’ Attendance and Productivity

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log (value add)</td>
<td>0.887***</td>
<td>6.212***</td>
<td></td>
</tr>
<tr>
<td>$\alpha + \gamma$</td>
<td>(0.101)</td>
<td>(2.011)</td>
<td></td>
</tr>
<tr>
<td>Log L</td>
<td>0.403***</td>
<td>-0.004***</td>
<td>0.422***</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>(0.015)</td>
<td>(0.001)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Log K</td>
<td>0.580***</td>
<td>0.002***</td>
<td>0.568***</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.000)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>Rainfall Index</td>
<td>-0.001**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temperature Index</td>
<td>0.003***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma$</td>
<td>2.201***</td>
<td>22.98***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.264)</td>
<td>(8.242)</td>
<td></td>
</tr>
<tr>
<td>P-value $\gamma = 1$</td>
<td>0.000</td>
<td>0.008</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>56786</td>
<td>56786</td>
<td>56786</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.743</td>
<td>0.743</td>
<td>0.743</td>
</tr>
</tbody>
</table>

Notes: Column (1) reports the OLS estimate of the effect of attendance on output and productivity; Column (2) reports the first stage regression and Column (3) reports the second stage. The dependent variable in column (1) is the natural logarithm of the firm’s value added; while the dependent variable in column (2) is the natural logarithm of the firm’s yearly attendance rate. Log L is the natural logarithm of the number of workers employed in the firm at the beginning of the year; Log K is the natural logarithm of the firm’s gross fixed capital at the beginning of the year; Rainfall and temperature indices are computed as the weighted average growing season cumulative rainfall and average temperature of the crops cultivated in the district, with weights equal to the relative importance of each crop, in terms of area sown. District level controls include yearly rainfall and temperature, they are included in all specifications. Firm level controls include: firm size, rural, ownership type, organisation type, two digit industry, they are included only in the second stage regressions. All specifications are based on yearly firm level data and include district and year fixed effects. Standard errors are clustered at the district level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$