The London School of Economics and Political Science

Workers' absences and productivity in the Indian registered manufacturing sector

Giulia Zane

A thesis submitted to the Department of Economics of the London School of Economics for the degree of Doctor of Philosophy, London, November 2016

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I declare that my thesis consists of about 27,000 words.

Statement of inclusion of previous work

I can confirm that Chapter 5 is revised version of the paper is submitted for my MRes degree in Economics at the LSE, awarded in July 2012.

Acknowledgements

First, I am deeply indebted to my supervisor Oriana Bandiera for her continuous guidance and encouragement during these years. Besides teaching me a great lot about rigorous research, she helped me through every single step of my PhD and pushed me to believe more in myself and become a stronger person.

I also greatly benefited from advice and feedback from all members of the LSE Development and Labour groups. In particular, I wish to thank Robin Burgess, Greg Fischer, Alan Manning, Maitreesh Ghatak, Tim Besley and Steve Pischke for taking the time to provide valuable suggestions that definitely improved my research work.

I acknowledge EIEF and the ESRC for their financial support and I thank the Indian Labour Bureau for sharing the data and providing assistance whenever needed.

The biggest thank you, however, goes to my friends and colleagues from the cool office: Dana, Roberto, Eddy, Shiyu, Michel, Pedro and Yunichi, who helped me in all possible ways during the last 12 months. Mentioning each person's contribution in a fair way would make a list almost as long as this thesis so forgive me for just saying I'll never forget all you did for me.

Other LSE PhD students, from other offices, also hugely contributed to both my research and my personal wellbeing in past years. Among many, I want to mention Marta, Munir, Albert, Miguel and Yu-Hsiang who shared with me some very difficult moments, but also many fun ones.

I also want to thank Nicoletta and David for helping me when most needed. And Moreno, as without him all this would probably never have happened.

Finally, I want to thank my family for their unconditional love and for always supporting my dreams.

Abstract

It is a well-established fact that increasing firms' productivity is a necessary step to achieve sustainable growth and development. In fact, low levels of productivity, in particular in the registered manufacturing sector, represent a major challenge for the Indian economy.

One key obstacle faced by these firms is the high level of employment protection, which makes it difficult to compete with the other sectors that largely rely on informal labour or flexible contracts.

High labour protection increases the incentive for workers to be absent from the factories whenever they have access to better job opportunities elsewhere. Moreover, India is undergoing a process of structural transformation, which is characterised by movement of workers from agriculture into manufacturing. During this process workers are often engaged in both sectors, particularly so across seasons. In fact, the lack of job opportunities during the lean agricultural seasons allows manufacturing firms to pay relatively low wages but, during the peak seasons, workers may find casual jobs in agriculture attractive and leave the firm temporarily.

Using firm level data, representative of the entire registered manufacturing sector, I find that absence rates are very high and negatively correlated with firms' productivity. In particular, I notice that absence rates tend to be highest when labour demand in agriculture is highest, i.e. during the harvest seasons. Using worker level data from a large jute mill, I find that this behaviour is most common among workers who are recent migrants from rural to urban areas who have access to agricultural jobs in their home villages.

I exploit exogenous shocks to agricultural productivity, that increase seasonality in agricultural labour demand, to estimate the effect of seasonal absences on firms' output. Finally, I develop a theoretical framework that that illustrates how seasonal absences can be interpreted as a consequence of asymmetric labour market rigidities between the two sectors and estimate the cost of these rigidities in terms of loss in manufacturing output and employment.

Contents

1 Introduction **8**

2 Manufacturing and Agriculture: two Interlinked Sectors **15**

2.1 The Manufacturing Labour Market 15

2.1.1 Labour Regulation and Workers' Absences 17

2.1.2 Links between Manufacturing and Agriculture . . 19

2.1.3 Description of Absence Data 20

2.2 The Agricultural Labour Market 23

2.2.1 Data and Descriptive Statistics 24

2.2.2 Empirical Strategy 29

2.2.3 Results 30

3 Theoretical Framework **41**

3.1 Workers' Attendance Decision 41

3.2 Agricultural Sector 42

3.3 Manufacturing Sector 43

3.3.1 Flexible Labour Market Outcome 44

3.3.2 Rigid Labour Market Outcome 45

3.3.3 Comparison Between Rigid and Flexible Labour Market Outcomes 46

3.4 Heterogeneous Workers 48

Appendices **52**

3.A Proofs 52

4 Seasonal Absences and Industrial Performance **55**

4.1 Overview 55

4.2 Data 57

4.2.1 Firm Level Data 57

4.2.2 Agricultural Data and Weather Shocks 59

4.3 Summary Statistics 60

4.4 Evidence of Seasonal Absences 61

4.5 Elasticity of Workers' Attendance with respect to Agricultural Wages 63
4.6 Effect of Seasonal Absences on Productivity 65
4.6.1 Empirical Strategy 67
4.6.2 Results 68
4.6.3 Robustness 68
4.7 Calibrating the Effect of Labour market Rigidities 69

5 Seasonal Absences and Rural-Urban Migration: Evidence from an Indian Jute Mill 80
5.1 Overview 80
5.2 Context 81
5.3 Data and Descriptive Statistics 83
5.3.1 Payroll Data 83
5.3.2 Output Data 85
5.3.3 Crop Calendar and Weather Shocks 85
5.3.4 Seasonal Absences 86
5.4 Empirical Strategy 87
5.5 Results 89

6 Conclusions 102

List of Figures

2.1 Absence Rates by District 22
2.2 Harvest and Sowing Seasons 26
2.3 Agricultural Wages 29
2.4 Weather Shocks and Agricultural Wages 33

3.1 Rigid vs Flexible Outcomes 48

4.1 Distribution of Monthly Absence Rates 71
4.2 Seasonal Absences 72
4.3 Agriculturas vs Manufacturing Wages 73
4.4 Gains from Flexible Labour Market 73

5.1 Number of Workers by District of Origin 92
5.2 Distribution Absence Rates Over Time 93
5.3 Distribution Workers' Absence Rates 93
5.4 Rice Harvest by Calendar Month 94
5.5 Seasonal Absences 94
5.6 Seasonality in Output 95

List of Tables

2.1 Workers' Absence Rates by State 35
2.2 Determinants of Workers' Absences 36
2.3 Summary Statistics Agricultural Sector 37
2.4 Agricultural Wage and Harvest Season 38
2.5 Weather and Crop Yield (First Stage) 39
2.6 Crop Yield and Agricultural Wages (2SLS) 40

4.1 Summary Statistics 74
4.2 Workers' Attendance - Monthly Level 75
4.3 Workers' Attendance - Yearly Level 76
4.4 Effect of Seasonal Absences on Productivity 77
4.5 Robustness Checks 78
4.6 Parameters Values 79
4.7 Calibration Results 79

5.1 Summary Statistics 96
5.2 Characteristics of the Migrants 97
5.3 Seasonality in Workers' Absences 98
5.4 Seasonal Absences and Agricultural Productivity 99
5.5 Seasonal Absences and Distance from Workers' Districts of Origin 100
5.6 Seasonal Absences and Productivity (2SLS) 101

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, whilst labour requirements in the manufacturing sector are generally constant over the year, in agriculture they are highly seasonal. Therefore, a labour surplus during the lean agricultural months is counterbalanced by a labour shortage during the peak season. Consequently, agricultural wages increase and may end up attracting manufacturing workers temporarily back to the fields.

The purpose of this thesis is to analyse the effect of seasonal movement of workers across sectors on manufacturing productivity. In particular, I evaluate the effect of seasonal absences of manufacturing workers during the peak season in agriculture, when their outside option is higher. Moreover, I analyse to what extent labour market rigidities, that prevent firms from adjusting wages seasonally, end up reinforcing these productivity effects, slowing down structural transformation by reducing movements of workers towards manufacturing.

Using newly assembled data for India, which provide rich information on the timing of the agricultural production cycle, crop yield, agricultural wages, weather as well as manufacturing output and workers' attendance, I produce novel evidence of seasonal fluctuations in the agricultural labour market and their causal impact on workers' absences and productivity in the manufacturing sector. I show that agricultural seasonality is enhanced by weather shocks that affect agricultural productivity and, as a consequence, agricultural labour demand and wages during the harvest season. Such shocks provide a source of exogenous variation that can be used to explain part of the year to year fluctuations in the workers' absences and quantify their impact on firms' productivity.

Workers' absences are a serious issue in this context. Indeed, every year they cause the loss of 9.25%[1] of working days in the Indian registered manufacturing sector, a very high rate if compared to 1.2% in the US and 3.3% in Canada[2]. Anecdotal evidence suggests that firms are greatly harmed by workers' absences[3] but, to my knowledge, no attempt has been made by the economic literature to quantify their cost.

Moreover, there is qualitative evidence that manufacturing workers in India are more likely to be absent during the peak agricultural seasons (Khurana et al. (2009) and Basariya (2015)) and that this is particularly common among workers with an agricultural background, who tend to return to their village of origin to take care of agricultural duties (Kumar et al., 2014). However, there has been no attempt to analyse this phenomenon quantitatively, nor to understand its consequences on firms' performance, much less to evaluate the extent to which labour market rigidities are responsible for such fluctuations and for the slowdown in the movement of workers from agriculture into manufacturing.

India is an ideal setting for this research: its economy is still dominated by the agricultural sector, which, according to the Population Census, in 2001 was employing 53% of the labour force, but its future growth crucially depends on the development of the manufacturing sector. Moreover, the size of the country allows to exploit local shocks to agricultural productivity, affecting local agricultural labour market outcomes (Jayachandran, 2006), while controlling for aggregate shocks that affect the whole economy and, in particular, output demand. Finally, the existence of heterogeneity in labour regulation across states makes it possible to analyse the role of labour protection in determining workers' attendance behaviour.

This thesis is divided into four parts. In the first part, I provide a description of manufacturing and agricultural labour markets in India and the linkages between the two sectors. I discuss the legal setting in

[1] Estimate based on ASI 2000-01 to 2007-08 data.

[2] The US figure comes from CPS 2012 and the Canadian figure comes from LFS 2011.

[3] For example Kumar (2010) argues that *"one of the most serious problems with which [Indian] industries are confronted today is that of absenteeism."* as it *"erodes the very potentiality, credibility and productivity of any company and organisation"*

which manufacturing firms operate and provide evidence on how labour market rigidities may play an important role in determining workers' attendance by showing correlations between the degree of employment protection and absence rates.

I use rich data on Indian agriculture to document both seasonal and year to year fluctuations of agricultural wages. I construct a district level monthly dataset including information on intensity of agricultural activity, agricultural output and weather shocks covering the 8 year period corresponding to the fiscal years 2000-01 to 2007-08. This is obtained by matching information on yield and cultivated area for the major crops with their district specific sowing and harvest calendar and data on weather realisations during the crop's growing season. I exploit the heterogeneity in crop harvest calendar across Indian districts to show that agricultural wages are higher when during the harvest period. Moreover, I use weather shocks as a source of exogenous variation in agricultural productivity and find that 1% increase in crop yield increases agricultural wages by 0.16% during the months of harvest.

In the second part, I develop a model to understand how fluctuation in agricultural productivity can affect manufacturing firms' performance in a context in which workers are allowed to move across sectors. Manufacturing firms choose labour and wages knowing that, to encourage workers attendance, they need to provide higher wages that would compensate them from the lost opportunity to work in agriculture. I compare the case in which firms are allowed to adjust wages and labour seasonally and in response to shocks to agricultural productivity, to the case in which labour market rigidities prevent these adjustments[4]. I show that, if absences are costly in terms of productivity and firms can adjust wages seasonally, it is optimal to increase wages to incentivise workers' attendance in the periods in which labour demand in agriculture is high. In particular, firms would do so to the point that there would be no seasonal fluctuation in attendance. On the other hand, when labour market

[4]This scenario is meant to represent the current institutional framework in which the firms object of this study operate. Indeed, while the agricultural sector typically employs casual labour for short term jobs, firms operating in the registered manufacturing sector are forced offers permanent contracts as the Contract Labour Regulation and Abolition Act of 1970 established that manufacturing companies are allowed to contract out only work which is non-core, non-perennial, and casual in nature. Moreover, manufacturing wages are usually fixed and firms have little possibility to sanction workers for being absent as they are subject to the Industrial Disputes Act of 1947.

rigidities prevent firms from adjusting wages seasonally, attendance rates would fluctuate giving origin to "seasonal absences". I show that this would generate lower average attendance and productivity, lower wages and potentially decrease manufacturing employment.

In the third part, I use firm level data representative of the entire Indian registered manufacturing sector, collected through 8 rounds of the Annual Survey of Industries (ASI), covering the period from 2000-01 to 2007-08. I enrich the main dataset by matching it with monthly data on workers' attendance, collected in a separate part of the survey, which, to my knowledge, was never used to perform econometric analysis. Exploiting differences in agricultural harvest calendar across Indian districts, I find that workers' attendance is lower during the labour intensive harvest months. I use exogenous weather shocks to agricultural productivity, that affect labour demand during the harvest season, to estimate the response of manufacturing workers' attendance to changes in their outside option. This approach allows me to compare workers' attendance, within district, across harvest seasons. I find that a 1% increase in yearly agricultural wages, caused by positive weather shock, results in a 0.14% decrease in yearly industrial workers' attendance, a result that is both statistically and economically significant. Finally, I estimate the effect of this phenomenon in terms of output loss, finding large results: 1% decrease in attendance rate, caused by a positive weather shock, reduces manufacturing output by over 6%. The interpretation of these findings is that seasonal workers' absences cause large disruptions in the firm production process, possibly because many workers leave simultaneously and are difficult to replace, which reduces firms' output much more than a change in the labour force of equal size.

I use these results to calibrate the model described above and estimate the cost of labour market rigidities in terms of output and manufacturing employment. My findings suggest that allowing firm to adjust wages and labour in response to temporary changes in agricultural productivity would increase average manufacturing output by 5.13%. Moreover, this increase in productivity will result in an increase in average manufacturing employment by 3.36%, without decreasing average wages. These results suggest that labour market rigidities not only decrease manufacturing productivity, but also slow down the process of structural transformation by keeping more workers in the agricultural sector.

Finally, I use personnel data from a large jute mill located in West Bengal to perform an insider econometric exercise which sheds additional light on the the phenomenon of seasonal absences and its impact on firms' outcomes. The jute industry provides an interesting setting for this study because, for historical reasons, it employs a large number of workers who migrated to West Bengal from neighbouring states (De Haan, 1997). These workers typically come from rural areas and move to the city in search of employment. However, in most of the cases, they leave their families behind, thus keeping strong ties with their villages of origin. I find that workers whose place of residence is a village outside the district in which the plant is located are much more likely to engage in seasonal absences than local workers. Taking advantage of high frequency output data, I estimate the impact of seasonal absences on firms' productivity using the timing of harvest as an instrument for workers' attendance. The results suggest that 1% decrease in workers' attendance during the harvest season decreases the firms' output by 1.6%. This more than proportional response suggests that the firm is unable to adjust its production process in response to changes in workers' attendance and that increasing the proportion of absent workers decreases the average productivity of those present by generating disruptions and bottlenecks.

This thesis contributes to the literature on structural transformation[5] by looking at how the interaction between agricultural and manufacturing sector affect industrial performance. An important question addressed by this literature is whether increases in agricultural productivity are necessary to foster growth in other sectors of the economy (Rosenstein-Rodan, 1943). Recent studies pointed out that, in order for improvement in agriculture to have a positive effect on industrial development, these changes must be "labour saving" (Bustos et al. (2016) and Foster and Rosenzweig (2004)). Consistent with these findings, changes in agricultural productivity that generate an increase in labour demand, such as positive weather shocks, may end up attracting workers back to the fields. Moreover, my model suggests that adding asymmetric labour market rigidities to this context amplifies this effect and slows down the structural transformation process by decreasing manufacturing employment.

The process of structural transformation was traditionally associated

[5] See Herrendorf et al. (2013) for a recent survey.

to the migration of workers from rural to urban areas (Harris and Todaro, 1970). A recent strand of literature, however, has pointed out that it is even more common for rural workers to seek employment in urban areas during the lean agricultural season and then return to their village of origin during the peak season (Bryan et al. (2014), Morten (2013), and Imbert and Papp (2016b)). This thesis contributes to this literature as it extends the concept of seasonal migration by showing that even workers with permanent jobs in the manufacturing sector are likely to return to the fields at the time of harvest.

The idea that seasonality in agricultural labour demand could harm manufacturing firms in the early stages of industrialisation finds support in the economic history literature. Sokoloff and Dollar (1997) argue that the higher degree of seasonality characterising agriculture in England, during the industrial revolution, favoured the development of cottage industry rather than more productive centralised plants that where prevalent in the US at the time. This is because centralised plants, being more capital intensive, require constant labour inputs and therefore suffer more from workers' absences. Similarly, Sokoloff and Tchakerian (1997) found that in 1860, US manufacturing firms located in counties where the dominant crop had a more seasonal labour demand had a significantly lower TFP than those located in counties specialised in less seasonal agricultural activities. This thesis provides a causal estimate of the effects of this phenomenon in an economy that, in present days, is still dominated by the agricultural sector.

Moreover, this thesis contributes to the literature on the consequences of labour market rigidities on manufacturing outcomes in India (Fallon and Lucas (1993), Besley and Burgess (2004), Ahsan and Pagés (2009), Aghion et al. (2008), Hasan et al. (2007), Adhvaryu et al. (2013), Chaurey (2015)) by noting that one of the mechanisms in which the high degree of labour protection characterising the sector may have a negative effect on productivity and employment is by providing provides incentives for workers to be absent from the workplace[6].

Finally, this thesis contributes to the literature on the effect of workers' absence on firms' productivity. Absences are usually considered a negative outcome, as they create an extra cost for the firm in terms of

[6] Ichino and Riphahn (2005)and Riphahn (2004) find that labour protection increases workers' absences in different settings.

reassignment of tasks to workers and substitutes (Allen, 1983). However, their effect on productivity, measured in terms of output per hour effectively worked, is not necessarily negative. In fact, present workers may work harder to compensate for the absent ones and productivity per hour effectively work may increase. Estimating the cost of workers' absence for the firm is, therefore, an empirical matter, which remains largely unexplored by the literature. The major difficulty in this exercise is a clear endogeneity problem: less productive firms are more likely to face higher absence rates. This may happen for several reasons: for example poor management quality may reduce both productivity and workers motivation; alternatively, firms facing a negative demand shock may encourage workers to be absent if they cannot dismiss them. One solution to this problem is to consider individual workers' attendance and productivity as did Miller et al. (2008) and Herrmann and Rockoff (2012) focusing on the effect of teachers' absence on students' outcomes. 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. This thesis solves this endogeneity problem by exploiting the fact that workers' absences in the manufacturing sector can be predicted by changes in agricultural productivity, which are caused by exogenous weather shocks. These shocks provide a source of exogenous variation that allows me to causally estimate the effect of seasonal absences on firms' productivity. Indeed, by using an instrumental variable approach, I am able to identify the local average treatment effect of changes in workers' absences that are caused by shocks to agricultural productivity, which, in turn, affect only seasonal absences.

The remainder of this thesis proceeds as follows: chapter 2 describes the relevant features of manufacturing and agricultural labour markets in India; chapter 3 proposes a theoretical framework; chapter 4 provides empirical evidence based firm level data representative of the entire sector; chapter 5 provides additional evidence and insights based on worker level data from a large textile plant; chapter 6 concludes.

2. Manufacturing and Agriculture: two Interlinked Sectors

2.1 The Manufacturing Labour Market

The empirical work at the basis of this thesis relies on firm level data representative of the whole Indian registered manufacturing sector. This includes all firms with more than 10 workers[1] involved in the manufacturing process that are officially registered according to the Factories Act of 1948. This sector is often referred to as the "organised" manufacturing sector and accounts for approximately two thirds of the Indian manufacturing output, which represents about 16% of Indian GDP. However, it employs only for about 20% of the manufacturing labour force as most manufacturing workers are employed in the informal sector (Sincavage et al. (2010), data covering the period 2000-06).

As in most developing countries, the manufacturing sector in India are characterised by low levels of productivity and slow growth (Bartelsman et al., 2013). These issues have been explained by the literature as a consequence of legal constraints (Besley and Burgess, 2004), poor management practices (Bloom et al., 2010) and misallocation of resources (Hsieh and Klenow, 2009). This thesis focuses on a less known phenomenon that partly explains this weak performance: the fact that firms' operating in this sector suffer from very high levels of workers' absence, which cause the loss of, on average, 9.25% of working days every year[2]. This rate is much higher than those recorded in developed countries, corresponding

[1]The 10 workers threshold applies to firms using electricity, for firms without electricity the threshold is 20 workers.

[2]Estimate based on ASI 2000-01 to 2007-08 data.

to, for example, 1.2% in the US and 3.3% in Canada[3].

In order to understand why absence rates are so high, it is important to consider the legal framework that regulates employment relationships in this context. While the majority of the Indian labour force is employed in the informal economy, workers employed in the registered manufacturing sector enjoy a very high level of employment protection under the Industrial Disputes Act (IDA) of 1947. Indeed, this act makes it extremely costly for firms to sanction workers, even in case of prolonged unjustified absences.

The literature has shown that high employment protection is likely to give origin to moral hazard problems resulting in shirking and absenteeism (Ichino and Riphahn (2005); Riphahn (2004)). However, no attention has been paid to the fact that it may also create incentives for workers to be absent whenever they have better job opportunities in other sectors. This is particularly relevant in a development context, where a large share of manufacturing workers come from a rural background and have easy access to jobs in agriculture. Since agricultural labour demand and wages vary seasonally, so will manufacturing workers' outside option and, therefore, their incentive to be absent.

The Annual survey of Industries (ASI), which is the main source of data on firms operating in the Indian registered manufacturing sector, collects firm level information on the number of days lost because of workers' absences[4] at a monthly frequency. The analysis reported in chapter 4 will use data from 8 rounds of the survey, covering the period 2000-01 to 2007-08 to provide causal evidence of the existence of the phenomenon of "seasonal workers' absences", defined as absences of manufacturing workers that occur during the period in which labour demand in agriculture is high, by exploiting heterogeneity in crop harvest calendar across districts. Moreover, it will exploit exogenous shocks to agricultural productivity to identify the effect of this phenomenon on firms' productivity.

[3] The US figure comes from CPS 2012 and the Canadian figure comes from LFS 2011.

[4] For a detailed description of the dataset refer to section 4.2.1

2.1.1 Labour Regulation and Workers' Absences

The aim of this section is to illustrate the legal framework in which registered manufacturing firms in India operate, how it differs across states and how it may affects workers' absence behaviour. In section 2.1.3 I will exploit differences in labour regulation across states to explain part of the geographic variation in workers' absences. In particular, I will show that absence rates are substantially higher in states with stronger employment protection regulation. This heterogeneity will be explored further in chapter 4 where I show that the elasticity of workers' attendance with respect to agricultural wages is higher in the states where workers are more protected, suggesting that employment protection not only increases the incentives to shirk but allows workers to take advantage of attractive outside opportunities.

Unlike the majority of the Indian labour force, regular[5] workers employed in the registered manufacturing sector enjoy a very high level of labour protection as their employment relationships are regulated by the Industrial Disputes Act (IDA) of 1947. This piece of legislation covers resolution of industrial disputes by setting up tribunals and labor courts and regulating hiring and firing of workers, closure of establishments, strikes and lockouts etc. An important feature of the IDA is that it virtually prohibits firing of workers, especially for large establishments (with more than 100 workers[6]) that are required to obtain government permission for dismissing even a single worker.

In practice, this restricts the possibility for firms to sanction workers even when they are absent from the workplace for a long period of time. In fact, even if the firm has the possibility to remove workers from their payroll if they take, long[7], unauthorised leaves, workers that appeal to a labour courts are usually reinstated (Kumar, 2010).

[5] These are workers directly employed by the firm on a regular basis, i.e. workers employed through contractors, casual workers and apprentices are excluded.

[6] State amendments changed this threshold to 50 in West Bengal and 300 in Uttar Pradesh.

[7] In principle workers who are absent for long periods could be removed from the payroll on the presumption that they have abandoned their job. However, the law does not clearly specify how long these absences need to be and the Supreme Court repeatedly ruled that firing workers under such presumptions is illegal and constitutes retrenchment. This implies that the firm has to hold an enquiry, prove the worker's misconduct and (if needed) seek government permission to terminate the employment relationship. Only for extreme cases, such as workers being absent from the workplace for over a year the Supreme Court overturned reinstatement sentences of labour courts admitting that the presumption of abandonment was justified (Kumar, 2010).

A rich literature has studied the effects of the IDA and, in particular, of its employment protection provisions contained in sections V-A and V-B, on firms' outcomes. Fallon and Lucas (1993) find that these provisions largely decreased employment in large manufacturing firms, i.e. those that are subject to the strictest labour protection provisions. Besley and Burgess (2004) exploited the fact that, although the IDA applies to all India, it has been amended several times by state governments, creating spacial variation in the degree of labour protection regulation. They considered all state level amendments of to sections V-A and V-B of IDA 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 and found that having a more "pro-worker" regulation reduced firms' output, employment, investment and productivity. These findings were confirmed by a Ahsan and Pagés (2009) who find that the negative effect of labour protection regulation on employment and output is larger in states where the IDA makes it harder to resolve labour disputes. Aghion et al. (2008) found that firms located in "pro-employer" states grew more as a consequence of delicensing and Hasan et al. (2007) showed that labour demand elasticities increased more in these states as consequence of trade liberalisation. Finally, Adhvaryu et al. (2013) found that firms operating in "pro-worker" states adjust less their labour force in response to local demand shocks and Chaurey (2015) found that in these states firms are more likely to use contractors to adjust their labour force in response to shocks, which can be explained by the fact that contract workers are not protected by the IDA.

The fact that a high degree of job protection may increase shirking and decrease attendance is not new to literature (Ichino and Riphahn (2005); Riphahn (2004)). However, this phenomenon is usually associated to the behaviour of public sector workers that, especially in developing countries, are not monitored effectively and get paid even when they are absent from the workplace (Chaudhury et al. (2006); Banerjee and Duflo (2006)). Things are different for privates sector workers as firms have high incentives to monitor their attendance and punish them from being absent. In fact, even firing is almost impossible, the Payment of Wages Act of 1936 allows Indian registered manufacturing firms to reduce the workers' wages by up to 100% when they are absent[8]. This thesis will show that, even for private sector workers, a higher degree of labour protection provides incentives to be absent from the workplace, whenever better employment opportunities are available.

2.1.2 Links between Manufacturing and Agriculture

This thesis focuses on "seasonal workers' absences" that I define as those absences that occur when labour demand in agriculture is high and workers leave the firm temporarily to take advantage of job opportunities in this sector. The purpose of this section is to explain the relevance of this phenomenon in a context of structural transformation, in which manufacturing and agricultural sectors are closely related, to the point we can say they share the same labour force.

The process of structural transformation is characterised by the moment of workers away from agriculture and into the manufacturing sector (Lewis, 1954). However, during this process, workers are often involved in both sectors, particularly so across season. This is because labour requirement in agriculture are low for the best part of the year, generating a labour surplus that can be exploited by the manufacturing sector to hire workers for relatively low wages, but they increase dramatically during the harvest season, causing many workers to return to temporarily to the fields.

Indeed, while the classical literature was based on the idea that structural transformation involves permanent moments of workers from rural to urban areas (Harris and Todaro, 1970), a recent literature has showed that it is even more common for rural workers to seek employment in urban areas during the lean agricultural season and then return to their village of origin during the peak season (Bryan et al. (2014), Morten (2013), and Imbert and Papp (2016b)). In this thesis I extend the concept of seasonal migration one step further, showing that even workers that enjoy permanent jobs in the manufacturing sector are likely to return to the fields during the harvest season.

Although there is abundant anecdotal evidence that manufacturing workers are more likely to be absent from the workplace during the peak

[8] The only exception is in case of statutory leave: the Factories Act of 1948 states that workers who work for more than 240 days in a calendar year are entitled to 1 day of paid leave for every 20 days of work. Paid leave should be authorised by the firm and scheduled in advance.

agricultural seasons (Khurana et al. (2009) and Basariya (2015)) and that phenomenon is driven by workers with an agricultural background who tend to leave tend to return to their village of origin to take care of agricultural duties (Kumar et al., 2014), this thesis is the first attempt to provide quantitative evidence of this phenomenon and to causally estimate its impact on industrial performance.

An important question addressed by the literature on structural transformation is whether increases in agricultural productivity are necessary to foster growth in other sectors of the economy (Rosenstein-Rodan, 1943). Recent studies pointed out that, in order for improvement in agriculture to have a positive effect on industrial development, these changes must be "labour saving" (Bustos et al. (2016) and Foster and Rosenzweig (2004)). In other words, only technological changes that decrease labour demand in agriculture would generate movement of workers towards other sectors of the economy producing positive "forward linkages". Consistent with these findings, temporary changes in agricultural productivity that generate an increase in labour demand, such as positive weather shocks, may end up attracting workers back to the fields.

In this thesis I show that this phenomenon has the additional effect to slow down the structural transformation process by generating substantial productivity losses in the manufacturing sector, which, in turn result in a lower level of manufacturing employment. The model proposed in chapter 3 highlights how this effect can be interpreted by the consequence of asymmetric labour market rigidities across sectors.

These findings are supported by the work of economic historians such as Sokoloff and Dollar (1997) and Sokoloff and Tchakerian (1997) who, looking at England and the US during the industrial revolution, found that higher degrees agricultural seasonality hindered the development of capital intensive centralised plants. They argue that this can be explained by the fact that capital intensive plants require constant labour input, which is hard to attain when workers are moving across sectors based on the agricultural calendar.

2.1.3 Description of Absence Data

The empirical evidence on workers' absence at the basis of this thesis relies on firm level data from part II of the Annual Survey on Industries (ASI). Although the ASI is the most widely used source of information on firms operating in the Indian manufacturing sector, researcher have so far utilized only part I of this survey, which contains detailed information on firms output and input use. Although ASI part II, which contains monthly information on absences and labour turnover, is collected regularly for the same set of firms, it is not publicly available and, to my knowledge, has never been use for research in economics[9]. The purpose of this section is to provide a general description of the data, with the objective to uncover some unknown facts.

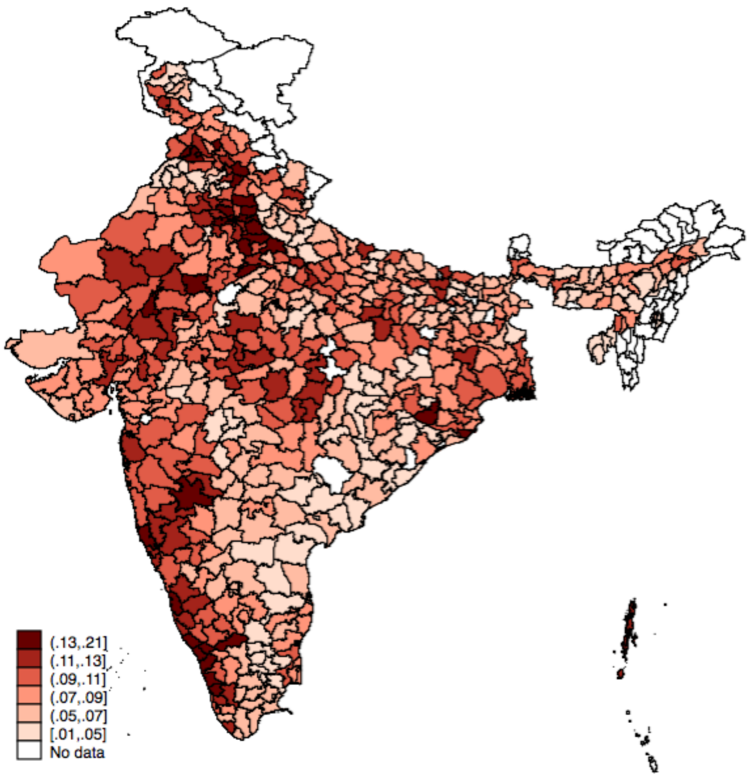
I use data from 8 rounds of the survey, covering the period 2000-01 to 2007-08 and I restrict the sample to the firms for which both part I and part II of the survey are available and can be successfully matched. The final dataset includes 211,406 firm-year observations.

The definition of workers' absence for the Indian legal system and for the ASI data is "failure of a worker to report for work when he/she is scheduled to work". This includes: absence with or without pay, with or without permission, sick leave and absence for personal reasons. It should not include statutory leaves as these should be planned in advance and the worker should not be scheduled to work. In practice, unscheduled absences may be treated as statutory leave, i.e. the workers receive their wages, as the management often fails to coordinate with workers and schedule vacations and find it easier to discount the days of absence from the workers' leave allowance[10].

[9] I thank the Indian Labour Bureau for kindly sharing the data with me.

[10] Anecdotal evidence collected during firm visits in India.

Figure 2.1: Absence Rates by District



Notes: The figure shows the spatial distribution of average absence rates across districts.

Figure 2.1 plots the spatial distribution of years average absence rates across Indian districts. It shows that absence rates are generally high across the country but there is a strong spatial correlation with certain states presenting a substantially higher level of absences. Table 2.1 reports summary statistics for workers' absence rates by state. It shows that the highest absence rates are observed in Kerala (12.66%) and the lowest in Chhattisgarh (4.88%). The overall average is 9.27%.

To assess whether absence rates are higher in states in which labour protection regulation is stronger I follow Adhvaryu et al. (2013) and classify as "pro-worker" the states that passed more amendments in "pro-worker" than in "pro-employer" direction. These are Maharashtra, Orissa and West Bengal[11]. Column 1 of Table 2.2 shows that ab-

[11] In 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 of Gujarat was criticised by the literature (Bhattacharjea, 2006)(ii) during the period considered some "pro-employer" reforms where implemented in the state.

sence rates are on average 1.5 percentage point higher in "pro-worker" states. Column 2 show that absence rates are on average lower in large firm (defined as firms with more than 100 regular workers), however the interaction between large and pro-worker is positive indicating a larger absence rate in large firms that operate in pro-worker states, which is where the workers enjoy the highest level of protection. These finding suggest that labour regulation may play an important role in determining workers' absence behaviour. Indeed table 2.1 shows that the states classified as "pro-worker" and in particular Maharastra indeed report an above average level of absences. However, this classification may not fully incorporate some important features of the labour markets, in fact, the highest levels of absence are recorded in Kerala, a state that, even if not classified as "pro-worker", has a long tradition of left wing governments and powerful labour unions.

Another interesting fact, pointed out by column 3 of 2.2, is that absence rates are 1.6 percentage point higher in urban areas than in rural areas and, as column 4 shows, this effect is stronger in districts in which the percentage of rural migrants is higher. As rural migrants are more likely to engage in seasonal absences, this result suggest that the phenomenon could explain a large part of the geographic variation in absence rates.

2.2 The Agricultural Labour Market

An important feature of the agricultural labour market in developing counties is the prevalence of short term casual contracts (Kaur, 2013). Their availability is highly seasonal as it depends on the agricultural cycle. Labour demand is higher during the months of harvest and much lower during the rest of the year. As a consequence, seasonal unemployment in rural areas is a common phenomenon (Morten (2013) and Bryan et al. (2014)).

Moreover, the amount of labour required for harvesting changes from one year to another, depending on agricultural productivity, which is turn affected by weather conditions (Jayachandran (2006)). If rains are abundant, agricultural yield is higher and more labour is required to harvest, which increases agricultural wages.

In this section I test whether seasonality and shocks to agricultural productivity are reflected into changes in agricultural wages. To do so I construct a monthly panel dataset of crop yield, weather shocks and agricultural wages at the district level for the period between 2000 and 2008.

While the causal relationship between crop yield and agricultural wages has already been established by the literature, little attention has been devoted to seasonality. Using data for India, 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 to what extent it affects agricultural wages in all months of the year. Moreover, little is known about the role of rainfall and temperature outside the monsoon season.

Since the empirical analysis proposed in this thesis relies on the existence of exogenous variation in monthly agricultural productivity, necessary to establish the existence of seasonal absences, I construct a monthly measure of "weather shocks" based on the crops harvested in each month and their growing season and I show that it has explanatory power even when controlling for monsoon season rainfall and temperature.

2.2.1 Data and Descriptive Statistics

Crop production data

I construct a district level monthly crop calendar based on the major crops cultivated in the district. I match crops to their district-specific sowing and harvesting calendar and then I aggregate the data into district level weighted averages, using the area cultivated under each crop as a weight. This allows me to obtain a monthly measure of the intensity of sowing and harvesting activities, which varies across districts.

The information about the agricultural production cycle comes from the 1967 Indian Crop Calendar published by the Directorate of Economics and Statistics. This publication contains information about the typical sowing and harvesting months of the most important crops at the district level. For each crop, I define as growing season the period delimited by the first month of sowing and the last month of harvest. To match 1967 districts with 2001 districts I refer to Kumar and Somanathan (2009), who provide a mapping of Indian districts over time. When the crop's sowing and harvesting information is not available for in a district, I impute it using the closest available data 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.

I match the crop calendar information with district level data on crop output and area sown for major crops for the period between 1999-00 and 2007-08, obtained from the Crop Production Statistics Information System website[12]. The same website provides district level information on land use over the same period. Combining these datasets allows me to construct monthly measures of crop yield and percentage of area harvested.

I calculate the share of cultivated area corresponding to each crop dividing the area under crop by the total area sown in the district in a particular year. I restrict the sample to crops that cover at least 1% of the total area sown in the district. Furthermore, I exclude from my analysis districts for which I have information about crop covering less than 50% of the total area sown. Finally, I keep only crops for which sowing and harvest dates are provided in the crop calendar described above[13].

I define crop yield as crop output divided by area under crop. For comparability across crops I normalize yields to have mean one. I then aggregate crop yields to obtain a district level measure, *log crop yield*

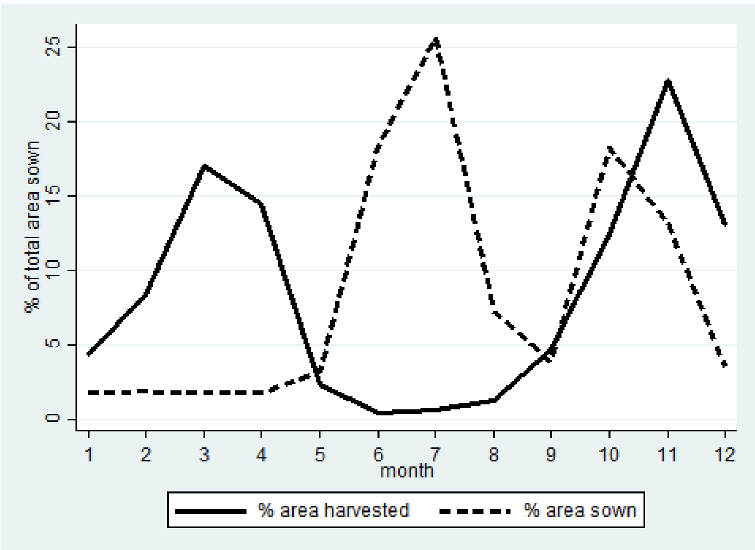
[12]The data were downloaded from the Crop Production Statistics Information System website: apy.dacnet.nic.in in November 2014.

[13]The excluded crops are: cashewnut, guar seed, khesari, maize (rabi), moth, oilseeds and soyabean.

index, constructed as weighed average of the log of yield of the crops harvested in the district in a particular month. The weights are equal to the share of the area under each crop[14]. If a crop is sown (harvested) over more than one month, I divide the area cultivated under this crop by the number of sowing (harvesting) months and impute it equally to each of them.

I use the area under each crop, combined with the crop's sowing and harvesting dates to construct a district level calendar, indicating the intensity of sowing and harvesting activities in each calendar month. Figure 2.2, plotting the average share of agricultural area sown and harvested in each calendar month, illustrates the main features of the Indian agricultural cycle. There are two main growing seasons: *rabi* and *kharif*. *Kharif* crops are sown between June and July, when the first monsoon rains arrive, and harvested between October and December; while *rabi* crops are sown between October and November and harvest in March and April.

Figure 2.2: Harvest and Sowing Seasons



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.

[14]Unlike Jayachandran (2006), who constructs a similar index using share of revenue originate by the crop as weight, I choose to use share of area sown as it better proxies the amount of labour required for harvesting.

Agricultural wages

The crop production and weather data are matched with district level information on monthly agricultural wages collected by the Directorate of Economics and Statistics of the Indian Ministry of Agriculture in a yearly publication called "Agricultural Wages in India" (AWI)[15]. I consider only male wages for occupations related to field labour. The data were originally collected at the centre level and, in 10% of the cases, multiple observations per district are available. In such cases I simply take the average of the observed wages. The original dataset covers 374 districts, however, I restrict the attention to the districts for which at least 12 monthly agricultural wage observations are available in the period between January 2000 and December 2007. After dropping a few districts for which agricultural information is not available I have a panel of agricultural wages for 272 districts. This panel is highly unbalanced and, on average, 49.5 monthly wage observations are available over a 96 months' period.

Weather shocks

I construct district-wise monthly weather shocks using use rainfall and temperature data collected by the Center of Climatic Research at the University of Delaware[16]. This dataset contains a time series of average monthly rainfall and temperature interpolated into a 0.5 by 0.5 degree latitude-longitude grid. Using a shapefile of Indian district in 2001, 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 (2015) I construct a crop specific measure of weather shocks, defined as cumulative rainfall or average temperature during the crop's growing season, that is defined as the period of time between the first month of sowing and the last month of harvesting. To aggregate them in a district level measure, I compute the weighted average of the shocks affecting all crops harvested in the district in a given month, using as weights the share of agricultural area under each crop.

[15]The data I am using were kindly shared with me by Thiemo Fetzer who digitised and prepared them for the paper Fetzer (2013).

[16]The data were downloaded from the Center of Climatic Research at the University of Delaware website: http://climate.geog.udel.edu/ climate/ in March 2014

Similarly, a yearly measure of weather shocks is obtained by aggregating the monthly measures in a weighted average, with weights equal to the share of area harvested in each month.

Descriptive Statistics

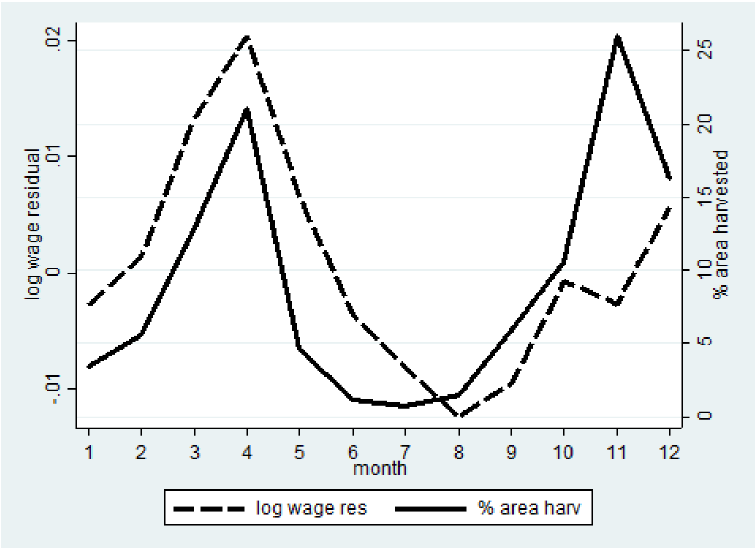
The dataset I constructed includes a monthly panel of district level harvesting and sowing intensity, crop yield index, weather information and agricultural wages. It covers 272 Indian district over a 8 years' period starting on January 2000 and and ending in December 2007. Table 4.1 reports summary statistics for the variables used to perform the analysis described in this chapter.

This panel dataset contains 24,124 district - month observations for crop calendar and weather information. However, the agricultural wage data is highly unbalanced and wages are available only for 55% of the sample: 13,278 observations. This is partly explained by the fact that agricultural wages in some districts are not collected during the lean agricultural season. Finally, crop yield is observed only in the months in which the crops are harvested, that is for 54% of the sample: 12,929 observations. This implies that rainfall and temperature indices, since they are based on the crops' growing seasons, are also available for these months. In order to deal with this issues I include in the analysis a number of fixed effects that allow me to capture the variation of interest.

The average amount of growing season rainfall is 704 mm; while the average temperature is 25.32°C. These figures can be compared with the average amount of rainfall in a month: almost 93 mm; and the average monthly temperature in the districts considered: 25.77°C. Since the average length of a crop's growing season in my dataset is 5.6 months, the amount of rain during the growing season months is above average and temperature is below average, which is what we would expect for India.

The correlation between agricultural wages and the share of area harvested in a calendar month is illustrated by figure 2.3. To remove the composition effect, which is relevant since the panel of agricultural wages is highly unbalanced, the dashed line represents the residual of a regression of $log(ag\_wage)$ in district and year dummies. The solid line represents the share of area harvested in the calendar month, as in figure , with the difference that here the sample is restricted to the observations for which agricultural wages are available. The graph shows that agricultural wages are higher in the months in which a higher percentage of the agricultural area is harvested.

Figure 2.3: Agricultural Wages



Notes: The variable log wage res is the residual from the regression of log of agricultural wages on district FE and year FE. The percentage of area harvested is calculated as the average of the observations for which the agricultural wage is non missing.

2.2.2 Empirical Strategy

In order to estimate to what extent agricultural wages follow the agricultural cycle I exploit the variation in harvest calendar across districts. This is generated by the fact that districts specialise in different crops that may differ in harvest calendar, but also to the fact that crops' harvest calendars vary across districts. I estimate the following equation:

$$log(ag\_wage_{tmd}) = \alpha harvest_{tmd} + \delta_d + e_m + v_t + u_{tmd} \qquad (2.1)$$

where $log(ag\_wage_{tmd})$ is natural logarithm of agricultural wage in district d in month m of year t; $harvest_{tmd}$ represents the percentage of the total agricultural land of district d that is harvested in month m of year t; δ_d, v_t and e_m represent district, year and calendar month fixed effects, respectively. If agricultural wages respond to changes in agricultural labour demand, they should be higher in the months in which a higher percentage of agricultural land is harvested. I cluster standard errors at the district level to account for the presence of autocorrelation.

The next step is to estimate the response of agricultural wages to shocks to agricultural productivity. The empirical model to be estimated is the following:

$$log(ag\_wage_{tmd}) = \beta_1 log\_yield\_index_{tmd} + e_{dm} + v_t + \epsilon_{tmd} \tag{2.2}$$

where $log\_yield\_index_{tmd}$ is a proxy of agricultural productivity in month m of year t in district d, in terms of crop output per acre. In order to obtain an index representing the yield of all crops harvested in the district in a month I normalize the yields of all crops to have mean 1 and then compute the weighted average of the log of all these yields, using as weights the share of the total area harvested in that month corresponding to each crop.

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

$$log\_yield\_index_{tmd} = \beta_2 weather\_index_{tmd} + e_{dm} + v_t + w_{tmd} \tag{2.3}$$

where $weather\_index_{tmd}$ 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-calendar month level in a weighted average, as for the yield index, with weights representing the relative importance of each crops in terms of agricultural area.

I include fixed effect for district-calendar month (e_{dm}) and year (v_t) in both first and second stage regression and I cluster standard errors at the district level to account for the presence of autocorrelation.

2.2.3 Results

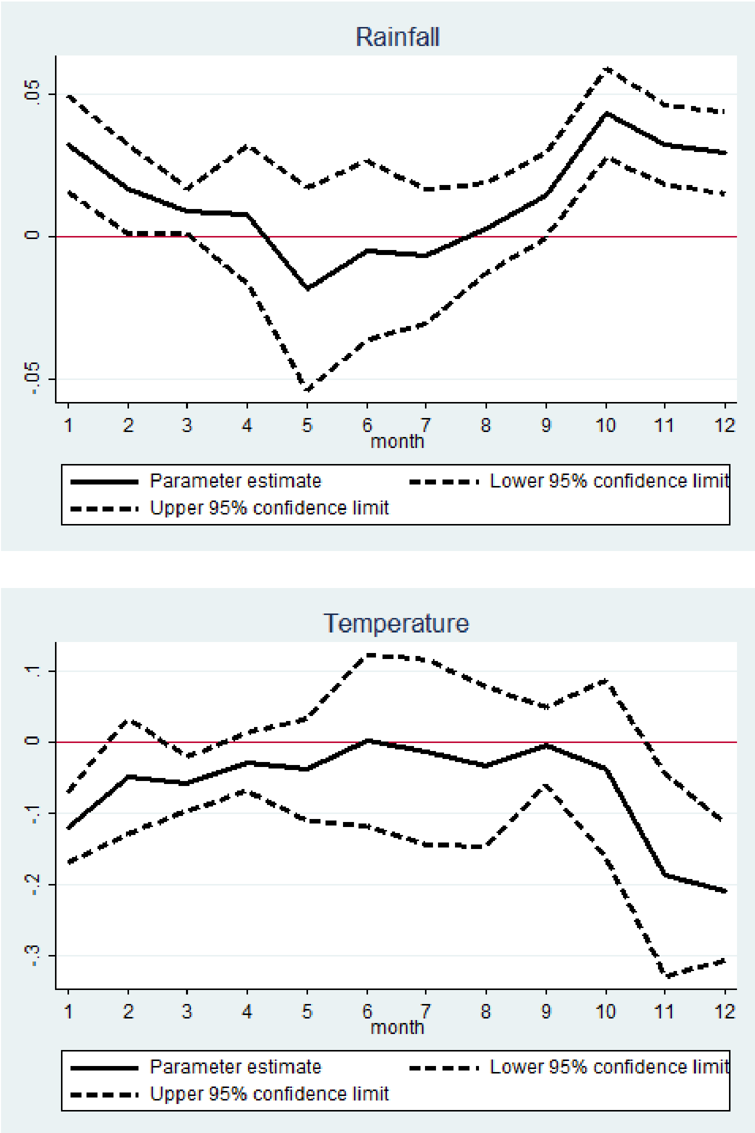
Table 2.4 reports the estimates for equation 2.1. The results show that agricultural wages are higher during the harvest season and the effect is positive and statistically significant. The magnitude of the coefficient suggests that going from 0 of agricultural land harvested to 100% increases agricultural wages by 3.3%. 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). Columns (3) and (4) show that the results are robust to the inclusion of calendar month fixed effects, thus exploiting only the differences in harvest calendar across districts. Finally, columns (2) and (4) suggest that the share of agricultural area sown in the month has no significant effect on agricultural wages.

Once I established that agricultural wages vary seasonally following the agricultural cycle, I test whether this variation is exacerbated by changes in agricultural productivity. By estimating equation 2.2 I compare agricultural wages within each harvest season across years and measure their response to changes in agricultural productivity. Since agricultural productivity and wages may be endogenously correlated I rely on exogenous weather shocks to obtain a causal estimate of the parameter of interest.

Table 2.5 reports the first stage results by estimating equation 2.3. The first 3 columns are estimated on the entire sample for which yield data are available; while the last 3 only on the sub-sample of observations for which also agricultural wages are non-missing. The effect of rainfall on crop yield is positive and the effect of temperature is negative, as expected. In particular, column (3) shows that a 100 mm increase in rainfall over the crops' growing season is estimated to increase crop yield by 2.8%, while 1°C increase in average temperature would reduce crop yield by 7.8%. Column (6) shows that the estimated effects of rainfall and temperature on crop yield are similar in magnitude for the sub-sample in which agricultural wages are available and they are statistically significant. However, the temperature coefficient is smaller 5.4% and less precisely estimated. I include in these regressions controls for contemporaneous monthly rainfall and temperature: the corresponding coefficient are small and non-significant in the specifications that include both growing season temperature and rainfall and have and removing them has a negligible impact on the coefficients of interest.

In order to verify that the effect of weather shocks on crop yield is relevant for the crops harvested in all months of the year I estimate equation 2.3 on one calendar month at a time, the estimated coefficients and confidence intervals are reported in figure 2.4. These results show that the effect of rainfall is positive and significant for the crops harvested between September and March, a period that includes almost the entire *kharif* and *rabi* seasons. The effect of Temperature is negative for all months of the year but it is statistically different from zero only between November and January. The lack of precision of the coefficients estimated for the central month of the year can be explained by the fact that these are based on a small number of observations. As figure 2.2.1 shows, in this period only a small percentage of the agricultural area is harvested and, therefore, crop yield data would be missing in most districts.

Figure 2.4: Weather Shocks and 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 interacted with calendar month dummies. 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, the regression includes controls for contemporaneous monthly rainfall and temperature.

Table 2.6 reports the results of the estimate of the effect of crop yield on agricultural wages. The first stage regression is the same as that reported in table 2.5, column (6). The second stage results, reported in column (2), shows that 1% increase in crop yield results in a 0.156% increase in agricultural wages[17]. This result is very similar to that estimated by the literature[18]. Column (3) shows the reduce form effect of weather on wages: the effect of rainfall is positive and significant and that of temperature is negative but not statistically different from zero.

The results reported in this section indicate that wages in agriculture are higher during the harvest and, across harvest seasons, they are higher when agricultural productivity is higher because of the realisation of positive weather shocks.

[17]The estimated effect of crop yield is similar, however less precisely estimated, when estimated for *kharif* and *rabi* crops separately. Moreover, it is robust to number of different specifications, including: measuring weather variables in terms of deviations from long term mean, including quadratic terms and controlling for monsoon rainfall (for *rabi* crops).

[18]Using shocks to yearly rainfall as instruments for crop productivity, Jayachandran (2006) shows that an increase in crop yield by 1% results in a 0.168% increase in agricultural wages.

Tables

Table 2.1: Workers' Absence Rates by State

State	Mean	Sd Dev	Obs
Kerala	12.66	9.61	6,840
Himachal Pradesh	12.39	6.82	2,986
Haryana	11.48	8.31	9,079
Maharashtra	11.19	7.48	25,375
Uttar Pradesh	10.79	9.10	17,220
Punjab	10.25	9.42	12,730
Rajasthan	10.21	7.66	8,987
Gujarat	9.78	6.84	19,110
West Bengal	9.18	6.87	10,521
Karnataka	8.57	6.15	12,681
Jharkhand	8.24	5.50	3,141
Madhya Pradesh	7.96	6.17	6,051
Orissa	7.88	6.69	3830
Bihar	7.87	6.43	2,191
Uttaranchal	7.57	4.96	2,868
Assam	6.81	4.69	4,084
Tamil Nadu	6.51	4.04	24,416
Andhra Pradesh	5.71	4.17	14,995
Chhattisgarh	4.88	3.32	3,333
All India	9.27	7.28	211,406

Notes: The table reports summary statistics for absence rates at the firm level. The data is obtained from 8 rounds of ASI part II. All manufacturing firms that could be matched with ASI part I dataset are considered. Small states and Union Territories are excluded.

Table 2.2: Determinants of Workers' Absences

	Dependent Var: Absence Rate (%)			
	(1)	(2)	(3)	(4)
Pro-worker	1.526***	1.445***		
	(0.058)	(0.065)		
Large		-0.866***		
		(0.042)		
Large*Pro-worker		0.737***		
		(0.100)		
Urban			1.653***	0.584***
			(0.044)	(0.076)
Share rural migrants				3.206***
				(0.342)
Share rural migrants*Urban				4.603***
				(0.272)
Year FE	Y	Y	Y	Y
Observations	211,406	211,406	211,406	211,406
Adjusted R^2	0.008	0.009	0.012	0.030

Notes: The dependent variable is yearly absence rate (mean 9.28). Pro-worker is dummy equal to one for the states of Maharastra, Orissa and West Bengal; Large is dummy equal to one if the number of regular workers employed by the firm is higher than 100; Urban is dummy equal to one if the plant is located in an urban area. Share rural migrants indicates the share of the districts' manufacturing workers who migrated from a rural area, this information is obtain from the 2001 Indian Census. Robust standard errors in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.3: Summary Statistics Agricultural Sector

Variable	Mean	St Dev	Obs	Source
Log agricultural wage	3.917	0.341	13,278	AWI
Share area harvested	0.088	0.159	24,124	CPSIS
Share area sown	0.087	0.151	24,124	CPSIS
Log crop yield index	-0.117	0.564	12,929	CPSIS and ICC
Rainfall index (100 mm)	7.044	5.992	12,929	Univ of Delaware,CPSIS and ICC
Temperature index (°C)	25.32	3.088	12,929	Univ of Delaware,CPSIS and ICC
Monthly rainfall (100 mm)	0.926	1.326	24,124	Univ of Delaware
Monthly temperature (°C)	25.77	5.26	24,124	Univ of Delaware

Notes: The data cover 272 Indian district over the period 2000-2007. 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. Crop yield is computed as the log of the crop output divided by its area planted, normalized to have mean 1. Cumulative rainfall and Average temperature are measured 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 2.4: Agricultural Wage and Harvest Season

	(1) log(ag wage)	(2) log(ag wage)	(3) log(ag wage)	(4) log(ag wage)
Share area harvested in the month	0.034***	0.034***	0.025***	0.025***
	(0.008)	(0.008)	(0.009)	(0.009)
Share area sown in the month		-0.010		0.002
		(0.006)		(0.007)
District FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Calendar month FE	-	-	Y	Y
Observations	13,278	13,278	13,278	13,278
Adjusted R^2	0.845	0.845	0.846	0.846

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 2.5: Weather and Crop Yield (First Stage)

	Dependent var: *log yield index*					
	Full sample			Non-missing ag wages		
	(1)	(2)	(3)	(4)	(5)	(6)
Rainfall Index	0.029***		0.028***	0.020***		0.021***
	(0.004)		(0.004)	(0.004)		(0.004)
Temperature Index		-0.080***	-0.078***		-0.050*	-0.054**
		(0.020)	(0.020)		(0.026)	(0.026)
District - month FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Observations	12,929	12,929	12,929	7,092	7,092	7,092
Adjusted R^2	0.638	0.634	0.641	0.674	0.670	0.675

Notes: The dependent variable 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-calendar month and year fixed effects as well as controls for contemporaneous monthly rainfall and temperature. Standard errors are clustered at the district level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.6: Crop Yield and Agricultural Wages (2SLS)

	FS	IV	RF	OLS
	log yield index	log(ag wage)	log(ag wage)	log(ag wage)
	(1)	(2)	(3)	(4)
Rainfall Index	0.021***		0.004***	
	(0.004)		(0.001)	
Temperature Index	-0.054**		-0.002	
	(0.026)		(0.008)	
log YI		0.156**		-0.010
		(0.065)		(0.011)
District - month FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Observations	7,092	7,092	7,092	7,092
Adjusted R^2	0.675	0.803	0.829	0.829

Notes: The dependent variable in column (1) is *log yield index*; while the depend 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 and year fixed effects as well as controls for contemporaneous monthly rainfall and temperature. Standard errors are clustered at the district level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

3. Theoretical Framework

I propose a simple model to understand how fluctuations in agricultural productivity can affect manufacturing firms' performance, in a setting in which workers are allowed be absent and work in agriculture. Since agricultural wages vary seasonally and in response to shocks to agricultural productivity, manufacturing the worker' outside option changes form period to period and so do their incentives to be absent. Manufacturing firms choose labour and wages knowing that, to encourage workers' attendance, they need to provide wages that are high enough to compensate them for the lost opportunity to work in agriculture.

I compare outcomes in two different scenarios: "flexible" and "rigid" labour market. Under the flexible labour market scenario, I assume that firms can adjust labour and wages every period, after observing the realisation of agricultural productivity. Instead, under the rigid labour marker scenario, I assume that firms can only offer contracts that are permanent and pay fixed wages. This implies that firms will maximise expected profits taking into account the fact that, when a positive shock affects the agricultural sector, a larger share of workers will be absent.

The rigid labour market scenario is meant to mimic the context in which firms in the Indian register manufacturing sector operate, described in chapter 2. Whereas, flexible labour market scenario should be interpreted as a benchmark that allows us to evaluate the effect of labour market rigidities on manufacturing sector outcomes such as workers' absences, productivity, labour and wages.

3.1 Workers' Attendance Decision

Assume that each worker consumes only one good c at price 1 and get utility $log(c)$. Workers have, as their only source of income, labour income and supply inelastically one unit of labour. If they work in the manufacturing sector they obtain a wage w_m. However, they have the option to be absent and work in agriculture for a wage w_{ag}. Doing so will give them the additional utility $\tilde{\psi}_i \geq 0$, which varies across workers and is unobservable to the employer. The utility that farmer is obtains from working in agriculture is: $log(w_{ag}) + \tilde{\psi}_i$. For simplicity, I define $\psi_i = e^{\tilde{\psi}_i}$ and assume it follows a Pareto[1,δ] distribution. The role of this parameter is to illustrate how some workers to obtain "extra utility" from working in agriculture as it allows them to return to the home village, visit their family or take care of their own land[1].

Therefore, worker i will decide to be absent and work in agriculture if:

$$log(w_{ag}) + \tilde{\psi}_i \geq log(w_m) \tag{3.1}$$

which is equivalent to:

$$\psi_i w_{ag} \geq w_m \tag{3.2}$$

This implies that, for each realisation of w_{ag}, the attendance rate faced by the firm will be[2]:

$$a(w_{ag}) = 1 - \left(\frac{w_{ag}}{w_m}\right)^{\delta} \tag{3.3}$$

where δ is the elasticity of workers' absence rate with respect to agricultural wages.

3.2 Agricultural Sector

I assume that the agricultural sector is competitive and that production, characterised by constant returns to scale, requires only labour input. The profits of the representative agricultural firm can be written as:

$$\pi_{ag} = \theta L_{ag} - w_{ag} L_{ag} \tag{3.4}$$

[1] I am focusing on an utility benefit rather than a transportation cost as workers that are employed in manufacturing and would have to pay a cost to return to the fields. If this parameter was negative it would be it hard to rationalise that workers often prefer agriculture even if wages, on average, are lower. Therefore, I choose to focus on the fact that, as the temporary migration literature has pointed out (Imbert and Papp, 2016a), utility from being in the village is a major determinant of workers decision to move across sectors.

[2] The CDF of a Pareto[1,δ], for $\psi \geq 1$, is $Pr(\Psi \leq \psi) = 1 - \left(\frac{1}{\psi}\right)^{\delta}$. The attendance rate faced by the firm given agricultural wages w_{ag} can be interpreted as the probability that the representative worker decided not to be absent or $Pr(\Psi \geq \frac{w_{ag}}{w_m})$.

where the parameter θ represents labour productivity in agriculture, which varies seasonally and is affected by weather shocks. Perfect competition and flexible labour market make sure that the agricultural wage in each period is equal to the current labour productivity: $w_{ag} = \theta$. Since the returns to scale are constant, firms are willing to hire any amount of labour for this wage.

3.3 Manufacturing Sector

The manufacturing firm's production function is assumed to be Cobb-Douglas, modified to take into account the role of workers' attendance. The firm's profits can be written in the as follows[3]:

$$\pi_m = A(a^\gamma L_m)^\alpha K^\beta - w_m L_m - rK \tag{3.5}$$

where a represents the workers' attendance rate; L_m and K represent labour and capital, respectively; w_m is manufacturing wage and r is rental rate of capital. I assume that capital is fixed and exogenous. The actual amount of labour used is aL_m, which corresponds to the number of workers employed multiplied by their attendance rate.

The coefficient γ represents the effect of attendance on output. In particular, if $\gamma = 0$ attendance has no effect on output, suggesting that the firm can fully adjust when workers are absent; whereas, if $\gamma = 1$ there is no adjustment and the effect of attendance is equivalent to the effect of a change in the number of workers; if $\gamma < 0$ the productivity of the hours actually worked decreases with attendance rate, as in the case in which the present workers exert extra effort to compensate for the absent ones; finally, if $\gamma > 1$ attendance has a positive effect on productivity because workers' absences cause some disruption in the production process.

The representative firm maximises profits by choosing both wages and labour, taking into account the fact that the workers have the option to work in agriculture. I start solving the model under the "flexible" labour

[3]Although I am assuming that workers earn the manufacturing wage only when present, here I am assuming this is not reflected in the firm's cost function. This is due to the following reasons reasons: (1) intuitively, it seems unrealistic that workers' absences would generate savings for the firm as they may have compensate to workers being absent by hiring substitutes or asking other workers to work overtime; (2) introducing absences in the cost term would complicate the firm's optimisation problem making it impossible to obtain a close form solution.

market assumption and then study how the results would change under the "rigid" labour market scenario. Comparing the two outcomes will allow me to estimate the cost of imposing labour market rigidities in a setting in which workers are allowed to be absent and move across sectors.

I solve the model focusing on the case in which workers' absences are costly for the firm, i.e. $\gamma > 1$ as this is the most interesting to study. The empirical analysis proposed in chapter 4 will test that this is the case.

3.3.1 Flexible Labour Market Outcome

Under the flexible labour market scenario, I assume that firms are allowed to choose wages and labour in each period, after observing the realisation of agricultural productivity and wages. Combing the first order conditions for profit maximisation I obtain optimal manufacturing wages, attendance and employment.

It can be easily shown that firms will set manufacturing wages $w_m^{FLEX}(\theta)$ that increase in agricultural productivity, θ, in order to incentivise workers' attendance by compensating them for their outside option in agriculture. In particular, manufacturing wages will be given by the following equation:

$$w_m^{FLEX}(\theta) = \theta(\gamma\delta + 1)^{\frac{1}{\delta}} \tag{3.6}$$

This result shows that there will be a linear relationship between manufacturing and agricultural wages (that in this model are equal to agricultural productivity, θ) and that manufacturing wages will be higher than agricultural wages, as workers gain utility from working in agriculture. Manufacturing wages will also be increasing in the cost of absences, γ, and in the elasticity of workers' attendance with respect to agricultural wages, δ.

Moreover, it can be shown that the adjustment in manufacturing wages, in response to changes in agricultural productivity, will be such that workers' attendance rate will remain constant at its optimal level:

$$a^{FLEX} = \frac{\gamma\delta}{1 + \gamma\delta} < 1 \tag{3.7}$$

Finally, since manufacturing wages are increasing in agricultural productivity, the optimal amount of labour employed, $L_m^{FLEX}(\theta)$, will be a decreasing function of θ and, as a consequence, so will the amount of output produced, $Y_m^{FLEX}(\theta)$. This implies that, when agricultural productivity increases, manufacturing firms will increase wages to incentivise attendance, but will also dismiss some workers, who will move to the agricultural sector. Therefore, manufacturing output and employment will decrease and agricultural output and employment will increase.

3.3.2 Rigid Labour Market Outcome

Under the assumption that firms cannot adjust wages and labour seasonally and in response to shocks to agricultural productivity, firms maximise expected profits, knowing that workers' attendance will change as agricultural wages fluctuate.

The representative firm's expected profits can be written as:

$$E_\theta(\pi) = AE_\theta\left(a(\theta)^{\gamma\alpha}\right) L_m^\alpha K^\beta - w_m L_m - rK \tag{3.8}$$

where the only component that is affected by θ is attendance[4]. Equations 3.9 and 3.10 report and the first order conditions for profit maximisation with respect to labour and wages.

$$\alpha\Gamma E_\theta(a(\theta)^{\gamma\alpha}) = w_m L_m \tag{3.9}$$

$$\gamma\delta\alpha\Gamma E_\theta\left(a(\theta)^{\gamma\alpha-1}(1-a(\theta))\right) = w_m L_m \tag{3.10}$$

where $\Gamma = AL_m^\alpha K^\beta$.

To obtain an analytical solution I assume that the parameter θ can take two values: θ_H with probability ρ and θ_L with probability $1-\rho$, and that $\theta_H > \theta_L$. An interpretation for this assumption is that agricultural productivity can be high when the rains are good, or low if there is draught. Firms must offer permanent contracts with fixed wages and cannot adjust them after observing weather realisation. An alternative interpretation is to consider seasonal fluctuations in agricultural labour requirements (and wages) and assume that the manufacturing firm is not allowed to adjust wages and employment seasonally.

A direct implication of the fact that wages are fixed is that attendance rates will be contingent on the realisation of θ. In particu-

[4] Combining equation 3.3 with the fact that agricultural wages will be equal to *theta* we know that the expression of attendance as a function of agricultural productivity is: $a(\theta) = 1 - \left(\frac{\theta}{w_m}\right)^{\frac{1}{\delta}}$.

lar, attendance will be lower when agricultural productivity is higher: $a^{FIX}(\theta_L) > a^{FIX}(\theta_H)$. Moreover, since labour is also fixed, changes in attendance will generate changes in manufacturing output, which will be lower when agricultural productivity is higher: $Y^{FIX}(\theta_L) > Y^{FIX}(\theta_H)$.

3.3.3 Comparison Between Rigid and Flexible Labour Market Outcomes

In this section I compare the results obtained under the two different scenarios, to determine the effect of labour market rigidities on manufacturing outcomes such as workers' attendance, wages, employment and output. The proofs of the following propositions are reported in appendix 3.A.

From the previous section we know that, under rigid labour market conditions, manufacturing workers' attendance will depend on the realisation of agricultural productivity. I will now compare the level of attendance in the two states of the world ($a^{FIX}(\theta_L)$ and $a^{FIX}(\theta_H)$) with the optimal level of attendance that would be achieved under flexible labour market conditions, a^{FLEX}.

Proposition 1

Under rigid labour market conditions, manufacturing workers' attendance is higher (lower) than the flexible labour market optimum when agricultural productivity is low (high):

$$a^{FIX}(\theta_L) > a^{FLEX} > a^{FIX}(\theta_H) \tag{3.11}$$

Moreover, average attendance under rigid labour market conditions is lower than under flexible labour market conditions if the elasticity of output with respect to workers' attendance is greater than one, that is if $\gamma\alpha > 1$.

Intuitively, firms operating under rigid labour market conditions do not have the possibility to stabilise the level of attendance across periods. Therefore, they offer wages that, while attracting a large share of workers when agricultural productivity is low, fail to keep attendance at the optimal level when agricultural productivity is high. Moreover, if absences are very costly in terms of productivity, i.e., they reduce output more than proportionately, the optimal level of attendance is high and too costly to achieve in rigid labour markets, even as an average between the high and the low states of the world.

In order to keep the level of attendance constant, manufacturing wages offered by firms operating under flexible labour market condition need to fluctuate mimicking those of the agricultural sector. It is interesting to compare them to the optimal (fixed) wage firms choose to offer under rigid labour market conditions.

Proposition 2

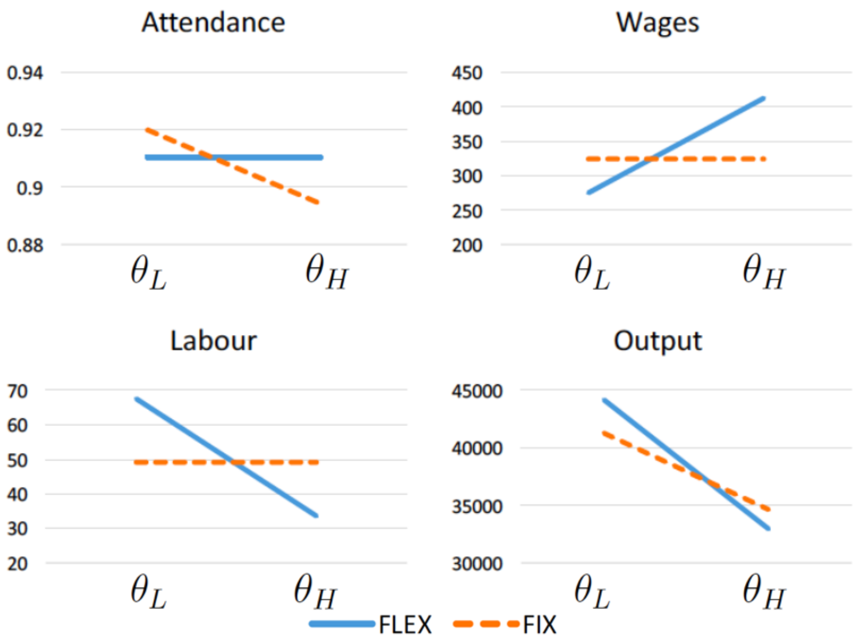
The optimal wage under rigid labour market conditions will be between the wage that firms offer, under labour flexible labour market condition, when agricultural productivity is high and the one they offer when agricultural productivity is low.

$$w_m^{FLEX}(\theta_L) < w_m^{FIX} < w_m^{FLEX}(\theta_H) \tag{3.12}$$

This results suggests that, while firms operating in flexible labour market conditions offer low wages and exploit the labour agricultural surplus during the lean agricultural seasons, firms in rigid labour markets cannot do so. In fact, flexible labour market conditions allow firms to hire workers who seek seasonal employment outside agriculture, e.g. seasonal migrants, which is what we often observe in the informal sector.

In order to determine whether average wages and employment are different across scenarios, it is necessary to make assumptions on the parameters of the model. To do so, I use the parameters estimated in chapter 4 and report the results in figure 3.1, which provides a graphical illustration of the main outcomes of this model. The figure shows that the effect of removing labour market rigidities would produce a large positive effect on employment output when agricultural productivity is low and a decrease in employment when agricultural productivity is high. This, however, is followed by a much smaller difference in the decrease in output, when agricultural productivity is high, between the two scenarios. This is due to the fact that, even if the amount of labour formally employed in rigid labour markets is constants, some workers move across sectors anyway, by being absent. This decrease in attendance has a negative effect on productivity and therefore reduces output.

Figure 3.1: Rigid vs Flexible Outcomes



Notes: the figure illustrate how labour market rigidities affect different manufacturing outcome according to the model.

3.4 Heterogeneous Workers

The model proposed above is based on the assumption that $\tilde{\psi}_i$, the parameter representing the worker's preferences towards agriculture, is unobservable to the firm. However, in some circumstances it is hard to argue that this is the case. For instance, rural migrants are much more likely to derive higher utility from working in agriculture if this means returning temporarily to their village or origin and visiting their families. In this section, I propose an extension to the model in which I allow firms to update their beliefs on workers' preferences for agriculture and offer different wages based on observable characteristics. The purpose of this exercise is to determine whether it would be optimal for firms to offer higher wages to workers who have higher incentive to be absent and work in agriculture and whether such wages should be high enough to ensure that the expected level of attendance is the same for all workers.

I assume workers' "preference for agriculture" can be decomposed in an observable, worker-specific component $\tilde{s}_i \geq 0$ and an unobservable component $\tilde{\psi}_i$. Therefore, workers' utility from working in agriculture becomes: $logw_{ag} + \tilde{s}_i + \tilde{\psi}_i$. To simplify, I define $s_i = e^{\tilde{s}_i}$ and $\psi_i = e^{\tilde{\psi}_i}$. As before, I assume that ψ_i is drawn from a random variable Ψ following Pareto[1,δ] distributions. Since s_i is observable, I allow firms to offer different wages to different workers depending on this parameter. Workers will be absent from the manufacturing firm and work in agriculture if the following inequality holds:

$$w_{im} \leq s_i \psi w_{ag} \tag{3.13}$$

where w_{im} is worker i's manufacturing wage. The expected attendance rate of worker i can be written as:

$$a_i(w_{ag}) = 1 - \left(s_i \frac{w_{ag}}{w_{im}}\right)^{\delta} \tag{3.14}$$

To incorporate workers' heterogeneity into the firm's optimisation problem, I rewrite the firm's profit function as follows:

$$\pi_m = A(\sum_i a_i^{\gamma} L_{im})^{\alpha} K^{\beta} - \sum_i w_{im} L_{im} - rK \tag{3.15}$$

This production function assumes complementaries between the different types of workers. This is necessary to obtain interior solutions. In fact, if workers were assumed to be perfect substitutes it would be optimal for the firm to select workers with the lowest level of observable preference for agriculture and supply side considerations would need to be made in order to explain workers' heterogeneity within a firm or even within the whole manufacturing sector.

As before, I first solve the problem under the "flexible" and "rigid" labour market scenarios separately and then compare the outcomes.

Flexible Labour Market Outcome

It is possible to show that, also in this case, firms operating under flexible labour market condition will set wages in order to achieve the optimal level of attendance: $a^{FLEX} = \frac{\gamma\delta}{1+\gamma\delta}$. However, now this implies offering wages that depend, not only not the current agricultural wage, but also on the worker's observable characteristics, represented by the parameter s_i as indicated by the following equation:

$$w_{im}^{FLEX}(\theta) = s_i\theta(\delta\gamma + 1)^{\frac{1}{\delta}} \tag{3.16}$$

In particular, wages will be increasing in s_i to incentivise the attendance of workers with higher preference for agriculture. On the other hand, it is possible to show that this heterogeneity in wages will be reflected by the heterogeneity in the amount of labour, in fact, workers with higher s_i are more expensive for the firm and fewer of them will be hired.

Rigid Labour Market Outcome

In the rigid labour market case firms are still forced offer permanent contract with constant wages that cannot be adjusted in response to changes in agricultural productivity but they are now allowed to offer different contracts to workers based on the observable parameter s_i. Firms maximise expected profits represented by the following equation:

$$E_\theta(\pi_m(\theta)) = AE_\theta\left((\sum_i a_i(\theta)^\gamma L_{im})^\alpha\right)K^\beta - \sum_i w_{im}L_{im} - rK \tag{3.17}$$

To simplify the calculation is I assume that that the probability of having a good shock to agricultural productivity is $\rho = \frac{1}{2}$ and study how wages and attendance will vary with respect to s_i. The proof of the following proposition is reported in appendix 3.A.

Proposition 3

Firms operating in rigid labour market conditions will offer higher wages to workers with higher preference for agriculture. $\frac{\partial w_{im}}{\partial s_i} > 0$ *However, the increase in wage will not be enough to ensure equal levels of attendance across workers. In particular, workers with higher preference for agriculture will display lower levels of attendance* $\frac{\partial a_i^{FIX}(\theta)}{\partial s_i} < 0$.

An implication of this result is that attendance rate for workers with higher preference for agriculture will be lower regardless the realisation of agricultural productivity, but the effect is amplified when agricultural productivity is higher.

The fact that workers with higher preference for agriculture earn higher wage and are absent more often (and therefore are less productive) both imply that the firm will decide to hire fewer of them.

Comparison Between Rigid and Flexible Labour Market Outcomes

This extension shows that the flexible labour market assumption, not only fails to explain the seasonal patten in workers' attendance we observe in the data, but it also fails to explain why workers' with different observable characteristics, such as the being migrants from rural areas, could display different levels of attendance.

Another interesting outcome is that, while in flexible labour markets workers with higher preferences for agriculture are paid higher wages that fully stabilise attendance, under rigid labour market conditions, even if paid higher wages, these workers are still more likely to engage in seasonal absences. This behaviour has negative consequences on productivity which, importantly, creates an additional disincentive for firms to hire them.

In conclusion, this model suggests that labour market rigidities have the additional effect of reducing movement of workers across sectors, therefore slowing down the structural transformation of the economy.

Appendix

3.A Proofs

Proof of proposition 1:

Under rigid labour market conditions, manufacturing workers' attendance is higher (lower) than the flexible labour market optimum when agricultural productivity is low (high):$a^{FIX}(\theta_L) > a^{FLEX} > a^{FIX}(\theta_H)$ *Moreover, average attendance under rigid labour market conditions is lower than under flexible labour market conditions if the elasticity of output with respect to workers' attendance is greater than one, that is if* $\gamma\alpha > 1$.

Let us define X as the ratio of attendance in rigid labour market condition under low agricultural productivity and high agricultural productivity:

$$X = \frac{a^{FIX}(\theta_L)}{a^{FIX}(\theta_H)} \tag{3.18}$$

From equation 3.3 we know that X can be rewritten in as: $X = \frac{w^\delta - \theta_L^\delta}{w^\delta - \theta_H^\delta}$ and $X > 1$ since $\theta_H > \theta_L$. Combining equation 3.9 and 3.10 and replacing the general expected value with the one obtained by assuming that θ can take only two values we obtain the following expressions:

$$a^{FLEX}\frac{\rho + (1-\rho)X^{\gamma\alpha - 1}}{\rho + (1-\rho)X^{\gamma\alpha}} = a^{FIX}(\theta_H) \tag{3.19}$$

$$a^{FLEX}\frac{\rho X + (1-\rho)X^{\gamma\alpha}}{\rho + (1-\rho)X^{\gamma\alpha}} = a^{FIX}(\theta_L) \tag{3.20}$$

where a^{FLEX} is workers' attendance rate under flexible labour market conditions, reported in equation 3.7. It is possible to notice that $X > 1$ implies $a^{FIX}(\theta_L) > a^{FLEX} > a^{FIX}(\theta_H)$.

Using equations 3.19 and 3.20 we can write average attendance rate

($\overline{a^{FIX}} = \rho a^{FIX}(\theta_H) + (1-\rho)a^{FIX}(\theta_L)$) in rigid labour market conditions as a function of a^{FLEX}:

$$\overline{a^{FIX}} = a^{FLEX}\left(\frac{\rho^2 + \rho(1-\rho)X^{\gamma\alpha-1} + (1-\rho)\rho X + (1-\rho)^2 X^{\gamma\alpha}}{\rho + (1-\rho)X^{\gamma\alpha}}\right) \tag{3.21}$$

It is possible to notice that the condition under which $\overline{a^{FIX}} < a^{FLEX}$ is that $X > 1$, which has been discussed above, and that $\gamma\alpha > 1$.

Proof of proposition 2:

The optimal wage under rigid labour market conditions will be between the wage that firms offer, under labour flexible labour market condition, when agricultural productivity is high and the one they offer when agricultural productivity is low. $w_m^{FLEX}(\theta_L) < w_m^{FIX} < w_m^{FLEX}(\theta_H)$

Replacing the $a^{FLEX} = \frac{\gamma\delta}{1+\gamma\delta}$ and $a^{FIX}(\theta_H) = 1 - \left(\frac{\theta_H}{w}\right)$, equation 3.19 can be rewritten as:

$$w_m^{FIX} = w^{FLEX}(\theta_H)\left(\frac{\rho + (1-\rho)X^{\gamma\alpha}}{(\rho + (1-\rho)X^{\gamma\alpha} + (1-\rho)\gamma\delta X^{\gamma\alpha}\left(\frac{X-1}{X}\right)}\right)^{\frac{1}{\delta}} \tag{3.22}$$

Similarly, equation 3.20 can be rewritten as:

$$w_m^{FIX} = w^{FLEX}(\theta_L)\left(\frac{\rho + (1-\rho)X^{\gamma\alpha}}{(1+\delta\gamma)\rho + (1-\rho)X^{\gamma\alpha} - \rho\gamma\delta X}\right)^{\frac{1}{\delta}} \tag{3.23}$$

It is easy to verify that $X > 1$ implies $w_m^{FLEX}(\theta_L) < w_m^{FIX} < w_m^{FLEX}(\theta_H)$.

Proof of proposition 3

Firms operating in rigid labour market conditions will offer higher wages to workers with higher preference for agriculture. $\frac{\partial w_{im}}{\partial s_i} > 0$ *However, the increase in wage will not be enough to ensure equal levels of attendance across workers. In particular, workers with higher preference for agriculture will display lower levels of attendance* $\frac{\partial a_i^{FIX}(\theta)}{\partial s_i} < 0$.

From the first oder conditions I obtain an implicit function for the optimal wages offered to worker i and I differentiate it with respect to s_i. It is possible to show that manufacturing wages will be increasing in s_i as in the flexible labour market case. In particular:

$$\frac{\partial w_{im}}{\partial s_i} = \frac{w_{im}}{2s_i} > 0 \tag{3.24}$$

The attendance rate of worker i, when agricultural productivity is equal to θ can be written as:

$$a_i^{FIX}(\theta) = 1 - \left(\frac{s_i \theta}{w_{im}}\right)^{\delta} \tag{3.25}$$

Using the result from equation 3.24 it follows that

$$\frac{\partial a_i^{FIX}(\theta)}{\partial s_i} = -\delta \left(\frac{s_i \theta}{w_{im}}\right)^{\delta} \left(\frac{1}{2s_i}\right) < 0 \tag{3.26}$$

4. Seasonal Absences and Industrial Performance

4.1 Overview

The purpose of this chapter is to provide quantitative evidence of the "seasonal absences" of workers employed in the Indian registered manufacturing sector and to estimate the effect on firms' productivity.

I define "seasonal absences" as those absences that take place during the peak agricultural seasons, when agricultural wages increase and workers have access to better outside options in the agricultural sector. Since manufacturing workers are typically employed on a permanent basis, with wages that do not adjust seasonally, they have the incentive to leave the firm temporarily and work in the agricultural sector, if the probability of being fired is low enough.

Section 2.1.1 provides a description of the legal framework regulating employment relations in this context. Specifically, it describes the main features of the Industrial Disputes Act (IDA) of 1947, the main piece of legislation covering resolution of industrial disputes and regulating hiring and firing of workers, closure of establishments, strikes and lockouts etc. Importantly, the IDA postulates that establishments employing more than 100 workers need to obtain government permission for dismissing even a sigle worker. This is true even in case of prolonged unjustified absences and in fact, workers who get dismissed for this reason are usually reinstated if they appeal to a labour court (Kumar, 2010). The intensity of labour protection provisions varies across Indian states as the IDA received numerous state amendments either in a "pro-worker" or in a "pro-employer" direction (Besley and Burgess, 2004). Following Adhvaryu et al. (2013), in chapter 2 I classify as "pro-worker" the states of Maharastra, Orissa and West Bengal[1] and I show that absence rates are on average 1.5 percentage point higher in these states, suggesting that workers are more likely to be absent when the threat to be fired is lower.

The analysis proposed in this chapter, is based on firm level data representative of the entire Indian registered manufacturing sector, collected through 8 rounds of the Annual Survey of Industries (ASI), covering the period from 2000-01 to 2007-08. They contain a rich amount of information on firms characteristics and productive activity, including input, output and capital stock. Moreover, they provide a unique source of data on workers' attendance at a monthly frequency, which, to my knowledge, have never been analysed before by researchers in economics.

I exploit the seasonal features of the agricultural labour market, described in section 2.2 to quantitatively document the phenomenon of seasonal absences among workers employed in the Indian registered manufacturing sector. Exploiting differences in agricultural harvest calendar across Indian districts, I find that workers' attendance is lower during the harvest months, which correspond to the peak season in agriculture. I then use exogenous shocks to agricultural productivity (i.e. weather shocks), that affect labour demand during the harvest season, to estimate the response of manufacturing workers' attendance to changes in their outside option. This approach allows me to compare workers' attendance, within district, across harvest seasons. I find that a 1% increase in yearly agricultural wages, caused by positive weather shock, causes a 0.14% decrease in yearly industrial workers' attendance, a result that is both statistically and economically significant.

Moreover, I estimate the effect of this phenomenon in terms of output loss, finding large results: 1% decrease in attendance rate, caused by a positive weather shock, reduces manufacturing output by over 6%. The interpretation of these findings is that seasonal workers' absences cause large disruptions in the firm production process, possibly because many workers leave simultaneously and are difficult to replace, which reduces firms' output much more than a change in the labour force of equal size.

[1] In 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 of Gujarat was criticised by the literature (Bhattacharjea, 2006)(ii) during the period considered some "pro-employer" reforms where implemented in the state.

Finally, I use the parameters estimates in this chapter to calibrate the model proposed in chapter 3 and compare the current outcome to a flexible labour market benchmark. The results suggest that removing labour market rigidities, which prevent firms from adjusting labour and wages in response to shocks to agricultural productivity, would increase average manufacturing output by 5.13% and average manufacturing employment by 3.36%, without decreasing average wages.

4.2 Data

4.2.1 Firm Level Data

This chapter is based on firm level data representative of the entire Indian registered manufacturing sector, collected through the Annual Survey of Industries (ASI)[2]. They cover all registered plants employing 10 or more workers using power and all plants employing 10 or more workers if they do not use power, operating in the manufacturing sectors. Firms employing more than 100 workers are surveyed every year, while 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[3] provides monthly information on workers' attendance 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[4]. These include 107,539 firm-year observations.

All variables expressed in nominal terms were deflated using the all India monthly CPI for Industrial workers[5].

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*

[2]See MOSPI (2014) for a complete description of these data.

[3]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.

[4]Part I data were dowloaded from the LSE India Data Centre website http://idc.lse.ac.uk/ and Part II data were obtained from the Indian Labour Bureau.

[5]The CPI data was obtained from the Indian Labour Bureau website http://labourbureau.nic.in/indtab.html.

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). These are *total output* defined as the 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; *total input* defined as the 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; *gross value added* defined as the difference between *total output* and *total input*; *profits* defined as *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[6]. 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

[6] 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, Nagalanda and Tripura), about 9% of the initial sample.

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[7]. Moreover, 2,623 firms report negative or zero output and are excluded from the analysis[8].

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

4.2.2 Agricultural Data and Weather Shocks

The firm level data presented above were matched with district level information about monthly characteristics of the agricultural labour market and weather conditions described in section 2.2 . The relevant variables are: agricultural wages, obtained from the "Agricultural Wages in India" (AWI) collected by the Directorate of Economics and Statistics of the Indian Ministry of Agriculture; harvesting timing and yield information for the major crops cultivated in the districts, obtained by combining from crop calendar information from the 1967 Indian Crop Calendar published by the Directorate of Economics and Statistics with data on crop output and area sown collected from the Crop Production Statistics Information System website; growing season rainfall and temperature computed using weather data collected by the Center of Climatic Research at the University of Delaware.

[7]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.

[8]Negative observed output is possible because it includes the change in stock of semi-finished products, which may be negative.

4.3 Summary Statistics

Firm production data

Table 4.1 panel C reports summary statistics for the main firm-level variables used in the analysis: capital, number of workers and value added. Both capital and number of workers are measured at the beginning of the fiscal year. Value added is computed as the difference between total output and the cost of materials, fuel, services and operating expenses. Since the variables are expressed in logs, they exclude all the instances in which capital or value added are zero or negative.

Absence data

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.

Absence rate is calculated the share of the man-days scheduled in a month that are lost due to workers' absence. Figure 4.1 shows the distribution of firm level monthly absence rates, the vertical line indicates the median of the distribution, which corresponds an absence rate of 7.66%. The figure shows that the distribution is highly skewed with a large number of firms reporting very high absence rates. The average absence rate is 9.7% as reported in table 4.1.

4.4 Evidence of Seasonal Absences

The model in chapter 3 predicts that, if firms are not allowed adjust wages and labour input in the short run, workers' attendance in the manufacturing would fluctuate across seasons. In particular, we expect workers' attendance to be lower when labour demand in agriculture is highest, i.e. during the harvest season and when agriculture is more productive.

Figure 4.2 plots monthly absence rates against the percentage of the district agricultural area harvested in the calendar month. The solid line represents absence rates when rainfall during the crops' growing season is high, that is over the 80th percentile of its long term (50 year) distribution; while the dashed line represents absence rates when rainfall during the crops' growing season is low, that is below the 20th percentile of its long term distribution. The figure shows that absence are increasing in the share of area harvested and more so when the amount of rainfall is high.

In order to test the hypothesis that attendance is lower during the harvest season formally, I exploit the fact that Indian districts specialise in different crops and that crop calendar varies across districts. This allows me to disentangle the effect of having a harvest in the district from other seasonal events that may affect attendance, such as national festivals or school holidays. The empirical model I estimate is the following:

$$log(attendance_{idmt}) = \alpha harvest_{dm} + \beta x_{idmt} + \delta_d + v_t + e_m + u_{idmt} \quad (4.1)$$

where $log(attendance_{idt})$ is the natural logarithm of the attendance rate in firm i, located in district d, in month m and year t; $harvest_{dm}$ 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's and district's characteristics; δ_d, e_m and v_t represent district, calendar 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.

$$log(attendance_{idmt}) = \alpha log(agr\_wage_{dmt}) + \beta x_{idmt} + \delta_d + v_t + u_{idmt} \quad (4.2)$$

where $log(agr\_wage_{dt})$ 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 2.4. Column (1) shows that attendance rates are significantly lower during the harvest months; 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 and that the effect is stronger for the workers that are more protected.

To address the concern that this may be driven by other events that happen during the harvest months, such as harvest festivals or weddings and to test whether workers' absences are caused by the availability of economic opportunities, I look at the response of workers' attendance to changes in the local agricultural wage.

According to the Payment of Wages Act of 1936, manufacturing workers may have to forgo the entire manufacturing wage if they are absent from the workplace. Since manufacturing wages are on average higher then agricultural wages, it may seem implausible that workers would choose to leave the manufacturing sector temporarily and work in agriculture. However, if we look at the distribution of the difference between the firms' average wage and the local district's agricultural wages, plotted in figure 4.3, we see that that, although the difference is in general positive, i.e. manufacturing wages are higher on average, often the difference is small and there are some circumstances in which agricultural wages are higher than manufacturing wages. Moreover, the overlap between the entire wage distribution is likely to be larger. Finally, as discussed in chapter 3, worker may obtain additional utility from working in agriculture, especially if this gives them the opportunity to visit their family or take care of their own land.

The estimated response of workers' attendance to changes in the local agricultural wage are reported in column (3) and (4) of table 2.4: as expected, the effect of an increase in agricultural wage on attendance is negative and statistically significant, suggesting that workers are absent more often when they have a better outside option in the agricultural sector. Also in this case, the effect is higher in pro-worker states, although the difference is not statistically significant[9]. The results are robust to the introduction of district-month fixed effects, highlighting the importance of changes in agricultural labour demand that occur from one year to another, within the same calendar month.

The estimate relationship between attendance and agricultural wages described above, however, is likely to biased due to endogenity problems. For instance, agricultural wages are likely to be endogenous to other factors affecting also the manufacturing labour market such as economic opportunities in other sectors. Moreover, data gaps and measurement error are likely to lead to make the estimate imprecise and biased towards zero.

As a solution to these problems I propose a two-sample two-stage estimation approach, described in the following section. 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, which allows also to improve reduce the number of data gaps. Indeed, in many cases agricultural wages are missing during the lean agricultural month, when few casual jobs are available. By averaging the available I can improve the quality of the data without losing the relevant variation for this exercise, which is that due to changes in agricultural productivity that occur from one year to another.

4.5 Elasticity of Workers' Attendance with respect to Agricultural Wages

The purpose of this section is to causally estimate the elasticity of workers' attendance with respect to agricultural wages. To do so, I propose a two-sample two-stage estimation approach using weather shocks as source

[9]The lack of statistically significance can be explained by the fact that agricultural wages presents many data gaps for the largest to the "pro-worker" states, Maharashtra, which reduces statistical power. Indeed, when replicating the estimate in column (2) of table 2.4 on the sub-sample for which agricultural wages are available the coefficient of the interaction Harvest*pro-worker is also not statistically significant.

of exogenous variation.

The analysis reported in section 2.2 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 though 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 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:

$$log(agr\_wage_{dt}) = \gamma weather\_index_{td} + \delta_d + \upsilon_t + e_{dt} \tag{4.3}$$

where $weather\_index_{td}$ 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:

$$log(attendance_{idt}) = \alpha log(\widehat{agr\_wage}_{dt}) + \beta x_{idt} + \delta_d + \upsilon_t + u_{idt} \tag{4.4}$$

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

The results are reported in table 4.3. Column (1) reports the results from the first stage regression, which is similar to that obtained using monthly level data in section 2.2, indicating that an increase in rainfall, which increases agricultural productivity, has a positive effect on agricultural wages, while an increase in temperature, which decreases agricultural productivity, has a negative effect on agricultural wages. Column (2) reports the second stage estimate for the elasticity of workers' attendance with respect to agricultural wages: the coefficient suggests that a 1% increase in the agricultural wage reduces yearly attendance by 0.142%. Finally, 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%.

These result show that there is a causal relationship between shocks to agricultural productivity, that affect manufacturing workers' outside option represents by potential income from working in agriculture, and their attendance. The fact that this relationship is stronger in the states where workers are more protected highlights the relevance of labour market rigidities in generating this phenomenon, consistently with the theory proposed in chapter 3.

4.6 Effect of Seasonal Absences on Productivity

The previous sections provide evidence the fact that workers' attendance in the Indian registered manufacturing sector is affected by shocks to agricultural productivity. In particular, I have showed that workers are more likely to be absent during the harvest seasons, that is when labour demand is highest, and in the years in which agricultural productivity and wages are higher. In this section I will estimate the effect of these "seasonal absences" on firms' productivity.

Although absences are usually considered a negative outcome for the firm, as they create an extra cost in terms of reassignment of tasks to workers and substitutes (Allen, 1983). However, their effect on productivity, measured in terms of output per hour effectively worked, is not necessarily negative. Indeed, the firm may be able to compensate for the loss in working time caused by absence by hiring substitutes of having the present workers work overtime. It is also plausible that when some workers are absent, the productivity of those present may increase if they work harder to compensate. On the other hand, if the firm is unable to find equally productive substitutes, absence may cause a loss in productivity. Moreover, absences may create disruptions, reduce workers' morale and take up management time that could otherwise be devoted to more productive activities.

The objective of this section is to estimate the effect of seasonal absences on firms' productivity. This is a particularly relevant question in developing countries, where manufacturing workers are likely to have an agricultural background and easy access to job opportunities in agriculture when labour demand in the sector is high. These absences may have a stronger effect on productivity, compared to regular absences, for several reasons: (i) having many workers absent at the same time is likely to generate to cause larger disruptions; (ii) finding substitutes in these periods may be more difficult; (iii) more productive workers may have access to better outside opportunities and are more likely to be absent in these periods.

To illustrate the problem formally, I rely on the model proposed in chapter 3, where I assume firms in the manufacturing sector are characterised by a Cobb-Douglas production function, modified to allow attendance to have an independent effect on productivity, as in the following equation:

$$Y = A(a^{\gamma}L)^{\alpha}K^{\beta} \tag{4.5}$$

where a = attendance rate, L is number of worker-days scheduled and K is capital. The actual number of worker-days worked will be aL: the number of worker-days scheduled multiplied by the attendance rate. The coefficient γ determines whether workers' absences have an effect on firms' output and productivity. In particular, if $\gamma = 0$ attendance rate has no effect on output, suggesting that the firm can fully adjust; whereas, if $\gamma = 1$ there is no adjustment and the effect of a change in attendance 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 $\gamma > 1$ absences have a negative effect on productivity of the days effectively worked, suggesting they create disruption in the production process. The remainder of this section will be devoted to the estimation of this parameter.

4.6.1 Empirical Strategy

The empirical model corresponding to equation 4.5 is the following:

$$log(Y_{idt}) = \alpha\gamma log(a_{idt}) + \alpha log(L_{idt}) + \beta log(K_{idt}) + e_{idt} \tag{4.6}$$

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; a_{idt} is attendance rate in firm i in year t and district d.

The coefficient of $log(a_{idt})$ combines the effect of a change in the amount of days actually worked and the effect a change in attendance per se. To obtain an estimate for the coefficient of interest γ I divide the attendance coefficient $\alpha\gamma$ by the labour coefficient α.

Since the purpose of the analysis is to estimate the effect of seasonal absences, rather than the effect of absences in general, I estimate equation 4.6 using a two-stages least square technique using exogenous shocks to agricultural productivity, provided by growing season rainfall and temperature, as instrument for attendance. The estimated coefficient will represent the local average treatment effect (LATE) of the change in attendance caused by changes in shocks to agricultural productivity.

Moreover, exploiting only the variation in absence that is caused by these exogenous shocks allows me to take care of the endogeneity issues related to this estimation: since attendance rates are likely to be lower in less productive firms or in firms facing a negative demand shock a simple OLS regression would lead to an overestimate of their effect. On the other hand, yearly absence rates a_{idt} is likely to be measured with error as it is obtained by aggregating the four monthly observations available in the dataset. This measurement error would probably bias the estimate of $\alpha\gamma$ towards zero.

The first stage regression in this case is the reduced form estimate for the analysis proposed in section 4.5:

$$log(a_{idt}) = \delta_1 rain_{td} + \delta_2 temperature_{td} + u_{itd} \tag{4.7}$$

where: $rain_{td}$ and $temperature_{td}$ are measures of growing season rainfall and temperature in year t and district d.

In order to obtain a consistent estimate of the other parameters in 4.6: α and β, and to address the concern that labour and capital may be determined after the shock to agricultural productivity takes place, I use number of workers at the beginning of the year as a proxy for L_{idt} and level of capital at the beginning of the year as a proxy for K_{idt}. As the raw materials and other inputs used for production is likely to depend on workers' attendance and its inclusion in the regression would bias the other coefficients, I measure Y_{idt} in terms of value-added.

4.6.2 Results

The results are reported in table 4.4 . Column (1) reports the OLS estimate of equation 4.6: the attendance coefficient suggest that 1% increase in attendance increases value added by 0.887%. The OLS estimate for γ is 2.2 and is statically greater than 1. The specification includes various firm level and district level controls as well as district and two-digit industry fixed effects. The standard errors are clustered at the district level.

Column (2) reports the first stage regression. It indicates that 100 mm increase in growing season rainfall decrease yearly attendance by 0.1% and 1°C increase in average growing season temperature increases yearly attendance by 0.3%. These estimates are obtained controlling for overall yearly rainfall and temperature to address the concern that there may be a direct effect of weather on attendance.

Column (3) reports the 2SLS estimate, the coefficient of attendance suggests that 1% increase in attendance (caused by a decrease in seasonal absences) increases value added by 6.21% a much larger effect of what obtained with the OLS estimation. These results suggest that seasonal absences have a large negative impact on firm productivity, indeed the corresponding estimate of γ is 14.73 and is statistically greater than 1 (p-value=0.002).

Column (4) reports the reduced form regression, showing that growing season temperature has a positive effect of value added, whereas the effect of growing season rainfall is negative (although not statistically different from zero).

4.6.3 Robustness

To confirm my estimate of the effect of seasonal absences on firms' productivity, I perform two robustness checks that I report in table 4.5.

First, I consider the possibility that shocks to agricultural productivity may affect firm's output, not only though workers' absences, but also by changing the number of contractors employed by the firm. In fact, like regular workers, contractors are also likely to move across sectors as pointed out by Colmer (2016) and Chaurey (2015). To address this concern I control for the number of days worked by contractor in a year. As this variable is affected by shocks to agricultural productivity it is a "bad control", which may invalidate my estimate. However, if its inclusion does not affect my coefficients of interest, I can conclude that changes in the contract labour use do not drive the effect that I find. Indeed, by comparing column (1), which reports my main estimate, and column (2) in table 4.5, it is possible to notice that the relevant coefficients are virtually unchanged.

Finally, I extend the analysis to firms who report zero capital stock, which were dropped from my main production function estimate. Given the magnitude of this subsample (over 32,000 firm-year observations) I am concerned that my estimate may not be representative of the entire sector. Therefore, I repeat the analysis by substituting the variable $log(K)$ with $log(K+1)$ and including these firms in the sample. The results for this exercise, reported in column (3), show that, although this leads to estimating an implausibly low capital coefficient, counterbalanced by an implausibly large labour coefficient, the estimate of the parameter γ does not change significantly and it is still possible to reject the hypothesis that is equal to 1.

4.7 Calibrating the Effect of Labour market Rigidities

In this section, I calibrate the model proposed in chapter 3 in order to estimate the cost of labour market rigidities in a context in which the workers' outside option fluctuates seasonally and in response to shocks to agricultural productivity. In particular, I compare the situation in which firms are allowed to adjust wages and employment in each period, after observing the realisation of agricultural productivity, to the situation in which firms are forced to offer permanent contracts with fixed wages.

Table 4.6 reports the parameters used for the model calibration. High and low agricultural wages are obtained by adding or subtracting one within district standard deviation from the average agricultural wage. The parameters α and γ are obtained from the production function estimation reported in section 4.6. The parameter δ, representing the elasticity of absence rate with respect to agricultural wages, is estimated using the same procedure as that used to compute the elasticity of attendance with respect to agricultural wages in section 4.5. The level of attendance in the rigid labour market scenario when agricultural wages are low, $a^{FIX}(\theta_L)$ is set to be equal to 0.91, which is the mean attendance rate in the months of no harvest. Finally the employment level in rigid labour markets L_m^{FIX} is set to be equal to 49.18, which is the median firm size in the sample.

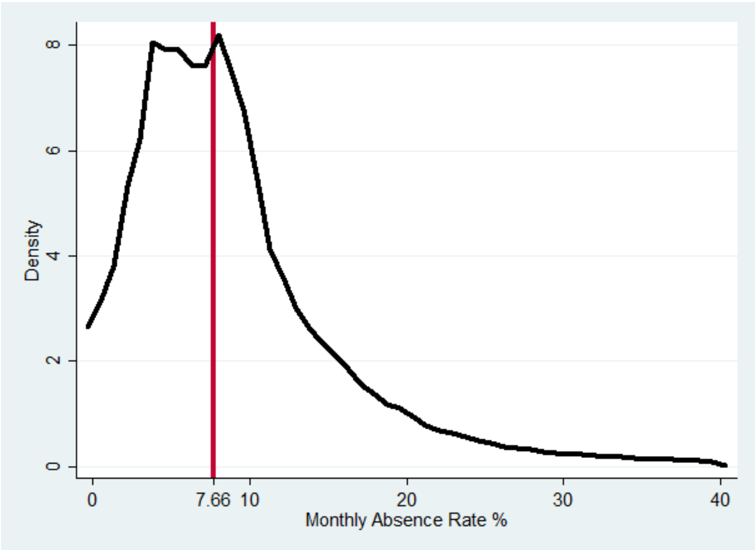
The results, reported in table 4.7, show that labour marker rigidities, coupled with seasonality in workers' outside option, cause a reduction in manufacturing wages, employment and output. Indeed, as absence increase, marginal productivity of labour decreases, and firm end up hiring fewer workers even if wages are lower. Therefore, not only firms lose in terms of output and profits, but also workers are penalised by having access to fewer jobs at lower wages.

According to my estimates, removing these rigidities would increase average manufacturing output by 5.13% and average manufacturing employment by 3.36%. Figure 4.4 plots the effect on output, labour and manufacturing wages as a function of the ratio between high and low agricultural wages. As this ratio increase, the gains from flexibility also increase.

Moreover, in this case more workers will be employed in manufacturing when the agricultural sector is less productive, resulting in a better allocation of resources and greater efficiency.

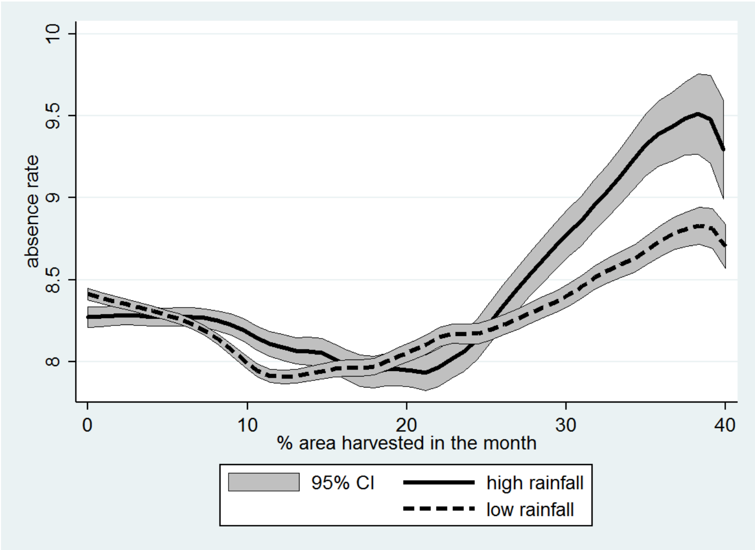
Figures

Figure 4.1: Distribution of Monthly Absence Rates



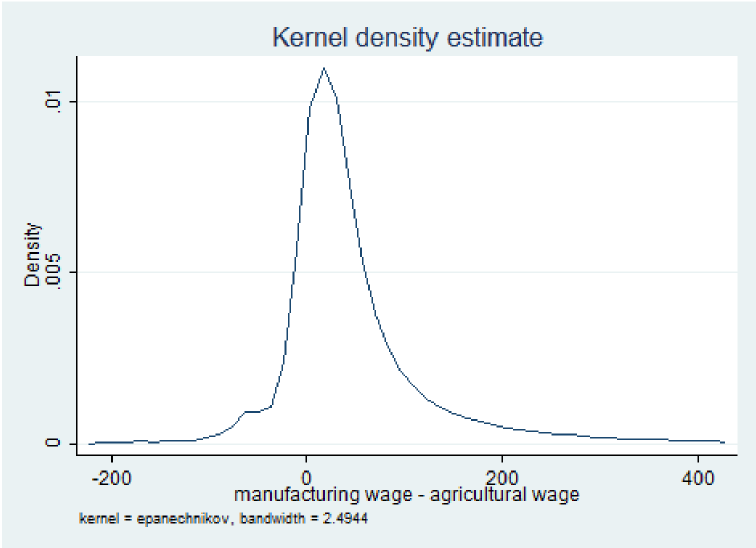
Notes: The figure plots the distribution of firms' monthly absence rates. The vertical line indicates the median absence rate in the sample, corresponding to 7.66%. Absence rate is defined as the share of the man-days scheduled in a month that are lost due to workers' absence. Absence data is reported for the months of March, June, September and December.

Figure 4.2: Seasonal Absences



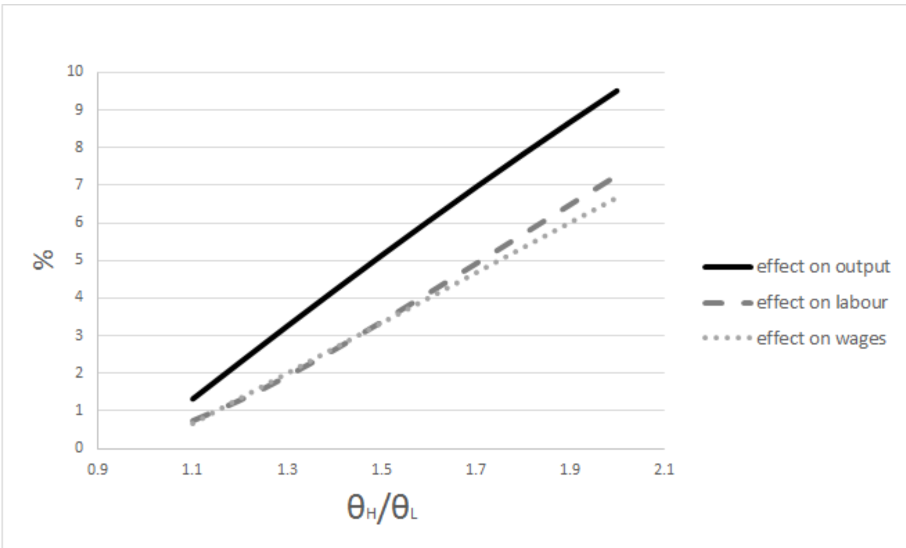
Notes: the graph represents the local polynomial regression of monthly absence rates on the percentage of the district's agricultural area that is harvested in the month. The measures of rainfall and temperature used are based on the most important crop, in terms of area sown, among those harvested in the district in the month. In particular, high rainfall is defined as the cumulative rainfall over the crop's growing season being above the 80th percentile of its long term (50 years) distribution. Similarly, low rainfall is defined as the cumulative rainfall over the crop's growing season being below the 20th percentile of its long term distribution.

Figure 4.3: Agriculturas vs Manufacturing Wages



Notes: the graph represent the distribution of the difference between the firms' average wage and the local district's agricultural wages. Although the difference is in general positive, i.e. manufacturing wages are higher on average, often the difference is small and there are some circumstances in which agricultural wages are higher than manufacturing wages.

Figure 4.4: Gains from Flexible Labour Market



Notes: The graphs plots the gains from moving from a rigid labour market to a flexible labour market, when the probability of having a positive shock to agricultural productivity is 20%, as a function of the ratio between the high and the low agricultural wage. The solid line represents the effect on output; the dashed line represents the effect on labour; the dotted line represents the effect on manufacturing wages.

Tables

Table 4.1: Summary Statistics

Variable	Mean	Standard Deviation	Observations	Source
Panel A - Firm level data - Monthly frequency				
Absence rate	0.097	0.079	404,222	ASI II
Panel B- Firm level data - Yearly frequency				
Log capital	17.48	2.063	59,027	ASI I
Log number of workers	3.302	1.482	107,539	ASI II
Log value added	15.56	2.144	104,215	ASI I

Notes: ASI I and ASI II stand for Annual Survey of Industries-Part I and II, respectively. Absence rate is computed as number of man-days lost due to absence divided by number of man-days scheduled.

Table 4.2: Workers' Attendance - Monthly Level

	(1) log(att)	(2) log(att)	(3) log(att)	(4) log(att)
Harvest	-0.002***	-0.001**		
	(0.001)	(0.001)		
Harvest*pro worker		-0.002*		
		(0.001)		
log(ag wage)			-0.010**	-0.009*
			(0.005)	(0.005)
log(ag wage) *pro worker				-0.004
				(0.007)
Observations	404222	404222	139815	139815
Adjusted R^2	0.127	0.127	0.152	0.152

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 4.3: Workers' Attendance - Yearly Level

	(1) Log(agr wage)	(2) Log(attendance)	(3) Log(attendance)
Rainfall Index	0.009**		
	(0.004)		
Temperature Index	-0.002***		
	(0.001)		
$\widehat{Log(agwage)}$		-0.142**	-0.111
		(0.067)	(0.073)
$\widehat{Log(agwage)}$* pro worker			-0.101*
			(0.060)
Observations	62601	107539	107539
Adjusted R^2	0.84	0.165	0.166

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; $\widehat{Log(agwage)}$ 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(ag wage) is estimated, following Murphy and Topel (1985). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.4: Effect of Seasonal Absences on Productivity

	(1) OLS Log(value add)	(2) FS Log(att)	(3) 2SLS Log(value add)	(4) RF Log(value add)
Log (attendance)	0.887***		6.212***	
$\alpha * \gamma$	(0.101)		(2.011)	
Log L	0.403***	-0.004***	0.422***	0.398***
α	(0.015)	(0.001)	(0.017)	(0.015)
Log K	0.580***	0.002***	0.568***	0.581***
	(0.011)	(0.000)	(0.012)	(0.011)
Rainfall Index (100 mm)		-0.001**		-0.001
		(0.000)		(0.004)
Temperature Index (°C)		0.003***		0.018**
		(0.001)		(0.005)
γ	2.201***		14.73***	
	(0.264)		(4.505)	
P-value $\gamma = 1$	0.000		0.002	
Observations	56786	56786	56786	56786
Adjusted R^2	0.743	0.743	0.711	0.742

Notes:Column (1) reports the OLS estimate of the effect of attendance on output and productivity; Column (2) reports the fist 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. Firm level controls include: rural location, ownership type, organisation type, two digit industry. All specifications are based on yearly firm level data and include district and year fixed effects. Standard errors are clustered at the district level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4.5: Robustness Checks

	IV	IV	IV
	(1)	(2)	(3)
	Log(value add)	Log(value add)	Log(value add)
Log (attendance)	6.212***	5.918***	11.179***
$\alpha * \gamma$	(2.011)	(1.905)	(3.315)
Log L	0.422***	0.448***	0.838***
α	(0.017)	(0.016)	(0.019)
Log K	0.568***	0.529***	
	(0.012)	(0.012)	
log(contractors)		0.050***	
		(0.002)	
log(1+K)			0.037***
			(0.002)
γ	14.73***	13.20***	13.33***
	(4.505)	(4.029)	(3.767)
P-value $\gamma = 1$	0.002	0.002	0.001
Observations	56786	56786	78983
Adjusted R^2	0.743	0.711	0.742

Notes: the table reports the 2SLS estimate of the effect of attendance on output and productivity, using as instruments 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 (1) reports the main specification; Column (2) adds the the control *log(contractors)*, which represents the natural logarithm of number of days worked by subcontractors in the year plus 1. Column (3) substitutes *log(K)*with *log(K+1)* in order to include firms that report zero capital stock. L is the number of workers employed in the firm at the beginning of the year; K is the firm's gross fixed capital at the beginning of the year; District level controls include yearly rainfall and temperature. Firm level controls include: rural location, ownership type, organisation type, two digit industry. All specifications are based on yearly firm level data and include district and year fixed effects. Standard errors are clustered at the district level.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4.6: Parameters Values

Calibrated Parameter		Value	Source/Target
High ag wage	θ_L	44	data
Low ag wage	θ_H	66	data
Probability $\theta = \theta_H$	ρ	0.2	
Length harvest season		0.33	
Labour coefficient	α	0.42	estimated
Attendance coefficient	$\gamma\alpha$	6.18	estimated
Elasticity of absence wrt ag wage	δ	0.69	estimated
Attendance when ag wage is low	$a^{FIX}(\theta_L)$	91.03%	data
Employment	L_m^{FIX}	49.18	data

Notes: $a^{FIX}(\theta_L)$ is equal to the mean attendance rate in the non-harvest months. $\theta_L = \overline{w_{ag}} - sd(w_{ag})$ and $\theta_H = \overline{w_{ag}} + sd(w_{ag})$where $\overline{w_{ag}}$ is mean agricultural wage and $sd(w_{ag})$ is the within district standard deviation of agricultural wages.

Table 4.7: Calibration Results

Flexible labour market

θ	attendance (%)	wage	employment	output
θ_L	91.03	100.00	52.59	102.71
θ_H	91.03	116.67	43.77	94.93
average	91.03	103.33	50.83	100.71

Rigid labour market

θ	attendance (%)	wage	employment	output
θ_L	91.03	100.00	49.18	100.00
θ_H	88.14	100.00	49.18	81.90
average	90.42	100.00	49.18	96.37

Notes:manufacturing wages are normalised to be equal to 100 in the rigid labour market case. Output is normalises to be equal to 100 in rigid labour market case when agricultural productivity is low.

5. Seasonal Absences and Rural-Urban Migration: Evidence from an Indian Jute Mill

5.1 Overview

In this chapter I use personnel data from a large jute mill based in West Bengal, India to provide detailed evidence of seasonality in workers' attendance and the consequences of this phenomenon on firms' output. The availability of information at the worker level allows me to understand which workers are more likely to engage in seasonal absences. Moreover, by matching high frequency attendance and output data I can establish a direct link between seasonal absences and output loss and determine to what extent the firm is able to adjust its production process in response to fluctuations in workers' attendance.

The jute industry provides an interesting setting for this study because, for historical reasons, it employs a large number of workers who migrated to West Bengal from neighbouring states (De Haan, 1997). These workers typically come from rural areas and move in search of employment, however, in most of the cases, they leave their families behind, thus keeping strong ties with their villages of origin. Indeed, over 60% of the workers employed in the plant object of this study have a permanent address in a district different from the one in which the firm is located and about 50% come from other states. Moreover, there is substantial heterogeneity in the districts of origin, which allows me to relate workers' attendance patterns to agricultural labour demand shocks in their home districts.

I find that rural migrants are much more likely to engage in seasonal absences than local workers. However, their attendance does not respond to temporary shocks to agricultural productivity. On the other hand, local workers' attendance is much less seasonal, but they are more likely to respond to weather shocks. This suggests that, while the former operate de facto as seasonal workers and once they pay the cost of leaving the firm they do not respond to small changes on agricultural productivity. In fact, the farther the district of origin is from the factory, the higher the workers' absence rate during the harvest season and the lower the response to positive shocks to agricultural productivity.

I estimate the impact of seasonal absences on firms' productivity using the timing of harvest as an instrument for workers' attendance. The results suggest that 1% decrease in workers' attendance during the harvest season decreases the firms' output by 1.6%. This more than proportional response suggests that the firm is unable to adjust its production process in response to changes in workers' attendance and that increasing the proportion of absent workers decreases the average productivity of those present by generating disruptions and bottlenecks.

5.2 Context

The analysis performed in this chapter relies on payroll data from a large jute mill located in West Bengal, India. This is a large textile plant that specializes in jute products. It is located within Kolkatas urban area and employs over 2,500 workers.

This type of factory is common in the area, as India is by far the world's greatest producer of jute and West Bengal accounts for over 70% of overall Indian raw jute production[1]. It follows that the jute manufacturing industry is concentrated in this area: out of the 78 registered jute mills operating in India, 63 are located in West Bengal[2].

The main items produced are jute sacks to contain food products, which are sold both for the domestic market and for export. Its level of profitability is average, and rather stable in the past few years. The firm is entirely owned and managed by family members, which is a feature of the majority of Indian textile firms (Bloom et al., 2012).

[1]Source: Indiastat

[2]Source: Office of the jute commissioner website: http://jutecomm.gov.in

The payroll data provides attendance information of both permanent and temporary workers as long as they have a formal employment relationship with the firm. There is no information about voucher workers and trainees.

These workers are hourly paid and they do not enjoy any overtime premium nor paid sick leave. According to their contracts each employee should work 8 hours per day for 6 days a week. Permanent workers can take up to 13 days off for paid statutory leave if they have worked for a minimum number of days during the year. They also enjoy other benefits established by the law such as pension funds, paid festival leaves, and high protection legislation. Temporary workers can be sent home by the management if they show up to work in a day but they are not needed (i.e. if there are no orders to be completed or no machines available) and they have few benefits and almost no employment protection.

Most of the workers employed in the plant have a permanent address in another district and about half of them come from states other than West Bengal. This is a common feature for this industry. In fact, for historical reasons, the jute industry located in West Bengal employs a large amount of workers who migrated in search of employment from a number of other northern Indian states, mainly Bihar. These are typically male rural "migrants" who leave their spouses and children in the village to take care of the land.

De Haan (1997) provides a good description of the livelihoods of jute mill workers and explains how they have remained virtually unchanged since the beginning of the twentieth century. Qualitative evidence suggests that maintaining such strong ties with their rural backgrounds, these workers are highly likely to take long leaves and return to their villages particularly during the harvest season. In fact, firms operating in this industry have historically faced very high absence rates (Chand and Banerji, 1952)during periods of peak agricultural activities.

The production process is mostly based on low skilled labour, although some occupations such as weaving and spinning require up to one year of training. This implies that workers cannot be easily substituted in case of absence or allocated to different occupations. De Haan (1999) argues that temporary workers are hired precisely with the objective to substitute the absent workers during periods of high absence rates.

5.3 Data and Descriptive Statistics

5.3.1 Payroll Data

This chapter relies on payroll data from a large jute mill for the period between September 2010 and June 2014. The original dataset contains attendance and wage information for all regular[3] workers employed during this period, 3,384 in total, at fortnight frequency.

It is an unbalanced panel as workers are reported in the data only if they work for at least one hour during the fortnight and no information is provided about accessions nor separations. Since the main outcome of interest is workers' attendance, and it is common for workers to be absent for a period of 15 days or more, I impute the missing observations as zero whenever a worker is absent from the dataset up to 6 consecutive periods (3 months) and then returns to the factory. Workers with attendance gaps longer than 6 consecutive fortnights are excluded from the analysis[4]. I also drop 58 workers whose average absence rate was above 52% (top 1% of the distribution) to remove outliers.

I merge the payroll data with the firm's master database that includes information on workers' characteristics, gender, date of birth, and district of residence. In this data cleaning process, I exclude 271 more workers for whom no district or origin information was available or who are registered as retired. The final dataset includes attendance information for 2,801 workers.

Panel A in table 5.1 reports summary statistics for some of the workers' characteristics included in the dataset. The average worker age is almost 37 years and the average tenure within the firm is 9.43 years; almost all of the workers are male (97%) and 82% are literate; the majority of workers are Hindu, but 14% are Muslim; in most cases workers have permanent contracts as the share of temporary workers is on average 8%.

As is the case for most jute mills in the area, the majority of workers are migrants from rural areas. In fact only 41% have a permanent residence within the Kolkata urban area[5], where the factory is located, and only 53% have permanent residence in West Bengal. The map presented

[3]These are all the workers employed full time in the factory. At each particular point in time there may have been up to 500 additional voucher workers or trainees for which no information is provided.

[4]Using this criterion I am dropping 254 workers.

[5]I define Kolkata urban area as the districts of Kolkata and Howra.

in figure 5.1 shows that there is substantial heterogeneity in the workers' district of origin. In fact, they are spread across various northern Indian states including Bihar, Uttar Pradesh, Orissa and Jharkhand. The average distance between Kolkata and the workers' district of origin is 239 km.

Workers attendance

The payroll dataset contains information about the number of hours worked by each worker at fortnight frequency. The working hours are classified into three categories: regular hours, overtime hours, and extra hours, where the latter indicates hours worked during the worker's day off. Since each worker is expected to work 8 hours a day, 6 days a week, with the exception of public holidays and statutory leave, I am able to estimate for each worker the number of hours scheduled in each fortnight. This allows me to calculate the workers' absence rate as 1 minus the share of hours worked over the total number of hours scheduled in a fortnight. Figure 5.2 shows the distribution of workers' absences by fortnight. Absence rates in this firm are in general very high, with a median level of 19%. However, there is a substantial heterogeneity across the period with absence rates ranging from 10% to 30%.

Characteristics of migrant workers

Table 5.2 reports summary statistics for workers characteristics by migration status, where workers are defined as rural migrants if their district of residence is outside Kolkatas urban area. The table shows that while the migrants are on average 1.42 years younger than local workers, they actually have 0.88 years longer tenure within the firm. Migrants are also 2 percentage points less likely to be temporary workers and they earn slightly higher wages on average. Local workers are 4 percentage points more likely to be female, 8 percentage points more likely to be literate, and 6 percentage points less likely to be Muslim. Most of the migrants come from states other than West Bengal and the average distance between Kolkata and their district of residence is 421 km. Absence rates are 5 percentage points higher for the migrants while overtime rates are identical. Figure 5.3 shows the distribution of workers' average absence rates by migration status. The entire distribution of average absence rates for rural migrants is shifted to the right, showing that the difference in means is not driven by few outliers but rather it is common for these workers to have lower levels of attendance. A further inspection of the data shows that the higher absence rates are mostly driven by migrant workers taking long leaves and then returning to the firm. Since migrants are less likely to be temporary workers, this cannot be explained by the nature of the contract.

5.3.2 Output Data

The plant is mostly engaged in production of jute fabric and sacks to contain food. While there could be differences in quality from one batch to another, the firm uses the weight of the produce to monitor output. This measure, originally collected on a daily basis, is aggregated at the fortnight level in order to be matched with payroll data. To ensure comparability across periods, given that there is some variation in number of days in which the firm operates, I calculate output per day. I exclude output produced on Tuesdays, which is when most workers have their weekly day off. Summary statistics are reported in panel B of table 5.1.

5.3.3 Crop Calendar and Weather Shocks

I combine these data with information about the local agricultural calendar and weather shocks in the workers district of origin. I focus only on the rice crop calendar for the following reasons: rice is the major crop in the states where workers employed in this plant are originally from; I have no access to land use data for the years covered by the analysis so I cannot construct a more accurate measure of crop harvest intensity and monthly weather shocks that I use in the other chapters of this thesis.

I obtain a district specific crop calendar for *rabi* and *kharif* rice from the 1967 Indian Crop Calendar published by the Directorate of Economics and Statistics. This allows me to determine when it is harvest season both in the workers district of origin and in the area in which the factory is located. Figure 5.4 shows that there is substantial heterogeneity in the calendar months in which rice is harvested across the workers district of origin. In particular, *rabi* rice is harvested between April and June and *kharif* rice is harvested mainly during the months of November and December. Importantly, there is no calendar month in which rice is harvested in all districts.

I construct district-wise monthly weather shocks using rainfall and temperature data collected by the Center of Climatic Research at the University of Delaware, combined with the district specific rice crop calendar. Following Donaldson (2015), I calculate cumulative rainfall and average temperature during the rice growing season, which is defined as the period of time between the first month of sowing and the last month of harvesting.

Summary statistics for these variables are reported in panel C of table 5.1. It is important to note that weather shock data are available only for the months in which rice is harvested and they refer to cumulative rainfall and average temperature affecting the relative growing season.

5.3.4 Seasonal Absences

Figure 5.5 plots average attendance rates for each calendar month by migration status. The solid line represents attendance of local workers, while the dashed line represents attendance of rural migrants. The shaded area indicates the calendar months in which the harvest of rice takes place in the region. During the harvest seasons, attendance is lowest for both groups but the pattern is much stronger for rural migrants. In fact, rural migrants' average attendance rates, while comparable to those of local workers outside of harvest season, drop below 75% in some of the harvest months.

Figure 5.6 shows that the firm's output follows the same seasonal pattern as workers' attendance. This suggests that the firm is not able to adjust its production process in response to workers' absences. Even if the firm could respond to workers' absences by hiring casual labour or by asking present workers to work overtime, the graph suggests this adjustment does not take place. In fact, the variation in average output per day appears to be even larger than that in workers' attendance. During harvest months there are some periods during which the plant produces less than 75 metric tonnes per day, much below the average of 85 and the maximum of almost 100[6].

[6] Further inspection of the data shows that this is not driven by few events such as power outages or strikes. The raw data shows that, even if there are some days of very low productivity, most of the daily observations are clustered around the period average

5.4 Empirical Strategy

The seasonal patterns in workers' absences could be driven by factors other than agricultural labour demand. In order to test the hypothesis that workers' attendance responds to changes in their outside option I use two different strategies. First, I look at to what extent their absence rates vary between harvest and non-harvest seasons. Then, I exploit exogenous shocks to agricultural productivity to assess to what extent absence rates, during the harvest season, respond to changes in agricultural labour demand.

In order to measure to what extent absence rates increase during the harvest season I estimate the following empirical model:

$$Absence_{idfmt} = \beta harvest_{dm} + v_t + w_i + \epsilon_{idfmt} \tag{5.1}$$

where: $Absence_{idfmt}$ is the percentage of working hours lost because of absence of worker i originally from district d in fortnight f, calendar month m and year t, calculated as is defined as 1 minus the share of regular hours worked over the total number of hours scheduled in the fortnight; $harvest_{dm}$ is a dummy indicating whether it is harvest season for rice in the workers' district of origin d in calendar month m. y_t and w_i represent year and worker fixed effect, respectively.

In order to account for heterogeneous responses in the absence rates of rural migrants, I interact $harvest_{dm}$ with a dummy indicating whether the worker's district of origin is outside Kolkata's urban area: $migrant_i$ in the following equation:

$$Absence_{idfmt} = \beta_1 harvest_{dm} + \beta_2 harvest^{*}_{dm} migrant_d + v_t + w_i + \epsilon_{idfmt} \tag{5.2}$$

The limited variation in rice harvest calendars in the area prevents me from including calendar month fixed effects in the regression and, therefore, from testing whether, within a calendar month, the increase in absence rates is due to harvest or other seasonal events.

Therefore, I exploit weather shocks to agricultural productivity to test whether absence rates are higher when agricultural productivity is higher and workers have better outside options in agriculture. To do so I estimate the following equation:

$$Absence_{idfmt} = \beta_1 Rainfall_{dmt} + \beta_2 Temperature_{dmt} + v_{dm} + w_i + \epsilon_{idfmt} \tag{5.3}$$

where $Rainfall_{dmt}$ and $Temperature_{dmt}$ represent cumulative rainfall and average temperature during the growing season of rice harvested in calendar month m of year t in the workers' district of origin d. To control for the fact that rice growing in different seasons, i.e. *rabi* or *kharif*, are exposed to completely different weather conditions and the fact that different districts have different soil and geographic structure, I include district time calendar month fixed effect v_{dm}.

Since weather shock data are only available for the months in which rice is harvested, this specification is looking at whether differences in agricultural productivity from one year to another are affecting absence rates during the harvest season.

The results reported in chapter 2 indicate that an increase in growing season rainfall has a positive effect on crop yield; while higher growing season temperature reduces crop yield. Moreover, a positive shock to agricultural productivity increases labour demand and, as a consequence, agricultural wages. Therefore, if workers' absence rates respond to changes in their outside option, I expect that an increase in growing season rainfall would increase absences and an increase in growing season temperature would decrease them.

To evaluate the effect of absences on output I need to estimate the production function proposed in chapter 3:

$$Y = A(a^{\gamma} L)^{\alpha} K^{\beta} \tag{5.4}$$

where Y represents the firm's output; A is total factor productivity; a is workers' attendance rate; L is the number of hours scheduled and K is capital. The parameter of interest is γ, which is the factor by which attendance affects labour productivity. In particular, if $\gamma = 0$ the firm can fully adjust its production process if workers are absent and changes in attendance have no impact on output; if $\gamma = 1$ changes in attendance affect output as much as changes in labour input; if $\gamma > 1$ absences decrease the productivity of the hours effectively worked. This function can be represented by the following empirical model:

$$log(Y_{fmt}) = log(A) + \alpha\gamma log(att_{fmt}) + \alpha log(L_{fmt}) + w_t + u_{fmt} \quad (5.5)$$

where Y_{fmt} represents firm's output in fortnight f, calendar month m and year t ; att_{fmt} is the average worker's attendance rate in period fmt and L_{fmt} is the total number of hours scheduled in period fmt. This model is estimated on period level data so one observation represents a fortnight. I include year fixed effects w_t to control for changes in capital stock and other unobservables.

In order to obtain a causal estimate of the parameter of interest I use an instrumental variable strategy in exploring the relationship between attendance rate and the timing of harvest discussed above. The first stage regression is represented by the following equation:

$$log(att_{fmt}) = \beta share\_harvest_m + w_t + e_{fmt} \quad (5.6)$$

where: $share\_harvest_m$ represents the share of workers in whose districts it is rice harvest season in calendar month m.

5.5 Results

Table 5.3 reports the results for estimation of equations 5.1 and 5.2 both with and without individual worker fixed effects. Column (1) shows that absence rates are on average 3 percentage points higher during the harvest months. Moreover, rural migrants' absence rates are 4.4 percentage points higher than those of local workers. Column (2) shows that the increase in absence rates is almost entirely driven by the behaviour of rural migrants. Indeed, rural migrants' absence rates are 5 percentage points higher during the harvest season, while the absence rates of local workers increase only by 0.66 percentage points. Column (3) and (4) show similar results when including individual worker fixed effects, suggesting that the results are not driven by differences in composition of the workforce across seasons. An alternative interpretation, however, is that rural migrants are de facto treated as seasonal workers and are not expected to show up when they are busy with outside agricultural jobs.

Table 5.4 reports the estimates of equation 5.3. Column (1) shows that 100 mm increase in growing season cumulative rainfall increases absence rates during the harvest by 0.23 percentage points. This implies that one standard deviation increase in rainfall (513 mm) would increase absence rates by 1.18 percentage points. Column (2) shows that 1 °C increase in growing season temperature decreased absence rates by 1.79 percentage point. Therefore, one standard deviation increase in temperature (1.18 °C) would decrease absence rates by 2.11 percentage points. Rainfall and temperature are likely to be negatively correlated so the coefficients decrease slightly when both are included in the same regression as reported in column (3). Columns (4) to (6) report the results of the interaction between the weather measures and the dummy variable indicating whether the worker is a rural migrant. None of the interaction coefficients are significantly different from zero and they have opposite sign with respect to the main coefficients. This suggests that the effect found in columns (1) to (3) was mainly driven by local workers and that while rural migrants are more likely to be absent during the harvest season, they do not respond to shocks to agricultural productivity.

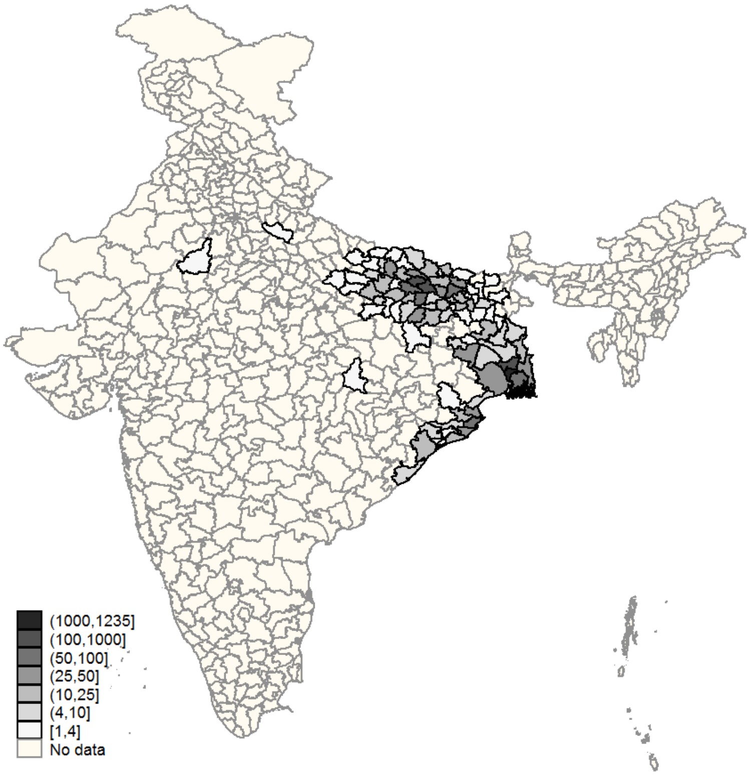
To interpret the results, I repeat the analysis including an interaction between the main independent variable and a measure of distance between the factory and the workers' district of origin. These results are reported in table 5.5. Column (2) shows that workers' absence rates increase in the distance between the factory and the workers' district of origin; while column (4) show that the workers' response to positive shocks to agricultural productivity decrease in this distance.

The estimates of the effect of seasonal absences on the firms' output are reported in table 5.6. Column (1) reports the OLS estimate for the effect of attendance on productivity. Attendance and productivity appear to be positively correlated. In particular, a 1% increase in attendance would increase output by 1.4%. This suggests that changes in attendance rates have a much larger impact than changes in the labour force. However, this analysis is based on data from a single plant and the variation in hours scheduled exploited to estimate this coefficient comes mainly from the number of working days available in a fortnight, which depends on the length of the month and public holidays. Column (2) reports the first stage regression that is the effect of harvest on workers' attendance. As already discussed above, attendance rates are lower during the harvest season, the relationship is negative and statistically significant. Column (3) reports the second stage regression. The results suggest that a 1% decrease in workers' attendance during the harvest season decreases the firms' output by 1.6%. The difference between this coefficient and the one reported in column (1) is due to the fact that the instrumental variable strategy adopted allow me to identify the effect of seasonal absences rather than absences in general. Although for this firm these are by far the most important sources of variation in workers' attendance, they may have a higher impact on firms' productivity as they cause a larger disruption in the production process in a period in which it is harder to find substitute workers since a large fraction of the labour force is busy with agricultural activities.

The estimates of the parameter of interest γ is 7.6 for the OLS regression and 9.6 for the IV estimation. In both cases the null hypothesis that is equal to 1 can be rejected at a 10% level of significance.

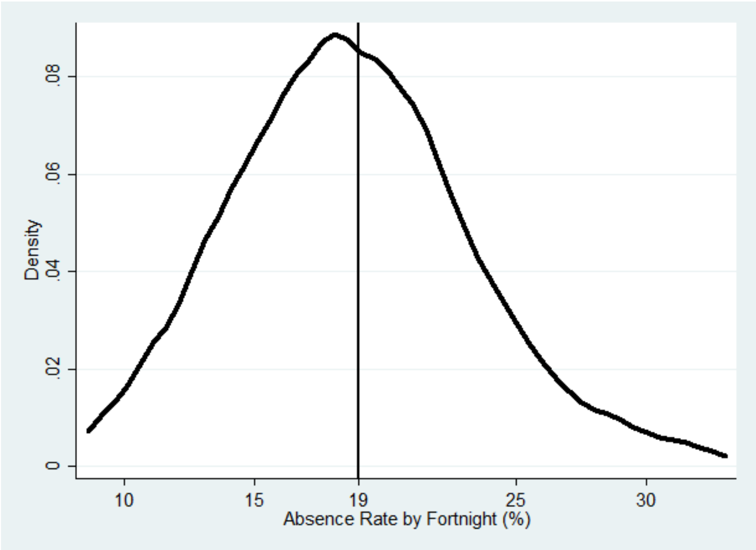
Figures

Figure 5.1: Number of Workers by District of Origin



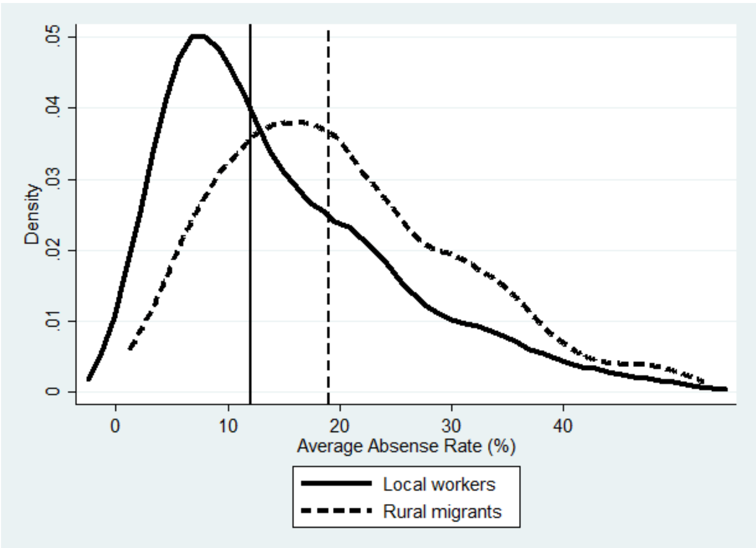
Notes: The figure represents the number of workers employed by the jute mill by district of origin on a map of India.

Figure 5.2: Distribution Absence Rates Over Time



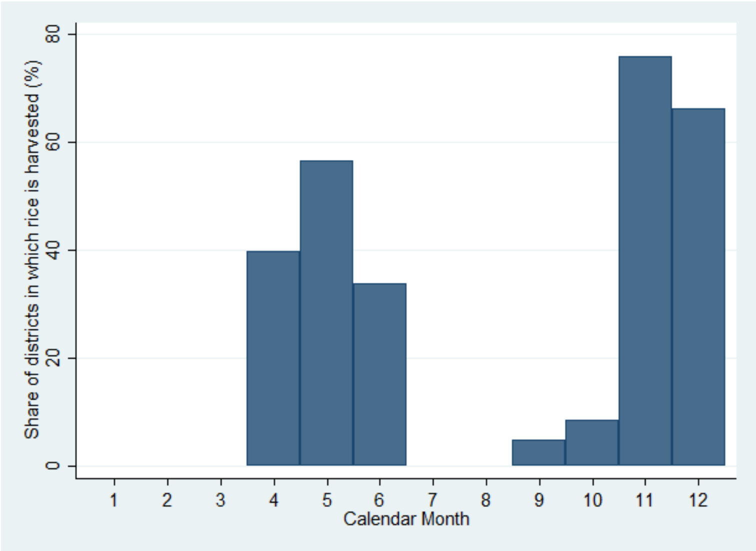
Notes: The figure plots the distribution of absence rates by fortnight, obtained by aggregating workers absence rated into fortnight averages. The vertical line indicates the distribution median, corresponding to 19%. Absence rate is defined as the 1 minus the share of regular hours worked over the total number of hours scheduled in a fortnight.

Figure 5.3: Distribution Workers' Absence Rates



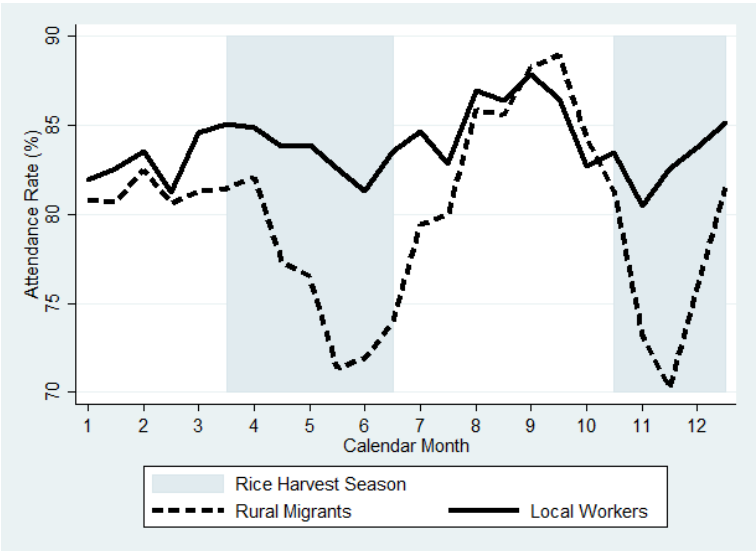
Notes: The figure plots the distribution of workers' averge absence rates by workers' migration status. Workers absence rated into individual averages. Rural migrants are defined as workers whose district of residence is outside Kolkata's urban area (defined as the districts of Kolkata and Howra). Absence rate is defined as the 1 minus the share of regular hours worked over the total number of hours scheduled in a fortnight.

Figure 5.4: Rice Harvest by Calendar Month



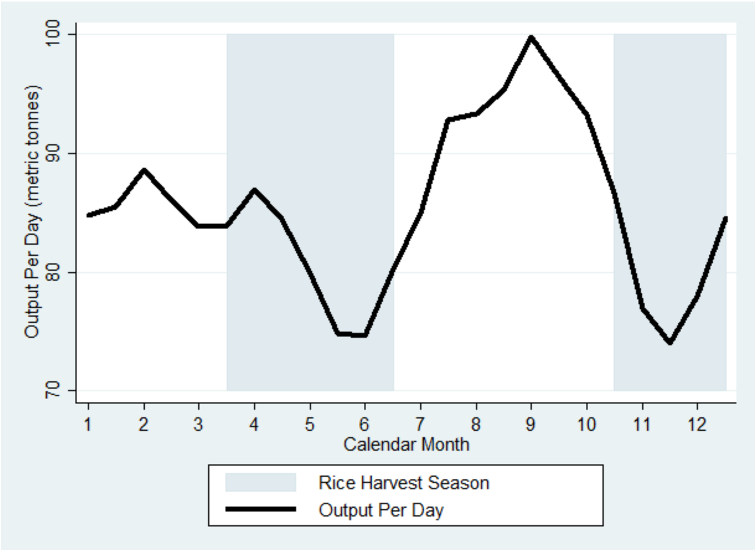
Notes: The figure represents the share of workers' district of origin in which rice is harvested by in each calendar month.

Figure 5.5: Seasonal Absences



Notes: The figure plots average attendance rates for each calendar month by migration status. The solid line represent attendance of local workers; while the dashed line represents attendance or rural migrants. Rural migrants are defined as workers whose district of residence is outside Kolkata's urban area (defined as the districts of Kolkata and Howra). The shaded area indicates the calendar months in which harvest of rice takes place in the area including the workers' districts of orgin.

Figure 5.6: Seasonality in Output



Notes: The figure plots average output per day for each fortnight. The shaded area indicates the calendar months in which harvest of rice takes place in the area including the workers' districts of orgin.

Tables

Table 5.1: Summary Statistics

Variable	Mean	Standard Deviation	Observations
Panel A: Worker's Data			
Age	36.66	8.75	206,307
Tenure	9.44	4.51	205,389
Female	0.02	0.16	206,598
Read/write	0.82	0.39	146,986
Muslim	0.14	0.34	148,705
Temporary	0.08	0.27	206,629
Local (district)	0.41	0.49	206,629
Local (state)	0.53	0.50	206,629
Distance to district of origin (km)	240.94	252.31	206,629
Absence rate (%)	18.62	27.10	206,629
Overtime rate (%)	0.08	0.16	206,629
Daily wage (INR)	273.88	56.12	206,629
Panel B: Production Data			
Output per day (metric tonnes)	82.65	9.38	92
Panel C: Crop Calendar & Weather Data			
Rice harvest	0.29	0.45	206,629
Growing season cum rainfall (100 mm)	6.64	5.13	60,618
Growing season avg temperature (°C)	25.74	1.18	60,618

Notes: Panel A reports summary statistics for a panel dataset of worker characteristics, attendance behaviour and wages in which the unit of observation is at the worker-fortnight level. This is an unbalanced panel covering 2,801 workers over the period between September 2010 to June 2014. The information about workers' characteristics comes from the firms workers' database collected at three points in time between July 2011 and July 2014. *Age* and *Tenure* are calculated respectively as the difference between the current year and the worker's year of birth and the year the worker joined the firm; *Read/write* is a dummy indicating whether the worker is able to read and write; *Muslim* is equal to one for Muslim workers, zero for Hindu workers; *Temporary* is equal to one for temporary workers, zero for permanent workers; *Local (district)* is a dummy indicating whether the worker's permanent address is within Kolkata's urban area; *Local (state)* is a dummy indicating whether the worker's permanent address is within West Bengal; *Distance to district of origin* is the distance between the centroid of the worker's district of origin and Kolkata; *Absence rate* is calculated as 1 minus the share of regular hours worked over the total number of hours scheduled in the fortnight; *Overtime rate* is defined as the number of overtime hours worked during the fortnight divided by the total number of hours scheduled; *Daily wage* is calculated as total payment received in the fortnight divided by number of (8-hour) days worked, if the worker worked zero hours during the fortnight, the observation is replace by the workers' average daily wage. Panel B reports firm's output per day, measured in terms of metric tonnes, aggregated at the fortnight level. Output produced on tuesdays is excluded as most workers have a dey off in these days. Panel C report summary statistics for rice harvest dummy and weather shocks. *Rice harvest* is a dummy indicating whether is harvest season for rice in the workers district of origin; *Growing season cum rainfall* and *Growing season avg temperature* represent cumulative rainfall and average temperature during rice growing season in the workers' district of origin, these measure are reported in the month in which rice is harvested.

Table 5.2: Characteristics of the Migrants

	Local Workers (obs: 54,188)	Rural Migrants (obs: 90,911)	
	mean (st dev)	mean (st dev)	diff (st error)
Age	38.32	36.90	1.42***
	(8.49)	(8.47)	(0.05)
Tenure	9.69	10.574	-0.88***
	(4.32)	(4.42)	(0.02)
Female	0.05	0.00	0.04***
	(0.21)	(0.06)	(0.00)
Read/write	0.87	0.79	0.08***
	(0.34)	(0.41)	(0.00)
Muslim	0.10	0.16	-0.06***
	(0.30)	(0.37)	(0.00)
Temporary	0.05	0.04	0.02***
	(0.22)	(0.19)	(0.00)
Local (district)	1.00	0.00	
	(0.00)	(0.00)	
Local (state)	1.00	0.16	
	(0.00)	(0.37)	
Distance to district of origin (km)	0.00	421.83	
	(0.00)	(190.15)	
Absence rate	0.15	0.20	-0.05**
	(0.21)	(0.30)	(0.00)
Overtime rate (%)	0.08	0.08	-0.00
	(0.16)	(0.17)	(0.00)
Daily wage (INR)	280.99	282.37	-1.38***
	(57.11)	(53.20)	(0.30)

Notes: The table reports summary statistics for the worker characteristics by migration status, where rural migrants are defined as workers whose district of residence is outside Kolkata's urban area (defined as the districts of Kolkata and Howra). The sample is restricted to those observations for which all variables are available. *Age* and *Tenure* are calculated respectively as the difference between the current year and the worker's year of birth and the year the worker joined the firm; *Read/write* is a dummy indicating whether the worker is able to read and write; *Muslim* is equal to one for Muslim workers, zero for Hindu workers; *Temporary* is equal to one for temporary workers, zero for permanent workers; *Local (district)* is a dummy indicating whether the worker's permanent address is within Kolkata's urban area; *Local (state)* is a dummy indicating whether the worker's permanent address is within West Bengal; *Distance to district of origin* is the distance between the centroid of the worker's district of origin and Kolkata; *Absence rate* is calculated as 1 minus the share of regular hours worked over the total number of hours scheduled in the fortnight; *Overtime rate* is defined as the number of overtime hours worked during the fortnight divided by the total number of hours scheduled; *Daily wage* is calculated as total payment received in the fortnight divided by number of (8-hour) days worked, if the worker worked zero hours during the fortnight, the observation is replace by the workers' average daily wage.

Table 5.3: Seasonality in Workers' Absences

	Dependent Variable: Absence Rate (%)			
	(1)	(2)	(3)	(4)
Rice Harvest	3.043***	0.665***	3.362***	0.653***
	(0.206)	(0.178)	(0.202)	(0.178)
Rice Harvest*Rural Migrant		4.362***		5.055***
		(0.383)		(0.375)
Rural Migrant	4.392***	3.059***		
	(0.427)	(0.421)		
Worker FE	-	-	Y	Y
Year FE	Y	Y	Y	Y
Observations	206629	206629	206629	206629
Adjusted R^2	0.010	0.011	0.150	0.152

Notes: The dependent variable is the share of working hours lost because of absence in a fortnight, it is calculated as 1 minus the share of regular hours worked over the total number of hours scheduled in the fortnight; *Rice Harvest* is a dummy indicating whether is harvest season for rice in the workers district of origin; *Rural Migrant* is a dummy indicating whether the worker's permanent address is outside Kolkata's urban area. Columns (1) and (2) include year FE. Columns (3) and (4) include year and worker effects. Standard errors are clustered at the worker level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5.4: Seasonal Absences and Agricultural Productivity

	Dependent Variable: Absence Rate (%)					
	(1)	(2)	(3)	(4)	(5)	(6)
Cum Rainfall (100 mm)	0.230***		0.172***	0.251***		0.201***
	(0.061)		(0.063)	(0.063)		(0.066)
Avg Temperature (°C)		-1.791***	-1.414**		-2.930***	-2.159**
		(0.622)	(0.649)		(0.868)	(0.907)
Cum Rainfall * Migrant				-0.068		-0.123
				(0.147)		(0.156)
Avg Temperature * Migrant					1.390	0.776
					(1.054)	(1.115)
Contemp Rainfall (100 mm)	0.302	0.447	0.409	0.317	0.610	0.537
	(0.487)	(0.492)	(0.492)	(0.488)	(0.513)	(0.514)
Contemp Temperature (°C)	-0.175	0.110	0.096	-0.177	0.209	0.166
	(0.204)	(0.227)	(0.228)	(0.204)	(0.242)	(0.242)
Worker FE	Y	Y	Y	Y	Y	Y
District*Calendar Month FE	Y	Y	Y	Y	Y	Y
Observations	60618	60618	60618	60618	60618	60618
Adjusted R^2	0.230	0.230	0.230	0.230	0.230	0.230

Notes: The dependent variable is the share of working hours lost because of absence in a fortnight, it is calculated as 1 minus the share of regular hours worked over the total number of hours scheduled in the fortnight; *Cum Rainfall* and *Avg Temperature* represent cumulative rainfall and average temperature during rice growing season in the workers' district of origin, these measure are reported in the month in which rice is harvested; *Migrant* is a dummy indicating whether the worker's permanent address is outside Kolkata's urban area; *Contemp Rainfall* and *Contemp Temperature* represent rainfall and temperature in the district in which the factory operates in the month in which production takes place. All specifications include worker district of origin*calendar month and worker effects. Standard errors are clustered at the worker level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5.5: Seasonal Absences and Distance from Workers' Districts of Origin

	Dependent Variable: Absence Rate (%)			
	(1)	(2)	(3)	(4)
Rice Harvest	0.653***	0.654***		
	(0.178)	(0.178)		
Rice Harvest*Rural Migrant	5.055***	-1.449***		
	(0.375)	(0.499)		
Rice Harvest*Distance		0.017***		
		(0.001)		
Cum Rainfall (100 mm)			0.201***	0.201***
			(0.066)	(0.066)
Avg Temperature (°C)			-2.159**	-2.161**
			(0.907)	(0.909)
Cum Rainfall * Migrant			-0.123	0.146
			(0.156)	(0.184)
Avg Temperature * Migrant			0.776	1.259
			(1.115)	(1.731)
Cum Rainfall * Distance				-0.001*
				(0.001)
Avg Temperature * Distance				-0.002
				(0.004)
Worker FE	Y	Y	Y	Y
Year FE	Y	Y	-	-
District*Calendar Month FE	-	-	Y	Y
Observations	206629	206629	60618	60618
Adjusted R^2	0.152	0.154	0.230	0.230

Notes: The dependent variable is the share of working hours lost because of absence in a fortnight, it is calculated as 1 minus the share of regular hours worked over the total number of hours scheduled in the fortnight; *Rice Harvest* is a dummy indicating whether is harvest season for rice in the workers district of origin; *Rural Migrant* is a dummy indicating whether the worker's permanent address is outside Kolkata's urban area; *Cum Rainfall* and *Avg Temperature* represent cumulative rainfall and average temperature during rice growing season in the workers' district of origin, these measure are reported in the month in which rice is harvested; *Migrant* is a dummy indicating whether the worker's permanent address is outside Kolkata's urban area; *Distance* is the distance between the centroid of the worker's district of origin and Kolkata. Standard errors are clustered at the worker level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5.6: Seasonal Absences and Productivity (2SLS)

	OLS	FS	IV	RF
	Log(output)	Log(attendance)	Log(output)	Log(output)
	(1)	(2)	(3)	(4)
Log(attendance)	1.400***		1.636***	
	(0.151)		(0.363)	
Log(hours scheduled)	0.185***	0.080*	0.170***	0.301***
	(0.063)	(0.044)	(0.064)	(0.072)
Rice Harvest		-0.069***		-0.112***
		(0.013)		(0.030)
Contemp Rainfall (100 m)	0.010	0.002	0.008	0.012
	(0.007)	(0.004)	(0.008)	(0.009)
Contemp Temperature (°C)	-0.002	-0.000	-0.001	-0.002
	(0.002)	(0.001)	(0.003)	(0.003)
γ	7.564**		9.621**	
	(2.917)		(4.768)	
P-value $\gamma = 1$	0.024		0.071	
Observations	92	92	92	92
Adjusted R^2	0.561	0.284	0.551	0.310

Notes: The dependent variable in column (1) (3) and (4) is the natural logarithm of output produced in the fortnight; while the depend variable columns (2) the natural logarithm of workers' attendance rate in the fortnight. Column (1) reports OLS estimates of the effect of attendance; column (2) and (3) report, respectively, the first and second stage of the 2SLS estimate of the effect of attendance on firms' output; column (4) reports the reduced form. All specifications include year fixed effects. Robust standard errors in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

6. Conclusions

This thesis studied the linkages between agricultural and manufacturing labour markets in a an economy in which the agricultural sector still employs the majority of the labour force. For the process of structural transformation to be completed, manufacturing firms must hire workers who have an agricultural background. However, these workers are more likely to be absent from the workplace when labour demand in agriculture is high.

To support this hypothesis, I show that fluctuations in agricultural wages, determined by seasonality and by exogenous shocks to agricultural productivity, are able to attract workers away from the factories and back to the fields. In particular, using on personnel data from large jute mill, I shows that this phenomenon is more prevalent among rural migrants who have strong ties with the agricultural sector.

Moreover, I find that this effect is stronger in firms facing a more stringent labour protection legislation, suggesting that workers take are more likely to take advantage of outside opportunities when they are less likely to lose their jobs.

Exploring exogenous shocks to agricultural productivity, namely growing season rainfall and temperature, I estimate the effect of seasonal absences on firms' productivity, finding a very large result: a 1% decrease in yearly attendance rate, caused by an increase in agricultural productivity, decreases output by 6%. This can be explained by the fact that in the periods of high labour demand in agriculture, a large amount of workers are likely to be absent simultaneously and finding replacement is harder. This causes severe disruptions in the production process and reduce also the productivity of the present workers.

I propose a theoretical model that explains the existence of this phenomenon as a consequence of the fact that firms are not able to adjust wages and labour force in response to shocks to agricultural productivity.

I use the parameters estimated in my empirical analysis to calibrate a simple model and evaluate the loss in efficiency caused by labour market rigidities. I find that in presence of seasonality in workers outside option, firms facing rigid labour market condition end up hiring fewer workers, paying lower wages and producing less than they would if they could adjust wages and employment seasonally.

While labour protection regulation is particularly stringent for firms operating in the Indian manufacturing sector, these findings apply to other settings in which labour market rigidities are asymmetric across sectors. Indeed, it is common to observe seasonal jobs in agriculture and permanent jobs in the manufacturing sector as a consequence of the different labour requirements characterising the production process.

The policy implications of this work are in line with that of literature on labour regulation and firms' performance (Besley and Burgess, 2004). In fact, this thesis provides evidence of a mechanism in which labour market rigidities may end up harming both firms and workers. In particular, I show that strong worker protection, combined with the lack of clear regulation absences, makes workers' attendance unpredictable for the firm, reducing the incentives to hire workers and, therefore, shrinking manufacturing employment.

The problem of workers' absences could be mitigated by improving human resources management and reduce the incentives for workers to be absent when they have a better outside option. The literature on management practices in this context (Bloom et al., 2012) suggest that there is large room for improvement. However, further research is needed to identify which measures could be taken by the firms to incentivise workers' attendance.

Finally, this thesis focuses only on one particular source of workers' absences. While seasonal absences appear to be important they do not fully explain why Indian firms face so high absence rates. Further research is needed to identify other causes of workers absences and to estimate their cost for the firm.

Bibliography

Adhvaryu, A., Chari, A., and Sharma, S. (2013). Firing Costs and Flexibility: Evidence from Firms' Employment Responses to Shocks in India. *Review of Economics and Statistics*.

Aghion, P., Burgess, R., Redding, S. J., and Zilibotti, F. (2008). The Unequal Effects of Liberalization: Evidence from Dismantling the License Raj in India. *American Economic Review*.

Ahsan, A. and Pagés, C. (2009). Are all labor regulations equal? Evidence from Indian manufacturing. *Journal of Comparative Economics*.

Allen, S. G. (1983). How Much Does Absenteeism Cost? *The Journal of Human Resources*.

Banerjee, A. and Duflo, E. (2006). Addressing Absence. *Journal of Economic Perspectives*.

Bartelsman, E., Haltiwanger, J., and Scarpetta, S. (2013). Cross-Country Differences in Productivity: The Role of Allocation and Selection. *American Economic Review*.

Basariya, S. R. (2015). Employee Absenteesim in Indian Industries. *International Journal of Science and Research*.

Besley, T. and Burgess, R. (2004). Can Labor Regulation Hinder Economic Performance? Evidence from India. *The Quarterly Journal of Economics*.

Bhattacharjea, A. (2006). Labour market regulation and industrial performance in India: A critical review of the empirical evidence. *Indian Journal of Labour Economics*.

Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., and Roberts, J. (2012). Does management matter? evidence from india. *The Quarterly Journal of Economics.*

Bloom, N., Mahajan, A., McKenzie, D., and Roberts, J. (2010). Why do firms in developing countries have low productivity? *The American Economic Review.*

Bryan, G., Chowdhury, S., and Mobarak, A. M. (2014). Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh. *Econometrica.*

Burgess, R., Deschenes, O., Donaldson, D., and Greenstone, M. (2014). The Unequal Effects of Weather and Climate Change: Evidence from Mortality in India. *Mimeo.*

Bustos, P., Caprettini, B., and Ponticelli, J. (2016). Agricultural Productivity and Structural Transformation: Evidence from Brazil. *American Economic Review.*

Chand, M. and Banerji, A. K. (1952). A statistical study of absenteeism in indian labour. *Sankhya: The Indian Journal of Statistics (1933-1960).*

Chaudhury, N., Hammer, J., Kremer, M., Muralidharan, K., and Rogers, F. H. (2006). Missing in Action: Teacher and Health Worker Absence in Developing Countries. *Journal of Economic Perspectives.*

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

Colmer, J. (2016). Weather, labour reallocation, and industrial production: Evidence from india. *Mimeo.*

De Haan, A. (1997). Unsettled settlers: Migrant workers and industrial capitalism in calcutta. *Modern Asian Studies.*

De Haan, A. (1999). The badli system in industrial labour recruitment: Managers' and workers' strategies in calcutta's jute industry. *Contributions to Indian Sociology.*

Donaldson, D. (2015). Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. *American Economic Review,* (forthcoming).

Fallon, P. R. and Lucas, R. E. B. (1993). Job security regulations and the dynamic demand for industrial labor in India and Zimbabwe. *Journal of Development Economics.*

Fetzer, T. (2013). Can Workfare Programs Moderate Violence? Evidence from India. *Mimeo.*

Foster, A. D. and Rosenzweig, M. R. (2004). Agricultural Productivity Growth, Rural Economic Diversity, and Economic Reforms: India, 1970-2000. *Economic Development and Cultural Change.*

Harris, J. R. and Todaro, M. P. (1970). Migration, unemployment and development: a two-sector analysis. *The American economic review.*

Hasan, R., Mitra, D., and Ramaswamy, K. (2007). Trade Reforms, Labor Regulations, and Labor-Demand Elasticities: Empirical Evidence from India. *The Review of Economics and Statistics.*

Herrendorf, B., Rogerson, R., and kos Valentinyi (2013). Growth and Structural Transformation. NBER Working Papers 18996, National Bureau of Economic Research, Inc.

Herrmann, M. A. and Rockoff, J. E. (2012). Worker Absence and Productivity: Evidence from Teaching. *Journal of Labor Economics.*

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and Manufacturing TFP in China and India. *The Quarterly Journal of Economics.*

Ichino, A. and Riphahn, R. T. (2005). The Effect of Employment Protection on Worker Effort: Absenteeism during and after Probation. *Journal of the European Economic Association.*

Imbert, C. and Papp, J. (2016a). Short-term Migration Costs: Evidence from India's Employment Guarantee. *Mime0.*

Imbert, C. and Papp, J. (2016b). Short-term Migration Rural Workfare Programs and Urban Labor Markets - Evidence from India. *Mime0.*

Jayachandran, S. (2006). Selling Labor Low: Wage Responses to Productivity Shocks in Developing Countries. *Journal of Political Economy.*

Kaur, S. (2013). Nominal Wage Rigidity in Village Labor Markets. *Mimeo.*

Khurana, A., Khurana, P., and Sharma, H. L. (2009). *Human Resource Management.* V.K. (India) Enterprises, New Delhi.

Krueger, A. B. and Mas, A. (2004). Strikes, Scabs, and Tread Separations: Labor Strife and the Production of Defective Bridgestone/Firestone Tires. *Journal of Political Economy.*

Kumar, C. R., Chakravarthy, K., and K.P.R.Rajesh (2014). A Review on Absenteeism. *International Journal of Computer & Organization Trends.*

Kumar, H. and Somanathan, R. (2009). Mapping Indian Districts across Census Years, 1971-2001. *Mimeo.*

Kumar, H. L. (2010). *Law Relating to Leave Holidays and Absenteeism in Industries.* Universal Law Publishing Co. Pvt. Ltd., New Delhi, 9th edition edition.

Kuznets, S. (1957). Quantitative Aspects of the Economic Growth of Nations: II. Industrial Distribution of National Product and Labor Force. *Economic Development and Cultural Change.*

Lewis, W. A. (1954). Economic Development with Unlimited Supplies of Labour. *The Manchester School.*

Miller, R. T., Murnane, R. J., and Willett, J. B. (2008). Do worker absences affect productivity? The case of teachers. *International Labour Review.*

Morten, M. (2013). Temporary Migration and Endogenous Risk Sharing in Village India. *Mimeo.*

MOSPI (2014). *Instruction Manual Annual Survey of Industries (Concepts, Definitions and Procedures).* http://mospi.nic.in.

Murphy, K. M. and Topel, R. H. (1985). Estimation and Inference in Two-Step Econometric Models. *Journal of Business & Economic Statistics.*

Riphahn, R. T. (2004). Employment Protection and Effort among German Employees . *Economics Letters*.

Rosenstein-Rodan, P. N. (1943). Problems of industrialisation of Eastern and SouthEastern Europe. *Economic Journal*.

Sincavage, J. R., Haub, C., and Sharma, O. (2010). Labor costs in India's organized manufacturing sector. *Monthly Labor Review*.

Sokoloff, K. and Dollar, D. (1997). Agricultural Seasonality and the Organization of Manufacturing in Early Industrial Economies: The Contrast Between England and the United States. *The Journal of Economic History*.

Sokoloff, K. L. and Tchakerian, V. (1997). Manufacturing Where Agriculture Predominates: Evidence from the South and Midwest in 1860. *Explorations in Economic History*.