Worker Absenteeism and Firm Outcomes: Evidence from the Indian Manufacturing Sector*

Sampreet Singh Goraya† Nandita Krishnaswamy§

Suanna Oh‡ Yogita Shamdasani¶

Preliminary and incomplete

Link to updated version

Abstract

Worker absenteeism is widely cited as a major impediment to private sector growth in developing countries, yet systematic evidence on its prevalence and consequences remains scarce. Using a nationally representative panel of Indian manufacturing firms, we document novel empirical findings on both the frequency of absenteeism and its impact on firm outcomes. High absenteeism rates are associated with smaller firm size and higher wages, with effects particularly pronounced in sectors with strong co-worker complementarities. Our in-depth survey of 206 firms in Odisha reveals that firms adopt costly strategies—such as hiring slack labor and rotating workers across tasks—to cope with frequent absences. To understand macroeconomic implications, we embed absenteeism risk into a model of firm dynamics with hiring frictions, wage bargaining, and firm entry/exit. The model rationalizes the empirical patterns and shows that absenteeism amplifies labor misallocation and substantially reduces aggregate productivity.

Keywords: Absenteeism, Productivity, Labor Turnover, Aggregate Output
JEL classification: *O11, L11, L25, M14*

*This draft: September 2025. We thank Anirvin Narayan, Sambhav Choudhury, Rucha Takle, Naman Agarwal, Guillermo Woo-Mora, and Matilde Casamonti for excellent research assistance. We thank Akhil Ilango for sharing the ASI Part II data from 2015 to 2019.

†Stockholm School of Economics. Email: sampreetgrg@gmail.com

‡Paris School of Economics. Email: suanna.oh@gmail.com

§Analysis Group. Email: krishnaswamy.nandita@columbia.edu

¶National University of Singapore. Email: yogita@nus.edu.sg

1 Introduction

Developing countries face a persistent wage employment challenge (Breza, Kaur and Shamdasani, 2021; Breza and Kaur, 2025). Wage employment rates are low, and even among those employed, work is often precarious, characterized by high turnover and little job security (Donovan et al., 2023). The dominant narrative in both academic and policy circles attributes this to a shortage of "regular" jobs, framing the problem as one of insufficient supply (World Bank, 2017; ILO, 2018). However, this supply-side view overlooks an important phenomenon: high levels of chronic absenteeism among wage workers (Adhvaryu et al., 2024). Worker absenteeism complicates the narrative of "too few regular jobs" due to the implication that even where wage employment exists, their quality and stability could be undermined by labor supply volatility.

Although the problem of worker absenteeism in manufacturing firms has been noted as early as the 1960s (James, 1960), systematic evidence on the extent of absenteeism and its implications for developing economies remain scarce. A large number of studies document high rates of absenteeism in health and education services (Chaudhury et al., 2006; Banerjee and Duflo, 2006; Callen et al., 2023). Public-sector absenteeism is often attributed to weak enforcement and limited disciplinary action. In contrast, the private sector is capable of tighter monitoring and imposing financial penalties; that is, workers typically do not earn wages on absent days. Nevertheless, a number of case studies demonstrate that absence rates are also high in private firms.[1] The challenge in addressing private-sector absenteeism is further compounded by the lack of available statistics from developing countries, particularly at the national level.

Using nationally representative firm data, this paper offers a comprehensive characterization of the absenteeism phenomenon in the Indian manufacturing sector and

[1] Adhvaryu et al. (2024) find that an average of 11% of workers are absent on any given day in a large Indian garment factory, with 9% of production lines experiencing absence shocks of at least 20%. A survey of 300 manufacturing firms in Uganda shows an average monthly absence rate of 15% (Cotton et al., 2004). These figures contrast with an absence rate of 2.8% in the US manufacturing sector (Bureau of Labor Statistics, 2025).

estimates its productivity consequences. We make use of newly available Annual Survey of Industries Part II data, which draws from a nationally representative sample of manufacturing establishments in India.[2] Through a careful data merging procedure, we create a unique monthly panel dataset of absenteeism spanning two decades (2000–2019), combined with annual establishment output data. We use this data to document novel empirical facts about the prevalence of absenteeism, as well as its connections to turnover, output, and wages. We supplement this analysis with primary data from in-depth surveys of 206 firms in Odisha, which provides additional insights into how firms cope with absenteeism. These findings inform our model of firm dynamics with absenteeism risk, which enables us to quantify aggregate impacts of worker absenteeism.

Our empirical analysis documents six key facts about absenteeism in the Indian manufacturing sector. Absenteeism is prevalent, largely unpredictable, and costly to establishments.

First, the average absenteeism rate is exceedingly high at 8.6%, with wide dispersion across regions.[3] A seasonal pattern shows absences rising during peak agricultural months, but its contribution to overall variation in absenteeism is economically insignificant, in contrast to large seasonal fluctuations in casual employment. Instead, we find significant regional differences: Northern and Western states record the highest rates, and states such as West Bengal and Maharashtra exceed the national average, consistent with the view that disciplining irregular workers is harder in pro-employee states (Besley and Burgess, 2004). At the same time, absenteeism remains high in pro-employer states such as Kerala and Rajasthan, suggesting that labor law leniency alone cannot explain observed patterns.

Second, while establishment fixed effects explain about 30% of the total variation—indicating persistent differences across establishments—absenteeism also has a large

[2] To our knowledge, this data has only been used in two unpublished papers (Zane, 2018; Krishnaswamy, 2019), which use seasonal patterns in agricultural employment and rainfall shocks respectively as instrumental variables to measure the impact of absences on firm output.

[3] The monthly average absenteeism rate is 8.62% whereas the yearly rate is similar at 8.52%.

transitory component. 12.5% of establishments experience month-to-month deviations as large as the mean, indicating the presence of unpredictable, time-varying absence shocks. This pattern aligns with our survey evidence that over 70% of worker absences are reported at most one day in advance. Furthermore, survey data indicate that workers are frequently absent due to illness and social obligations—including weddings, funerals, helping neighbors, and visiting relatives—which are difficult to forecast from the firms' perspective.

Third, we examine the extent to which establishments adjust their production schedules in anticipation of higher absenteeism within a year. On average, we find nearly zero covariance between months with more scheduled worker-days and months with lower absenteeism. This suggests there is limited reallocation of labor in response to monthly variation in worker absences, potentially caused by rigid production processes. However, we find substantial heterogeneity across establishments: the covariance is larger among bigger establishments, establishments with more flexible production technologies, and establishments with stronger managerial capacity.

Fourth, absenteeism is associated with substantial workforce turnover across months within establishments, reflected in monthly new hiring and separation (i.e. quits and dismissals). Absenteeism may force establishments to replace missing workers, while expectations of high future absenteeism may also lead them to hire buffer workers. To better understand firm responses, we ask firms how they deal with worker absences in our survey. A substantial share of firms report using strategies to cope with absenteeism which include hiring migrant workers (40%), hiring slack labor (17%), and rotating existing workers across time/tasks (19%).[4] These findings suggest that the costs of absenteeism extend beyond lost productivity to personnel management costs as establishments must continually reconfigure their workforce to sustain operations.

Fifth, absenteeism carries serious productivity costs. A 1% increase in absenteeism is associated with a 0.15% decline in sales and a similar decline in value added, con-

[4] The strategy of internally substituting workers across rotation spots is also studied in Adhvaryu, Gauthier, Nyshadham and Tamayo (2024)

trolling for state, sector, and year fixed effects. Even with establishment fixed effects, effects remain statistically significant: a doubling of the absence rate from one year to the next within a firm reduces output by 1.6%. These adverse productivity impacts are larger in sectors with strong co-worker complementarities—proxied by an O*NET index of teamwork and communication across workers—where output elasticities are roughly 30 percent higher than the average.

Sixth, higher absenteeism is associated with higher wages for permanent workers. It is also associated with higher reliance on contract workers, although wages of contract workers remain unaffected.[5] This pattern is consistent with the argument put forward by James (1960) that the absenteeism problem leads to the rise of a large casual labor market in India.

To interpret these patterns and quantify their macroeconomic implications, we develop a model of firm dynamics with both productivity and absenteeism risk. Absenteeism shocks are modeled as an AR(1) process with both persistent and transitory components. Firms choose the size of their workforce, but workforce adjustments are costly. These labor adjustment costs capture co-worker complementarities, where replacing an existing worker with a new hire disrupts production. They also partly reflect hiring frictions, as firms must devote resources to finding suitable matches to replace absent or departing workers.

The model also incorporates wage bargaining, which creates a positive link between absenteeism and wages. Higher absenteeism raises the marginal product of labor, leading to upward pressure on wages. However, workers only capture part of this surplus – this depends on their bargaining weight. Thus, the bargaining process ensures that wages rise with absenteeism, amplifying its cost for firms. Finally, we allow for endogenous firm entry and exit to study the long-run consequences of absenteeism for firm dynamics. Absenteeism lowers the value of entry by increasing costs

[5]Permanent workers are defined as "direct workers" that have long-term contracts, while contract workers are temporary workers typically hired through labor contractors. Details are provided in Section 2.1.

associated with workforce disruptions and hiring frictions. As a result, absenteeism dampens firm dynamism and reduces overall economic activity.

We calibrate the model parameters to match key features of the Indian manufacturing data. We begin by targeting the employment distribution across firms, using fixed costs to discipline the bottom half of the distribution and productivity persistence and dispersion to capture the dominance of large firms. Parameters governing firm entry are chosen to replicate observed entry rates and the relative size of entrants. Hiring and separation costs are set to match worker reallocation patterns, while absenteeism parameters are calibrated to the mean and dispersion of absenteeism in the data. Finally, standard values from the literature are used for technology, preferences, and bargaining power.

The calibrated model implies that eliminating absenteeism would substantially raise aggregate manufacturing productivity. We find that doubling the long-run absenteeism rate reduces aggregate output by 10%. While most of this decline reflects the mechanical effect of having fewer workers present, about one-third is explained by a reduction in aggregate productivity itself. This productivity channel is economically meaningful and highlights how absenteeism generates inefficiencies beyond the mechanical loss of labor input. The decline in productivity also feeds into higher output prices. In high absenteeism environments, firms face acute labor shortages and raise nominal wages to attract and retain workers. However, these wage increases do not translate into higher welfare, as output prices rise even more, causing real wages to fall.

Absenteeism imposes substantial efficiency costs on firms. Hiring costs rise as firms repeatedly adjust their workforce to cover for frequent absences, and these adjustments are themselves disruptive, leading to further production inefficiencies. At the aggregate level, absenteeism increases labor misallocation, as reflected in greater dispersion in the marginal product of labor. A central mechanism is wage dispersion: absenteeism shocks reduce effective labor, raise the marginal product of remaining workers, and push up their wages. As a result, firms facing higher absence rates pay higher wages, relative to otherwise identical firms. This mechanism implies that absenteeism can generate firm-specific wage premia even in the absence of technological heterogeneity, providing a novel channel for the emergence of wage inequality across firms.

Finally, absenteeism dampens firm growth and dynamism. The variance of employment growth declines in high-absenteeism environments, consistent with firms hoarding labor and entrants starting at smaller scales to buffer against attendance shocks. Taken together, these results show that absenteeism not only reduces output and productivity but also contributes to misallocation and lower dynamism in the manufacturing sector.

While our analysis suggests that reducing absenteeism could generate sizable productivity gains, achieving this would require concrete grasp of the underlying causes. Firms in our survey attribute worker absences largely to social obligations including attending weddings, funerals, and religious festivals (61%), illness (56%), and agricultural work (22%).[6] Goraya, Oh and Shamdasani (2025) delve further into this issue, showing that the time spent on social obligations is a key driver of workers' need for flexible absences and is essential for maintaining access to network benefits such as informal insurance. Consequently, substantially reducing absenteeism would require diminishing the role of such networks, which would be financially difficult and potentially undesirable from a broader social perspective. Therefore, this research informs our understanding of absenteeism both as a significant barrier to firm growth and as a complicated policy challenge.

This paper makes both empirical and theoretical contributions to the literature. First, the paper extends our understanding of absenteeism in developing countries by providing a detailed picture of the extent and consequences of absenteeism in the manufacturing sector. Recent micro-level evidence demonstrates that seasonal absen-

[6] Participants can list multiple reasons for worker absenteeism.

teeism has a negative impact on output (Zane, 2018; Krishnaswamy, 2019). Establishments make considerable efforts to reduce absenteeism and and incur significant costs in the process. Engström and Holmlund (2007) show that higher wages and benefits can expand the applicant pool and potentially reduce absenteeism, pointing to a trade-off between wage costs and attendance. More recently, Adhvaryu, Gauthier, Nyshadham and Tamayo (2024) emphasize the role of relational contracts in managing absenteeism shocks—line managers may informally "trade" extra workers across teams to cover unexpected shortfalls. We present aggregate statistics that speak to broader patterns in absenteeism as well as provide a comprehensive overview of coping strategies that firms undertake.

Second, our findings have implications for an important body of research that documents persistent heterogeneity in productivity across establishments (Syverson, 2011). This dispersion is higher in developing countries, and the left tail is fatter (Hsieh and Klenow, 2009). Our findings highlight a novel factor that explain productivity differences: the extent of labor utilization. Absenteeism drives a wedge between measured labor productivity (output per worker hired) and true labor productivity (output per worker present). Failing to account for absenteeism leads to biased inference in the measurement of productivity differences.

Third, to the best of our knowledge, this is the first paper to quantify the macroeconomic costs of absenteeism using a model incorporating dynamics, which are essential for capturing absenteeism risk from the firm's perspective. An old literature on absenteeism in developed countries emphasizes the role of non-convexities in assembly-line production when estimating effects of absenteeism (see Weiss, 1985; Barmby, Sessions and Treble, 1994; Coles and Treble, 1993; Coles and Treble, 1996). Our contribution differs both in modeling approach and in focus. Whereas previous studies have relied on static frameworks, our model embeds absenteeism shocks in a dynamic, general equilibrium framework with heterogeneous establishments. This approach captures not only the direct productivity effects of absenteeism but also its interaction with establishment decisions over the long run, such as hiring buffer workers, adjusting wage offers, and responding to anticipated absence risk. In doing so, our analysis bridges the gap between micro-level studies of absenteeism and its macroeconomic implications for labor markets in low-income economies.

Finally, our paper connects to a small but growing literature on imperfect competition and labor turnover across labor markets in developing countries (Amodio, Medina and Morlacco, 2022; Donovan, Lu and Schoellman, 2023; Amodio and de Roux, 2024). For example, establishments' exercise of monopsony power in Indian manufacturing has been found to depress wages considerably (Brooks, Kaboski, Li and Qian, 2021). Further, Cosar, Guner and Tybout (2016) show that introducing search frictions and wage bargaining into a heterogeneous establishment model can generate realistic employment dynamics and wage dispersion in emerging markets. These insights motivate our model's treatment of wage and employment outcomes as the result of bargaining under hiring frictions, rather than a competitive clearing. This allows us to analyze how wages, hiring and establishment entry respond to absenteeism shocks. Overall, our analyses suggest that worker absenteeism is key to understanding labor market dynamics in developing countries.

2 Data and Measurement

2.1 Annual Survey of Industries

We use the Annual Survey of Industries (ASI), administered by India's Ministry of Statistics and Programme Implementation each financial year (April–March). The ASI is a census of all formally-registered large manufacturing establishments and a random sample of small manufacturing establishments.

We construct our analysis dataset using multiple components of the ASI. First, we use the annual ASI cross-section data which provides comprehensive establishment-year level information including balance sheets, income statements, and establishment characteristics. Second, we incorporate monthly ASI Part II data, which contains establishment-month level information on worker attendance, absenteeism, hiring and separation. Third, we use the annual ASI panel data which contains the same set of variables as the cross-section data in addition to a consistent establishment identifier that remains fixed across years for the same establishment.

An important challenge we face is that the panel data cannot be merged directly with the cross-section data due to a lack of consistent establishment identifiers in the two datasets. To overcome this challenge, we carry out a fuzzy match using observable establishment characteristics, following Martin, Nataraj and Harrison (2017). We rely on the panel data primarily to obtain a consistent establishment identifier over time in our merged data set, which in turn enables the use of establishment fixed effects in the analysis.

We exclude a few survey years due to data quality issues and structural inconsistencies. Specifically, we drop 1999–2000 as it was missing establishment identifiers, 2005–06 as it duplicates 2004–05, 2009–10 as it is missing Part II data, and 2011-12 as the shift from calendar to financial year in the Part II data created ambiguities. We also exclude 2008–09 and 2010–11 due to poor alignment in common variables across the cross-section and Part II data, reflecting potential inconsistencies in identifiers.

Worker Types. Formal manufacturing plants operate under the Industrial Disputes Resolution Act (IDRA), 1947, which governs dispute resolution, hiring and firing, and worker rights in the event of closure. Most provisions apply to *direct* workers, who are typically engaged on daily-wage or salaried contracts. In practice, establishments can adopt a "no work, no pay" rule for unauthorized absences. An exception is statutory earned leave: workers who complete at least 240 days of work in a calendar year accrue one day of paid leave for every 20 days worked, which must be authorized by the establishment and scheduled in advance. In our context, most absences are *unscheduled* (see Section 2.2).[7]

[7]In addition, once workers cross the 240–day threshold, they become eligible for benefits such as medical insurance and provident-fund contributions (Zane, 2018).

While our monthly absenteeism statistics pertain to *direct* workers, establishments also hire *contract* workers, who are typically outside the scope of IDA provisions that apply to direct workers. The use of contract labor has grown in popularity among employers (Bertrand, Hsieh and Tsivanidis, 2024), and is pervasive among large establishments but uncommon among small ones, reflecting fixed and intermediary costs of hiring through labor contractors. In our sample's median establishment, contract workers account for 12.8% of total worker-days. Contractors are engaged to supply a stipulated *headcount* each day rather than specific individuals. Field interviews indicate that penalties for shortfalls are limited, so contract labour provides an imperfect hedge against absenteeism.

2.2 Establishment Survey

We conduct an in-depth survey with managers of manufacturing establishments in four districts in Odisha, India. We build our survey sample using the 2018–19 ASI sampling frame, focusing on the top ten industries by number of registered establishments. The survey has two aims: (i) to benchmark our hand-collected measures of worker absenteeism against ASI records to validate data quality, and (ii) to collect unique data (not captured in the ASI) on coping strategies establishments use to limit exposure to absenteeism-induced labor shortfalls. In particular, we have three modules on coping strategies: a) buffer or rotating workers, b) hiring workers via labor contractors, and c) hiring migrant workers. Our survey sample consists of 206 establishments in 4 districts. These establishments are small to mid-sized manufacturing firms founded in the late 2000s (mean $\approx$ 2007), employing about 37 workers on average, of which two-thirds ($\approx$ 24) are unskilled.

2.3 Co-worker Complementarities Index

We utilize the O*NET database which offers detailed descriptors of occupational characteristics to construct measures of co-worker complementarities by occupation.[8] In particular, we use the Work Context module which includes scores on key dimensions such as teamwork, impact, communication, and contact. Each score ranges from 1 to 5 and reflects the importance of the respective attribute for job performance, as assessed through standardized occupational surveys. We construct an occupation-wise composite index of co-worker complementarities by averaging standardized scores across the four dimensions.

3 Stylized Facts

In this section, we document six stylized facts on absenteeism in India's manufacturing sector.

3.1 Prevalence of Absenteeism

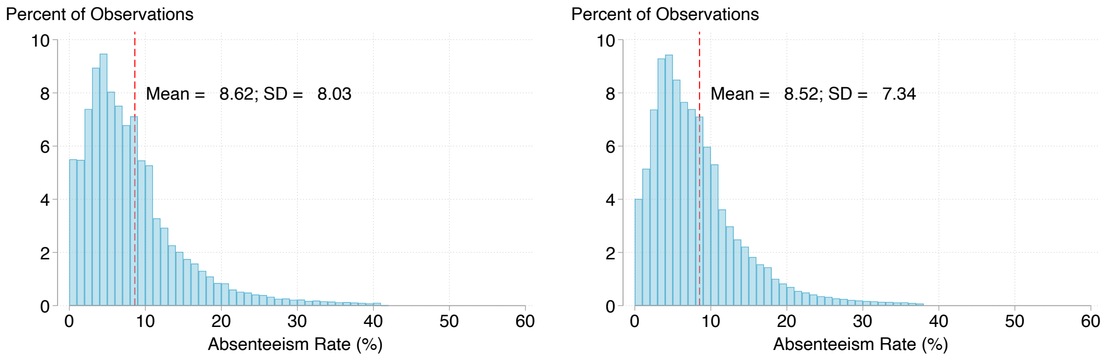
Fact 1*: The average absenteeism rate is 8.6% with substantial regional variation — ranging from 3.7% to 14.3% — across states.*

Figure 1a plots monthly absenteeism rates across establishments in India. The average is 8.62% with a fat right tail, with establishments at the top percentile experiencing up to 40% absenteeism. These moments remain practically unchanged when we aggregate the data to the annual level, as illustrated in Figure 1b.[9] The yearly average of 8.52% is almost three times as large as the absenteeism rate reported in the U.S. manufacturing sector (BEA).

[8]While one would ideally use data specific to India, such data do not exist. We thus rely on this U.S. based database, under the assumption that skill requirements for a given occupation are broadly comparable across countries.

[9]The yearly absenteeism rate is computed as $A_{ft} = \sum_m s_{m,ft} A_{m,ft}$, where $s_{m,ft}$ is the proportion of worker-days scheduled to work in a given month m and $A_{m,ft}$ is the monthly absenteeism rate.

Figure 1: Absenteeism Rates



(a) Monthly

(b) Yearly

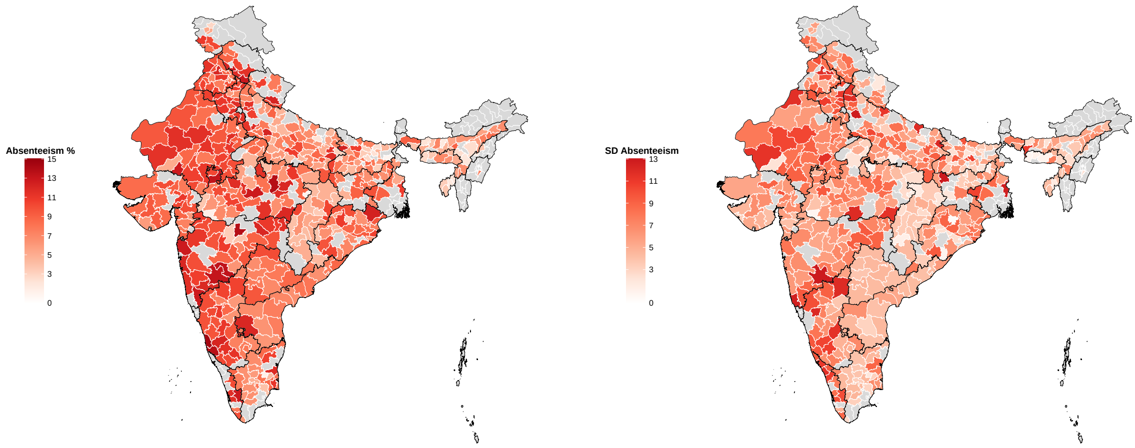
Note: Figure illustrate the distribution of monthly and yearly absenteeism rates across establishments in India between 2000-2019. Absenteeism rate is defined as the ratio of man-days lost to man-days scheduled. The red dotted line indicates the sample mean.

Is absenteeism geographically concentrated? We find substantial spatial dispersion — Figures 2a and 2b visually illustrate the mean and standard deviation of absenteeism across establishments, by district.[10] Northern and Western states tend to exhibit the highest absenteeism rates. Interestingly, states such as West Bengal and Maharashtra experience above-national-average absenteeism (see Appendix Table A.2), consistent with the view that disciplining irregular workers is more difficult in pro-worker states (Besley and Burgess, 2004). However, this interpretation is challenged by the fact that pro-employer states such as Kerala and Rajasthan also show absenteeism rates well above the national average. These findings suggest that the leniency of labor laws alone cannot fully explain observed absenteeism patterns.

India remains partly an agrarian economy, and workers often shift to agricultural employment during peak farming seasons. This pattern may introduce seasonal fluctuations in absenteeism. Figure 3 presents monthly absenteeism rates separately by state. While there is some evidence of seasonality, the pattern is not uniform across

[10]We repeat this exercise at the state level in Appendix Figure B.1. We document a strong positive relationship between the mean and dispersion of the absenteeism rate at both the district and state level in Appendix Figure B.2.

Figure 2: Geography of Absenteeism



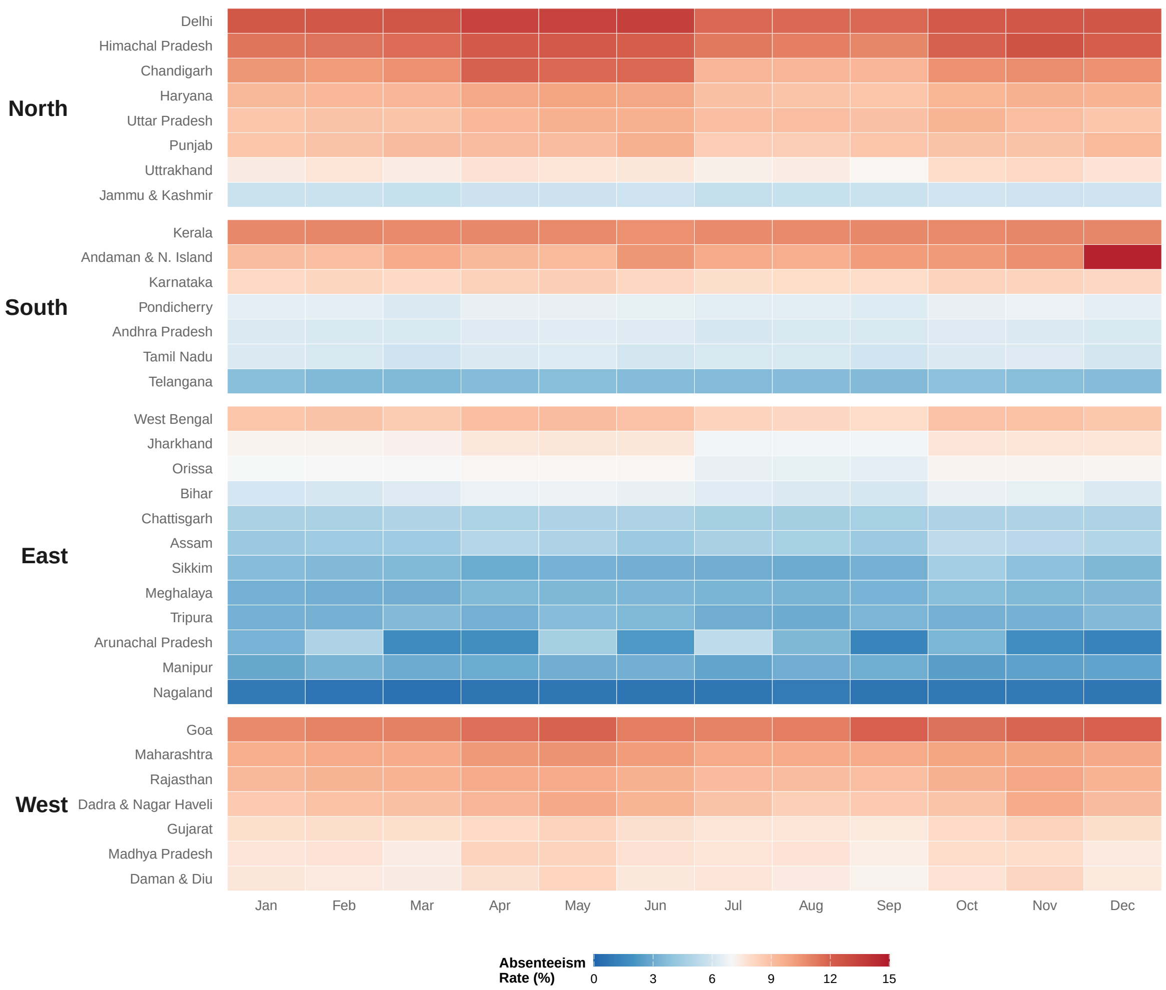
(a) Mean absenteeism (b) Standard deviation of absenteeism

Note: Figure plots means and standard deviations of monthly absenteeism rates across districts in India between 2000 and 2007. The black solid lines demarcate state boundaries.

regions. In Northern states such as Punjab, Haryana, and Uttar Pradesh, absenteeism tends to rise in April-May and October-November, which broadly coincide with the Rabi (wheat) harvesting and Kharif (rice) harvesting seasons respectively. Some states (e.g., Delhi and Kerala) exhibit persistently high absenteeism throughout the year. In the Eastern region, absenteeism remains relatively stable even during peak agricultural periods, undermining the notion of a strong nationwide seasonal effect. Overall, while some regions show absenteeism spikes during key agricultural months, seasonality only partially explains the total observed variation, indicating that other structural or institutional factors—such as establishment practices, labor market norms, or unexpected disruptions to labor supply—are likely important determinants of absenteeism risk for establishments.

Comparison of ASI and Own Survey Data. To validate our absenteeism measures, we compare firm-reported rates from our survey with those recorded in the ASI. As shown in Appendix Figure C.1, the distributions align closely, with absenteeism levels falling within a similar range across both sources.

Figure 3: Timing of Absenteeism



Note: Figure illustrates monthly averages of absenteeism by state between 2000 and 2019. States are grouped by region – North, South, East or West. Within each region, states are organized according to descending mean absenteeism rates.

3.2 Variation in Absenteeism

***Fact 2**: Establishment-level fixed effects account for 30% of the total variation in monthly absenteeism rates, indicating persistent differences across establishments. However, there remains substantial variation in absenteeism even within an establishment — 12.5% experience absenteeism shocks that reach twice the sample mean, underscoring the significance of unpredictable, time-varying shocks.*

Table 1 presents the extent of residual variation in absenteeism rates after accounting for different sets of fixed effects. Column (1) reports the standard deviation of the residuals (σ_e), while columns (2) through (6) show the share of observations with residuals exceeding 1, 2, 3, 5, and 8.52 p.p., respectively. Without any fixed effects, the standard deviation of absenteeism is large ($\sigma_e = 8.03$), and 39.4% of establishment-year observations deviate from the mean by more than 5 percentage points. Including establishment fixed effects reduces this substantially ($\sigma_e = 5.67$), which suggests that establishment-specific factors explain 30% of the variation in observed absenteeism. Further inclusion of year, year-state or year-sector fixed effects yield marginal reductions in residual variation, with the standard deviation stabilizing around 5.6 and the share of large residuals (e.g., >5pp) remaining virtually unchanged. These results highlight the presence of substantial idiosyncratic or establishment-time-specific shocks to absenteeism that cannot be captured by fixed establishment characteristics or broader temporal or sectoral factors.[11]

Next, we examine the relative importance of persistent versus transitory components in explaining establishment-level variation in absenteeism by estimating the relationship between current and lagged absenteeism rates. In particular, we estimate the following regression: $A_{ft} = \mu_a + \rho_a A_{ft-1} + \xi_{ft}$, where ρ_a captures the degree of persistence and ξ_{ft} represents transitory shocks. In Table 2, we find that absenteeism is highly persistent at the monthly level, with $\rho_a = 0.829$. However, when we move to

[11] We find similar results when considering yearly absenteeism rates (see Appendix Table B.1).

Table 1: Residual Variation in Absenteeism

	σ_e (1)	$\|e\| > 1$ (2)	$\|e\| > 2$ (3)	$\|e\| > 3$ (4)	$\|e\| > 5$ (5)	$\|e\| > 8.62$ (6)
No FE	8.030	0.866	0.743	0.631	0.394	0.125
Establishment FE	5.671	0.710	0.514	0.375	0.206	0.082
Establishment FE + Year FE	5.657	0.708	0.510	0.371	0.204	0.082
Establishment FE + Year×State FE	5.625	0.706	0.507	0.368	0.203	0.081
Establishment FE + Year×Sector FE	5.650	0.708	0.510	0.371	0.204	0.081

Note: This table summarizes residual variation in monthly absenteeism from a regression with fixed effects. Each row lists the relevant set of fixed effects included in each regression. Column 1 reports the standard deviation of the residuals (denoted as σ_e). Columns 2 to 6 report the proportion of observations that have a residual e with absolute values exceeding 1, 2, 3, 5 or 8.62% (mean monthly absenteeism rate).

yearly absenteeism, persistence declines to $\rho_a = 0.447$. This decline suggests that transitory shocks play a larger role in explaining year-to-year variation, relative to month-to-month variation. Overall, while establishment-specific factors account for part of the variation, unpredictable, time-varying shocks emerge as the dominant driver of absenteeism.

Table 2: Persistence in Absenteeism

	Log Absenteeism Rate	
	Monthly (1)	Yearly (2)
L.log(Absenteeism Rate)	0.829	0.447
	(0.001)	(0.004)
	[0.000]	[0.000]
Year FE	Yes	Yes
Sector FE	Yes	Yes
State FE	Yes	Yes
R^2	0.720	0.323
Dep Var Mean	1.689	1.841
N	1,527,579	157,631

Note: This table examines the persistence of absenteeism rates over time. The dependent variable is log absenteeism rate, and the key independent variable is the lagged log absenteeism rate. Column (1) uses monthly data where each observation represents an establishment-year-month, while Column (2) uses annual data where each observation represents an establishment-year. Standard errors are clustered at the establishment level and reported in parentheses. P-values are reported in square brackets.

Direct evidence on the unpredictable nature of worker absences. We use our own survey data to shed light on the timing of worker absences. Over 70% of absences are reported at most one day in advance (Appendix Figure C.3). This unpredictable nature of absenteeism leaves establishments with limited scope to effectively smooth labor input over time and consequently, absenteeism, which may help explain why absenteeism rates are far less persistent across time than other firm outcomes such as size or productivity.

3.3 Intertemporal Substitution of Production and Absenteeism

Fact 3*: Establishments appear unable to mitigate absenteeism risk through intertemporal reallocation: the near-zero covariance between monthly scheduled worker-days and absenteeism rates indicates that firms have very limited scope to adjust production schedules across months.*

As shown earlier, absenteeism exhibits a substantial degree of persistence across months, suggesting that establishments may be able to anticipate periods of high absenteeism. If establishments can reliably forecast such patterns, associated productivity losses could, in principle, be mitigated by rescheduling production across the months. To formally assess whether establishments are able to reallocate production in response to seasonal absenteeism, we exploit the following identity:

$$\begin{aligned} A_{ft} &= \sum_m s_{m,ft} A_{m,ft} \\ A_{ft} &= \overline{A_{ft}} + \underbrace{\sum_m \left(s_{m,ft} - \overline{s_{ft}}\right)\left(A_{m,ft} - \overline{A_{ft}}\right)}_{\text{Reallocation Term}} \end{aligned} \tag{1}$$

The first term in the decomposition $\overline{A_{ft}}$ reflects the average absenteeism rate across months, while the second term captures a covariance component that measures how production scheduling correlates with absenteeism. A negative covariance indicates that establishments shift production toward months with lower absenteeism i.e., they reallocate labor inter-temporally to reduce exposure to absenteeism risk. In contrast, a positive covariance suggests that establishments allocate more labor during high-absenteeism months, potentially as a hedge. This pattern is consistent with rigid production systems, such as Just-in-Time techniques, where deviations from the production schedule are costly.

We plot the distribution of the reallocation term in Figure 4. On average, establishments exhibit negligible reallocation of worker-days across months, indicating limited responsiveness to fluctuations in absenteeism. However, there is considerable heterogeneity. Nearly half of the establishments display a positive reallocation term, suggesting that more labor is allocated during months with higher absenteeism—consistent with the view that modern manufacturing establishments operate under rigid production schedules, making deviations costly and difficult to implement.

Conversely, a substantial share of establishments exhibit a negative reallocation term, implying some movement of labor away from months with elevated absenteeism risk. Nonetheless, even the most responsive establishments are only able to reduce absenteeism by approximately 1 percentage point relative to the sample average of 8.62%, highlighting the limited scope for intra-year reallocation as a mitigation strategy.

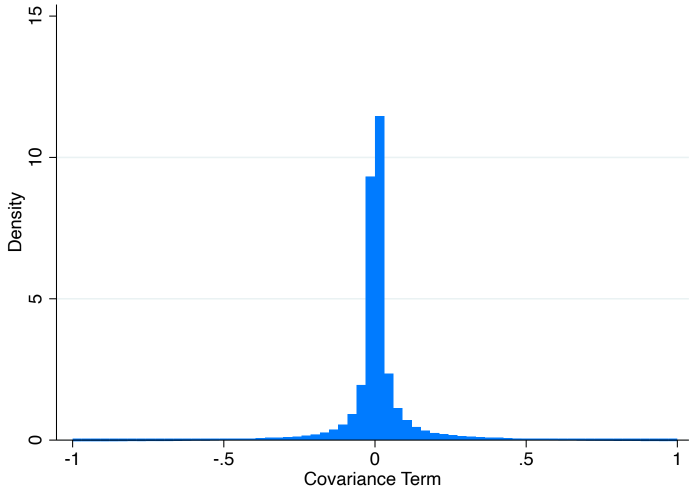
Table 3: Effect of Covariance Mean on Establishment Attributes

	log(Scheduled Workers)	log(VA)	Contract Share (%)	Manager Share (%)
Covariance Mean	-0.146	-0.182	-0.626	0.767
	(0.011)	(0.017)	(0.167)	(0.086)
	[0.000]	[0.000]	[0.000]	[0.000]
Year FE	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Establishment Category FE	Yes	Yes	Yes	Yes
Dep Var Mean	9.086	16.890	12.704	12.251
N: establishment-year	346.840	346.840	346.813	322.165

Note: This table examines the relationship between establishments' responsiveness to absenteeism (as measured by the covariance between absenteeism and worker scheduling) and various establishment attributes. The dependent variables are as follows: natural log of the number of workers who are scheduled to show up for work (column 1), natural log value added (gross sales minus total inputs) (column 2), share of contract workers and managers as a percentage of total employment (columns 3 and 4). Standard errors are clustered at the establishment level and reported in parentheses. P-values are reported in square brackets.

Given heterogeneity in establishments' ability to cope with absenteeism, we con-

Figure 4: Intertemporal Substitution of Scheduled Workers



Note: This figure shows the distribution of within-establishment covariance between worker absenteeism and scheduling proportions for the period 2000 to 2019. The covariance term is calculated as the sum of deviations from establishment-year means: $\sum$(absenteeism $-$ mean_absenteeism) $\times$ (scheduling_proportion $-$ mean_scheduling_proportion). Scheduling proportions represent each unit's share of total scheduled man-days within a establishment-year. Sample is restricted to covariance values between -1 and 1.

duct a regression analysis to examine which establishment characteristics are associated with better reallocation behavior. Table 3 presents results. We find that larger establishments—those that schedule more worker-days annually or produce higher levels of output—are more successful in mitigating absenteeism risk by reallocating worker-days toward months with lower absenteeism. Moreover, establishments that employ a higher share of contract workers appear to manage the schedules of their direct workers more effectively. This may partly reflect underlying differences in establishment size, as larger establishments are more likely to engage contract labor. Finally, drawing on the literature on managerial practices(Bloom and Van Reenen, 2007), we find that establishments with a greater number of managers per worker are also more effective in reallocating labor in response to absenteeism. Together, these findings suggest that establishments do exert effort to adapt to absenteeism risk. However, on average, such strategies yield only modest reductions in absenteeism exposure.

3.4 Labor Turnover and Absenteeism

***Fact 4**: Labor turnover measures —hiring and separation of workers — are positively associated with absenteeism.*

Table 4: Absenteeism, Hiring and Separation

	Hiring rate		Separation rate	
	(1)	(2)	(3)	(4)
Absenteeism Rate	0.016	0.010	0.028	0.018
	(0.001)	(0.001)	(0.001)	(0.001)
	[0.000]	[0.000]	[0.000]	[0.000]
Year-Month FE	Yes	Yes	Yes	Yes
Sector FE	Yes	No	Yes	No
State FE	Yes	No	Yes	No
Establishment FE	No	Yes	No	Yes
R^2	0.050	0.228	0.033	0.216
Dep Var Mean	1.369	1.363	1.168	1.163
N:establishment-month	1,490,510	1,487,181	1,490,510	1,487,181

This table illustrates the relationship between absenteeism and labor churning. The dependent variables are as follows: hiring rate, calculated as the number of accessions in a month divided by number of workers on the first day of the month (columns 1 and 2), and separation rate, calculated as the number of separations (excluding those due to death or retirement) divided by the number of workers on the first day of the month (columns 3 and 4). Standard errors are clustered at the establishment level and reported in parentheses. P-values reported in square brackets.

Absenteeism may indicate a need to hire workers to replace missing manpower. Table 4 investigates the relationship between absenteeism and labor turnover, focusing on hiring and separation rates. Columns 1 and 2 report regressions where the dependent variable is the hiring rate, while columns 3 and 4 focus on the separation rate. Across all specifications, absenteeism is positively and significantly associated with both hiring and separation, suggesting that establishments respond to higher absenteeism with greater labor turnover.

In columns 2 and 4, which include year-month and establishment fixed effects, a 1 percentage point increase in the absenteeism rate is associated with a 0.010 percentage point increase in the hiring rate and a 0.018 percentage point increase in the separation rate.[12] These effects are economically meaningful given the average hiring and separation rates (roughly 8–9%).

Overall, results indicate that absenteeism is strongly correlated with both increased hiring and separation activity, pointing to a high degree of labor churning in response to worker absence. This suggests that absenteeism imposes not just productivity costs but also disruptive personnel costs, as establishments are forced to continually adjust their workforce to maintain operations.

3.5 Establishments' Output and Absenteeism

***Fact 5**: Establishments with higher than mean absenteeism rates experience lower output, with an estimated output elasticity of –0.016. The impact is more pronounced in sectors with high co-worker complementarities–where the absence of one worker disrupts others–yielding a higher output elasticity of –0.021.*

We find that absenteeism is strongly associated with lower levels of output in Table 5.[13] In our preferred specification with establishment and year fixed effects, we find that a 1% increase in absenteeism is associated with a 0.016% decline in value added (column 2). This estimate imply that a doubling of the absenteeism rate within a firm is associated with a 1.6% decline in output. From Table 1, 12% of establishments experience year-to-year swings in absenteeism as large as their own mean, highlighting the relevance of these elasticities.

Appendix Table B.5 shows similar results when using sales instead of value added as the output measure. Taken together, these findings indicate that absenteeism imposes substantial productivity costs on establishments.

We exploit sectoral heterogeneity in co-worker complementarities to examine how the impact of absenteeism varies across production environments. We hypothesize

[12] We find similar results when using yearly instead of monthly data (Appendix Table B.3, columns 2 and 4. As expected, the magnitudes are much larger, 1 p.p. increase in the absenteeism rate leads to a 0.50 p.p. increase in the hiring rate and 0.68 p.p increase in the separation rate.

[13] Note that for facts 5 and 6, we can only use yearly data as output data is not available at the monthly frequency.

Table 5: Absenteeism, Output and Co-worker Complementarities

	log(VA) (1)	log(VA) (2)	log(VA) (3)	log(VA) (4)
log(Absenteeism Rate)	-0.152	-0.016	-0.047	-0.010
	(0.007)	(0.003)	(0.009)	(0.004)
	[0.000]	[0.000]	[0.000]	[0.009]
ONET Composite Index=1 × log(Absenteeism Rate)			-0.104	-0.011
			(0.008)	(0.005)
			[0.000]	[0.026]
Year FE	Yes	Yes	Yes	Yes
Sector FE	Yes	No	No	No
State FE	Yes	No	No	No
Establishment FE	No	Yes	No	Yes
Dep Var Mean	16.903	17.198	16.903	17.198
N:establishment-year	342,589	282,505	342,589	282,505

Note: This table illustrates the relationship between absenteeism, co-worker complementarities, and establishment output. The dependent variable is the natural log of mean monthly gross values less mean monthly total inputs. The O*NET composite index is derived by standardizing and taking the average of O*NET scores from different dimensions (teamwork, communication, contact and impact). This index is then transformed into a binary variable, using the median as a cutoff. Standard errors are clustered at the establishment level and reported in parentheses. P-values are reported in square brackets.

that in sectors where production relies more heavily on team-based processes, shocks to absenteeism are more damaging to firm productivity. In such settings, the productivity of one worker depends critically on the presence and performance of others–for example, in assembly-line production, where a bottleneck at a single stage can disrupt the entire workflow. To capture this, we construct an index of co-worker complementarities based on occupational characteristics that reflect the degree to which a task requires teamwork, communication, and interpersonal contact. We interact absenteeism with this measure to test whether the productivity effects of absenteeism are amplified in sectors with higher co-worker complementarities.

Establishments operating in high-complementarity environments experience significantly larger output losses from absenteeism (column 4). The interaction term is negative and statistically significant, indicating that the productivity cost of absenteeism increases with the degree of co-worker dependency. In sectors with high complementarities, the implied elasticity of absenteeism on output rises to -0.021, underscoring amplified disruptions in team-based production settings. In Appendix Table B.4, we examine individual dimensions of co-worker complementarities and find that sectors relying most on worker-to-worker communication are the most affected.

3.6 Wages of Workers and Absenteeism

***Fact 6**: The wage elasticity of absenteeism (for direct workers) is 0.02 within an establishment. Establishments exposed to higher absenteeism are associated with a higher share of contract workers–with an elasticity of 0.355%.*

Table 6: Absenteeism and Wages

	log(Direct Employee Wage) (1)	Contract Employee Share (2)	log(Contract Employee Wage) (3)
log(Absenteeism Rate)	0.022	0.355	0.003
	(0.001)	(0.050)	(0.002)
	[0.000]	[0.000]	[0.266]
Year FE	Yes	Yes	Yes
Establishment FE	Yes	Yes	Yes
Dep Var Mean	15.717	12.815	15.716
N: establishment-year	341,557	341,937	95,149

Note: This table illustrates the relationship between absenteeism and wages. The dependent variables are as follows: log wages per worker for those under direct employment (column 1), percentage of workers under contract employment (column 2), and log wages per worker for those under contract employment (column 3). Standard errors are clustered at the establishment level and reported in parentheses. P-values are reported in square brackets.

Table 6 explores the relationship between absenteeism rates and establishment-level wage outcomes, using establishment-year level data. Column (1) shows that higher absenteeism is associated with higher wages for directly employed workers: a 1% increase in absenteeism is linked to a 0.022% increase in direct worker wages, statistically significant at the 1% level. This is consistent with the interpretation that establishments may compensate higher absenteeism risk with higher wages to retain workers.

A positive wage elasticity of absenteeism helps rule out certain underlying mechanisms that might otherwise explain worker absence. In particular, a firm demand-side explanation, grounded in efficiency wage theory, posits that firms optimally choose absenteeism levels by offering wages that incentivize regular attendance. Under this view, more modern firms—especially those that rely on Smithian specialization, where each worker performs a narrowly defined task within a tightly coordinated production process—would offer higher wages in exchange for lower absenteeism. This mechanism implies a negative correlation between wages and absenteeism.

However, our finding of a positive relationship between wages and absenteeism points to a different interpretation: absenteeism appears to be primarily driven by labor supply shocks faced by workers. These include unanticipated personal or social obligations, seasonal employment opportunities in agriculture, and illness. Evidence from our establishment survey supports this view: around 65% of reported absences are attributed to these three reasons (see Appendix Figures C.2a–C.2b). In such cases, firms may raise wages simply to retain workers on the factory floor. This explanation is more consistent with absenteeism being a worker-side constraint rather than a strategic decision by firms.

Next, we examine how absenteeism among direct (permanent) workers influences a firm's reliance on contract labor. Given the observed positive correlation between wages and absenteeism for direct workers, we hypothesize that firms may respond by increasing the share of contract workers, who provide a more flexible and often lower-cost alternative. However, due to imperfect substitutability between direct and contract workers–especially in tasks requiring firm-specific skills–this substitution may not fully offset the productivity losses caused by absenteeism.

Results in column 2 support this hypothesis: a 1% increase in absenteeism is associated with a 0.00355 percentage point increase in the share of contract workers, suggesting that firms adjust their workforce composition in response to absenteeism shocks. However, there is no statistically significant relationship between absenteeism and the wages of contract workers (column 3), indicating that while absenteeism influences hiring decisions, it does not significantly affect wage-setting for contract workers. Taken together, these findings imply that firms adapt to absenteeism through a combination of higher wages for direct workers and greater use of flexible labor arrangements, though such strategies may only partially mitigate the underlying productivity disruptions.

Taking Stock. So far, our analysis shows that absenteeism in the Indian manufacturing sector is widespread and unevenly distributed across regions, sectors, and firms. While part of this variation reflects persistent, establishment-specific factors, much of it stems from unpredictable, time-varying shocks. Firms have limited ability to adjust production schedules to offset these risks, and absenteeism leads to sizable productivity losses–especially in sectors where workers' output depends heavily on their co-workers. In response, firms raise wages for direct workers and increase reliance on contract labor, but these adjustments only partially mitigate the costs. The macroeconomic implications, however, cannot be fully inferred from this reduced-form evidence. We therefore turn to a quantitative model to assess the broader economic impact of absenteeism.

4 Model of firm dynamics

This section provides a theory that rationalizes the facts on worker absenteeism, wages, and firm productivity and allows us to study the aggregate implications of worker absenteeism.

4.1 Production Technology and Marginal Product of Labor

The firm's endogenous state variable is N_t, the number of incumbent workers at the beginning of period t. We also include an exogenous *absenteeism state* $\xi \in \mathcal{E}$, where $1 - exp(\xi)$ is the fraction of workers who are *absent* in period t. We model it as a Markov process $\xi_t = \mu_a(1-\rho_a) + \rho_a\xi_{t-1} + \sigma_a\varepsilon_t$, where ρ_a is the persistant parameter and $\varepsilon_t \sim N(0,1)$ is an iid shock and σ_a is standard deviation. Thus, a part of the absenteeism state is predicted by the firms, which we label as *anticipated absenteeism*. We assume the long-run mean to be zero $\mu_a = 0$.

The firm produces output using a Cobb-Douglas production function with labor as the only input and returns to scale $\alpha < 1$. Let z denote total factor productivity, which also follows a Markov process $z_t = \rho_z z_{t-1} + \sigma_z \varepsilon_t^z$, where ρ_z is the persistant parameter and $\varepsilon_t^z \sim N(0,1)$ is an iid shock and σ_z is standard deviation.[14] We also assume that there is an exogenous separation rate ρ_t between firms and workers. Given N_t workers at the start of the period and absenteeism shock ξ_t, the effective incumbent workers are $exp(\xi)(1-\rho_t)N_t$.

After observing their absenteeism and productivity state, firms may hire outside workers H_t. However, hiring is frictional, which makes incumbent and newly hired workers imperfectly substitutable, thus generating tangible effects of absenteeism on firm value. We define output as

$$Y_t \;=\; exp(z_t)\,(\underbrace{exp(\xi_t)(1-\rho_t)N_t + H_t}_{\text{Effective Labor } L_t})^{\alpha}, \tag{2}$$

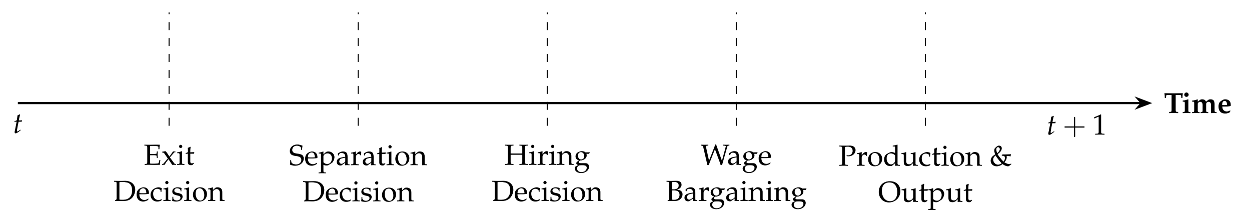
where $0 < \alpha < 1$ ensures diminishing returns to labor (e.g., reflecting a fixed capital or land input) and $(1-\rho_t)$ is the worker retention rate. This implies the marginal revenue product of effective labor is

$$F_L(N_t) = \text{MPL}_t \;=\; \frac{\partial Y_t}{\partial L_t} \;=\; exp(z_t)\,\alpha\, L_t^{\alpha-1}$$

which decreases as L_t rises. In a high-absence state, the marginal product of each present worker rises, because fewer active workers are sharing the workload. In other words, when some workers are absent, each person who does show up becomes more essential to production. This feature has critical implications for wages and hiring, as we discuss next.

[14]It is straightforward to allow some correlation between z and ξ, however, for simplicity, we starting with no correlation.

Figure 5: Timeline of Events Within a Period



Note. The figure illustrates the sequence of events within a period from t to $t+1$. First, wage bargaining occurs, followed by production and output. Subsequently, firms make separation decisions, and finally, hiring decisions are made. The bottom arrow depicts the transition to the next period, where the resulting state variables N_{t+1} (employment) and s_{t+1} (state) are realized.

4.2 Wage Bargaining with Replacement Costs

Our baseline model adopts a simple wage-setting protocol in which wages are determined entirely ex post—that is, after all hiring, and separation decisions have been completed—through a bargaining process between the firm and its workers. When there are hiring frictions–such as recruitment costs or the loss of output when a worker quits–firms earn quasi-rents. Workers can appropriate a share of these rents through bargaining. As a result, wages tend to increase with firm productivity, but not one-for-one, since threat points–such as a worker's outside option–limit the share of surplus that workers can extract.

The sequence of events within a period is as follows. At the start of the period, firms decide whether to stay in the market or exit. Then, productivity and worker absenteeism are realized. Then the worker separation takes place. Based on this information, the firm makes hiring decisions. Once these are finalized, a bargaining stage begins in which the firm negotiates wages with its workforce. After wages are determined, production occurs, agreed wages are paid to present workers, and the period concludes. See Figure 5 for a visual summary of this timing protocol.

The wage bargaining stage proceeds as follows. The firm bargains bilaterally with each worker over the flow wage for the current period, following the protocol proposed by Brügemann, Gautier and Menzio (2019) and used in Elsby and Gottfries (2022). The firm sequentially negotiates with each worker under the possibility of a temporary breakdown in negotiations. The bargaining game is constructed such that each worker is strategically symmetric. Brügemann, Gautier and Menzio (2019) characterize an equilibrium in which all workers within the firm receive the same wage. This wage coincides with that implied by the marginal surplus-sharing rule of Stole and Zwiebel (1996).

The relevant surplus being bargained over is the marginal flow surplus, reflecting the credible threats each party can issue. In this environment, the threat of a permanent breakdown in negotiations is not credible, since the firm and worker will wish to resume bargaining in future periods. Instead, disagreement is associated with a temporary disruption in production. Because wages are renegotiated every period, future turnover and wages are independent of the currently agreed wage. The flow surplus barganing rule takes the following form:

$$\eta \left[MPL_t - w - w_L L + \kappa_t \right] \; = \; (1-\eta) \left[w - b \right], \tag{3}$$

where, κ_t is the replacement cost, and and w_L is the marginal impact of additional worker on wages.

Proposition 1: The equilibrium wage take the following form

$$w_t = (1-\eta)\, b + \eta\, \kappa_t + \frac{\eta}{1-\eta(1-\alpha)} MPL(N_t, \xi_t, z_t). \tag{4}$$

The resulting wage equation reflects several familiar forces: it is increasing in the worker's outside option, the firm's cost of replacing a worker, and the marginal product of labor. Consistent with existing literature, more productive firms tend to pay higher wages, all else equal.

This approach to wage determination offers several advantages. First, it aligns closely with established models in the literature. Second, because bargaining is limited to the current period's flow wage—which is renegotiated in each period—past wages do not influence future wages. This substantially reduces the number of state variables the firm must track, simplifying the firm's dynamic problem.

Corollary 1: *Wages are higher in high absenteeism states*. In periods of high absenteeism, the marginal product of each present worker increases–since the absence of others makes the contribution of those who remain more critical. As a result, the surplus $\Delta\mathcal{V}$ that each incumbent worker can claim through bargaining is larger.

Intuitively, an incumbent who shows up during a period of widespread absence becomes essential to maintaining operations, thereby gaining additional bargaining leverage and securing a higher wage for that period. This aligns with empirical findings that the loss of hard-to-replace coworkers increases firms' willingness to pay the remaining workers.

Importantly, this mechanism implies that wages are pro-cyclical with respect to absenteeism shocks: a rise in absenteeism–despite representing a negative labor productivity shock–leads to higher wages for the remaining workers. This is not paradoxical once we recognize that the shock creates scarcity in effective labor. The firm effectively "shares the pain" by awarding a premium to those who continue working, because their labor becomes indispensable. Empirical evidence supports this prediction. For example, Jäger, Heining and Lazarus (2022) documents that after an unexpected loss of coworkers, surviving workers experience wage gains on average. Similarly, during positive surplus shocks as in Kline, Petkova, Williams and Zidar (2019), incumbents capture a portion of the expanding surplus. Our model treats absenteeism in a comparable way–as a labor supply shock that increases the wage of incumbents.

4.3 Firm's Dynamic Problem

Employment Law of Motion. We suppress the time subscript from here onwards. The firm's workforce evolves according to:

$$N' = (1 - \rho)N + H, \tag{5}$$

where ρ is the exogenous separation rate. The term $(1 - \rho)N$ represents the survivors (incumbents who remain employed into $t + 1$), and H are the new hires. No explicit endogenous separation–layoffs or quits related to w–is included. The motivation for this assumption comes from high firing costs that firms face in India due to extensive labor laws, Besley and Burgess (2004) and Bertrand, Hsieh and Tsivanidis (2024). However, we do allow costless firing when firms exit the market, which will become clearer when we present the dynamic problem. Allowing flexible firing costs both for incumbents and exiters is relatively straightforward.

Let $V(N, \xi)$ be the firm's value (maximized the expected present value of profits) at the start of period t, given it has N workers, the absenteeism state is ξ, and productivity z. The firm chooses how many workers to hire $H \geq 0$ in period t. Workers hired in t will incur the cost c_h. If absent workers are paid some sick-leave benefits $\underline{b}$, then the wagebill is:

$$W = \underbrace{w * (exp(\xi)(1 - \rho)N \quad + H)}_{\text{wages paid to on-site workers}} + \underbrace{\underline{b} * (1 - \rho) * N}_{\text{wages paid to off-site workers}} \quad . \tag{6}$$

If we had assumed the firm pays all N_t workers regardless of attendance, then the wage bill would be $w((1 - \rho_t)N + H)$, and if we assume $\underline{b} = 0$, then wagebill only consists of on-site workers. These two scenarios represent the extreme cases. In the latter scenario, a firm has some cost relief when workers are absent, though it still loses their output. Given these components, the firm's Bellman equation is:

$$V(N,\xi,z) = \max_{w,H\geq 0}\Big\{ \underbrace{p\cdot exp(z)(\exp(\xi)(1-\rho)N+H)^{\alpha} \;-\; W - p\frac{\kappa}{2}\Big(\frac{H}{N}\Big)^2 N - p\cdot cf}_{\text{current period profits}} \\ + \underbrace{\beta\max\Big(\int V(N',\xi',z')dF(\xi',z'|\xi,z),0\Big)}_{\text{continuation value}}\Big\}, \tag{7}$$

where $\beta \in (0,1)$ is the firm's discount factor and p is the price of the output. The expectation is taken over the next period's exogenous states ξ' and z'(since N' is determined by (5) once H_t is chosen). The firm anticipates future absenteeism shocks in determining the value of having a certain workforce N'. Importantly, the firm takes as given that if it has N_t workers, it must pay them w_t each as determined by bargaining. Higher N_t generally means a lower marginal product and thus a slightly lower wage per worker (due to diminishing returns), but since b provides a floor, wages may not drop too far. In making hiring decisions, the firm anticipates how future wages w' will adjust with employment N'. The term $-\frac{\kappa}{2}(H_t/N_t)^2 N_t$ is the hiring cost, that increase if H_t is a large fraction of N_t. We capture these costs as a disruption in the production process, while the newly hired workers are trained or allocated within the firm.

The continuation value takes into account that if the future value of production is too low, then firms prefer to exit the market. Therefore, the decision to exit is made after the current period's production and before the next period's information set is revealed. Here, the value of exit is assumed to be zero, thus, there are no layoff costs for the firm.

Optimality Conditions. The solution of the Bellman problem in Equation (7) yields a hiring policy $H = H^*(N,\xi,z)$, an implied employment path is $N' = n(N,\xi,z)$ and an exit policy $\chi(N,\xi,z)$. We can characterize the optimality conditions as follows.The following first-order condition (FOC) for hiring holds (assuming an interior solution for H^*):

$$\kappa \frac{H}{N} = MPL - w - w_L * L + \beta \mathbb{E}\Big[V_N\big(N', \xi', z'\big)\Big], \tag{8}$$

The left side is the marginal cost of hiring one more worker. The right side is the expected marginal benefit: the increase in firm value next period from having one extra worker. Equation (8) thus equates marginal hiring cost to the discounted marginal value of labor. This is the Euler-like condition governing optimal employment dynamics.

Next, we can characterize $V_N(N, \xi)$ using the *envelope condition*. Because $N' = (1-\rho)N + H^*$, we have $\partial N'/\partial N = (1-\rho) + (\partial H^*/\partial N)$. However, by the envelope theorem, we can ignore the $\partial H^*/\partial N$ term (since H^* was chosen optimally). Intuitively, the direct effect of N on continuation value is through retention of incumbents: an extra worker today leads to $(1-\rho)$ extra worker (in expectation) next period, since a fraction ρ would leave. Thus, $\partial N'/\partial N \approx (1-\rho)$ in the envelope condition. Using this, we have:

$$V_N(N, \xi, z) = \underbrace{exp(\xi)(1-\rho)\Big(MPL - w - w_L * L\Big)}_{\text{current-period marginal profit}} + \beta\,(1-\rho)\,\mathbb{E}\big[V_N(N', \xi', z')\big]. \tag{9}$$

Proposition 2: The optimal replacement rate is defined by equation that states that the marginal value of an extra worker today equals the immediate increase in profit from having that worker in the current period (additional output minus wage cost, if the worker is present) plus the expected continuation value if that worker stays into next period. The continuation value is discounted and adjusted by the probability $(1-\rho)$ that the worker is still with the firm next period (i.e. not separated at end of t).

$$\kappa \frac{H_t}{N_t} \approx V_N(N, \xi, z)/(1-\rho) \tag{10}$$

Equations (8) and (10) together determine the optimal hiring policy and the marginal value of labor. They mirror the job creation condition in search models: in steady state, the marginal cost of hiring equals the present value of a filled job.

4.3.1 Firm Entry

We assume there is a mass M of potential entrants that are ex-ante identical. An entrant firm must pay the entry cost $c_e > 0$ to set-up the plant and draw $G(\xi, z)$. We assume that new entrants start with a $\underline{N}$ of workers, that we will discipline by matching the size of entrants.

$$\beta \int V(\underline{N}, \xi, z) dG(\xi, z) \leq p \cdot c_e \tag{11}$$

In equilibrium, the free entry condition in Equation (11) holds such that the value of entry is equal to the sunk cost of entry c_e.

4.4 Aggregation

Distribution of the firm. Let $\psi\left(N', \xi', z' \mid N, \xi, z\right)$ denote the transition from (N, ξ, z) to (N', ξ', z'). Using the law of motion for employment in Equation (5) and hiring policy $H(N, \xi, z)$, we can write next period employment policy as $n(N, \xi, z)$. Stationary distribution $\mu(N', \xi', z')$ then solves linear system of the form:

$$\begin{aligned}
\psi\left(N', \xi', z' \mid N, \xi, z\right) &= F\left(\xi', z' \mid \xi, z\right) \mathbb{1}\left[N' = n(N, \xi, z; p)\right] \mathbb{1}[\chi(N, \xi, z; p) = 0] \\
\mu\left(N', \xi', z'\right) &= \int \psi\left(N', \xi', z' \mid N, \xi, z\right) d\mu(N, \xi, z) + M \int dG\left(\xi', z'\right) \mathbb{1}\left[N' = \underline{N}\right]
\end{aligned} \tag{12}$$

Total output is given by

$$\begin{aligned}
Y(\mu, M; p) = & \int \left[f(L(N, \xi, z; p), \xi, z) - c_f - v_h\right] d\mu(N, \xi, z) \\
& + \quad M \int \left[f(L(\underline{N}, \xi, z; p), \xi, z) - c_f - v_h\right] dG(\xi, z),
\end{aligned} \tag{13}$$

where v_h are the hiring costs. One can integrate $\nu_h(\mu, M, p)$ to get the aggregate hiring costs. We define $\mathcal{B}(\mu, M; p)$ as total sick-leave benefits paid to the absent workers. The labor demand and wagebill are given y

$$
\begin{aligned}
L^d(\mu, M; p) &= \int n(N, \xi, z; p) d\mu(N, \xi, z) + M \int n(\underline{N}, \xi, z; p) dG(\xi, z) \\
\mathcal{W}(\mu, M; p) &= \int w(N, \xi, z; p) L(\underline{N}, \xi, z; p) d\mu(N, \xi, z) + \int w(\underline{N}, \xi, z; p) L(\underline{N}, \xi, z; p) dG(\xi, z) \\
&\quad + B(\mu, M; p),
\end{aligned} \tag{14}
$$

and the aggregate profits are given by

$$
\Pi(\mu, M; p) = pY(\mu, M; p) - \mathcal{W}(\mu, M; p) - \mathcal{B}(\mu, M; p) - Mpc_e \tag{15}
$$

4.5 Households

A representative household is endowed with an L^s amount of labor. The household consists of many workers who are employed by firms at different wage rates. For simplicity, we assume that the labor supply is fully inelastic and the *real income* is defined as:

$$
\mathcal{C} = \frac{\mathcal{W} + \mathcal{B} + \Pi}{p} \tag{16}
$$

The household budget constraint is satisfied every period.

4.5.1 Discussion on the microfoundations of worker allocation across firms

So far, we have assumed that there exists a degenerate distribution of wages, and all workers can not move to the highest-paying firm. Here, we provide a microfoundations that can sustain such an equilibrium. We assume that a representative household consists of many heterogeneous workers v who choose to work for a firm f drawn from the distribution as described above. Their indirect utility function is:

$$
u_f(v) = \frac{\varphi_f \, \epsilon_f(v) \, w_f}{p}, \tag{17}
$$

where w_f is the wage, p is the consumption goods price index, φ_f captures amenities offered by firms, and $\epsilon_f(v)$ is an idiosyncratic amenity draw that is specific to each worker v and firm f. Idiosyncratic amenities are drawn independently for each worker and firm from the Fréchet distribution, $F(e) = \exp\left(-e^{-\kappa}\right),$. These amenities shocks can be interpreted as search frictions as well. The shape parameter $\kappa > 1$ regulates the dispersion of idiosyncratic amenities and determines the wage elasticity that captures the responsiveness of worker allocation to firm-level wages. Using the properties of the Fréchet distribution, the probability a worker chooses to worker in a firm f is given by $\ell_f = \frac{(\varphi_f w_f)^{\kappa}}{\sum_s (\varphi_s w_s)^{\kappa}}$. This is also the share of workers allocated to firm f. One can also include unemployment with the outside option b. Therefore, we can choose a distribution of $\varphi_f \in \Phi$ that will generate an allocation of workers that coincides with the stationary distribution of employment and wages mentioned above.

4.6 Stationary Competitive Equilibrium

A stationary equilibrium consists of an output price $p^* \geq 0$, a mass of entrants $M^* \geq 0$, and a measure of incumbents, μ^*, such that: a) wages are set by bargaining protocol as specified in Equation 4; b) firms maximize lifetime discounted values as in Equation (7); c) The labor market clear: $L^d\left(\mu^*, M^*, p^*\right) = L^s$ and goods market clear: $pC = pY - pM_e c_e$; d) there is a stationary distribution of firms: $T\left(\mu^*, M^*, p^*\right) = \mu^*$; and e) free entry condition holds as in Equation (11).

5 Quantitative Analysis

5.1 Calibration

The model requires us to provide values for fourteen parameters. We divide the model parameters into two major groups: Externally and internally calibrated parameters. The values for externally calibrated parameters are either normalised to one or taken from the literature due to difficulties in calibrating them with existing data. Meanwhile, the fitted parameters are calibrated by matching certain moments in the data to their counterparts in the model. Next, we illustrate the fixed parametric values.

Externally Calibrated Parameters. All parameter values are reported in Panel B of Table 7. We set the returns-to-scale parameter to $\alpha = 0.85$ following Midrigan and Xu (2014), and the discount factor to $\beta = 0.92$ to match the average interest rate of 8.69% in India. The worker's bargaining power is calibrated to $\eta = 0.447$, consistent with Elsby and Michaels (2013). The worker's outside option and the wage received during absence are normalized to one, though we conduct robustness checks allowing for alternative values of the wage paid to absent workers.

Internally Calibrated Parameters. All parameter values are reported in Panel A of Table 7. There are three sets of parameters that needs to be internally calibrated: a) firm dynamics, b) absenteeism and c) productivity. All these parameters are internally calibrated (jointly) by matching specific moments as mentioned below.

Targeted Moments We begin by calibrating the parameters that directly shape the employment distribution across firms. The fixed cost parameter c_f is chosen to match the employment share of the bottom 50 percent of firms. In the model, higher fixed costs raise the share of output produced by smaller firms, while greater productivity persistence ρ_z and dispersion σ_z fatten the right tail of the productivity distribution, thereby increasing the employment share of the largest firms. Accordingly, we calibrate ρ_z and σ_z to match the employment shares of the top 5 and top 10 percent of firms.

We next turn to parameters governing firm entry and worker reallocation. The size of new entrants is set to match the relative scale of entrants to incumbents, and the entry cost is calibrated to reproduce the entry rate of 8.8% reported in Hsieh and Klenow (2014). Hiring costs are chosen to match the observed hiring rate of 24% in the data, while the separation rate–exogenous in the model–is directly set to 20%.[15] Finally, the absenteeism parameters, persistence ρ_a and dispersion σ_a, are calibrated to

[15] Yearly hiring and separation rates are computed by multiplying the corresponding monthly rates by 12; see Appendix Table B.2.

match the mean and standard deviation of the absenteeism rate in the data.

Table 7: Fitted Model Parameters and Targeted Moments

Parameter	Value	Description	Moment	Model	Data
Panel A. Internal Calibration					
Firm Dynamics					
c_f	3.1	Fixed cost	Employee share: Bottom 50%	15.2%	5.4%
c_e	13.4	Entry cost	Entry Rate	8.8%	8.8%
n_e	5	Entrant size	Relative entrant size	0.50	0.43
κ	13.9	Hiring costs	Hiring Rate	30%	24%
ρ	0.15	Separation rate	Separation Rate	20%	20%
Absenteeism					
ρ_a	0.45	Persistence in Absenteeism	Mean Absenteeism	8.5%	8.5%
σ_a	0.56	Dispersion in Absenteeism	S.D. Absenteeism	6.8%	7.3%
Productivity					
ρ_z	0.85	Persistence in Productivity	Employee share: Top 5%	52%	65%
σ_z	0.29	Dispersion in Productivity	Employee share: Top 10%	42%	78%
Panel B. External Calibration					
α	0.85	Returns to scale			
β	0.92	Discount rate			
η	0.44	Worker bargaining power			
$\underline{w_o}$	1	Worker outside option			
b_{sick}	$1 \times \underline{w_o}$	Wage for absent workers			

Notes. This table reports the calibrated model parameters and the targeted empirical moments. Panel A (Internal Calibration) lists parameters chosen to match moments in the data. Panel B (External Calibration) shows parameters taken from the literature: returns to scale α from Midrigan and Xu (2014), discount factor β set to match an average interest rate of 8.7 percent in India and bargaining power η from Elsby and Michaels (2013).

5.2 Results

Impact of Absenteeism on Productivity. The calibrated model is used to quantify the aggregate costs of worker absenteeism. We conduct a series of counterfactuals by varying the parameter μ_a in the Markov process, which governs the long-run level of absenteeism in the economy. Because absenteeism rates are assumed to follow a log-normal distribution, changes in μ_a affect not only the mean but also the dispersion of absenteeism rates faced by firms, thereby altering the level of risk faced by firms. We choose values of μ_a that span the range of average absenteeism rates observed in the data, roughly 3–15 percent, see Panel A Table 8. All outcome variables are normalized to the baseline economy.

Table 8: Aggregate Outcomes and Misallocation Measures

$exp(\mu_a)$	36.8	100	182
	(1)	(2)	(3)
Panel A. Implied Absenteeism Rate			
Mean	3.1%	8.6%	15.3%
S.D.	2.7%	7.5%	13.7%
Panel B. Aggregate Outcomes			
Real Consumption	106.1	100.0	92.0
Total Non-absent Workers	104.3	100.0	94.4
Aggregate Productivity	101.7	100.0	97.4
Price of Output	95.5	100.0	106.8
Average Nominal Wage	97.5	100.0	103.9
Average Real Wage	102.0	100.0	97.3
Hiring cost (% of Output)	93.0	100.0	110.0
Panel C. Misallocation Measures			
Variance (Employment Growth)	102.0	100.0	99.3
Variance (Wages)	91.3	100.0	115.8
Variance (Marginal Product of Labor)	100.1	100.0	101.6

Notes. This table reports model-implied aggregate outcomes and misallocation measures under different levels of absenteeism risk, parameterized by $\exp(\mu_a)$. All values are normalized to the baseline case in Column (2). Panel A reports the implied mean and standard deviation of absenteeism rates. Panel B shows aggregate outcomes including employment, productivity, prices, and wages. Panel C presents measures of misallocation based on the variance of employment growth, wages, and the marginal product of labor.

We define aggregate output as $Y = AL$, where A is aggregate productivity and L is the number of non-absent workers. Doubling the long-run absenteeism rate reduces output by 10 percent (Panel B, Table 8). Most of this decline is a mechanical consequence of having fewer non-absent workers, which is not the main focus of this paper. More importantly, aggregate productivity A also falls by 2.6 percent, accounting for about one-third of the total decline in output. These losses represent a meaningful barrier to the development of India's manufacturing sector.

Lower productivity is reflected in higher prices: the price of output rises by 6.8 percent. Although average nominal wages increase, real wages decline because prices rise more than proportionally. Consequently, overall worker welfare–measured by real consumption—falls.

Absenteeism also raises efficiency costs. Hiring costs faced by firms increase by almost 10 percent relative to the baseline, reflecting production disruptions from replacing absent workers. Together, higher absenteeism risk and hiring costs increase resource misallocation, as indicated by a 2 percent rise in the dispersion of the marginal product of labor. Thus, absenteeism itself is an important driver of misallocation of labor in developing economies.

Absenteeism and wage variation across firms. The rise in the marginal product of labor is driven by greater wage dispersion. Attendance shocks ξ cause effective labor to fluctuate according to $L = e^{\xi}(1 - \rho)N + H$. A negative shock (many absences) reduces L, raising the marginal product of labor since MPL $\propto L^{\alpha-1}$ with $\alpha - 1 < 0$. With fewer effective workers, each becomes more valuable at the margin. Through bargaining, workers capture part of this increase, so firms experiencing more absences end up paying higher wages to the remaining workers than otherwise identical firms without shocks. In short, absenteeism-induced fluctuations in L translate into wage variation across firms. This mechanism can generate wage inequality even when firms have the same technology and hire identical workers.

This implies that wage disparities across firms need not only reflect differences in technology or worker ability; they can also arise from differences in labor utilization rates. Firms with higher absenteeism pay higher wages than those with lower absenteeism. In effect, absenteeism operates like an idiosyncratic productivity shock, providing a novel explanation for firm-specific wage premia observed in the data.

Firm growth and employment dynamics. Finally, another results that links absenteeism and labor misallocation is the reduction in labor reallocation in the economy. We find that the variance of employment growth declines by 0.7 percent in high-absenteeism environments. One explanation is that firms facing frequent absences grow more cautiously, "hoarding" labor and carrying buffer workers to ensure against disruptions. Entrants may also choose to grow slowly if they anticipate the need to hire excess staff. These dynamics imply that absenteeism not only increases misallocation but also dampens the reallocation forces that drive growth.

6 Conclusion

Worker absenteeism is often cited by firms in developing countries as a major constraint, yet systematic evidence on its prevalence, causes, and consequences has been limited. This paper provides new evidence from Indian manufacturing firms, combining a nationally representative panel with a purpose-built survey to document both the scale of absenteeism and the ways firms cope with it. We show that absenteeism is widespread and highly variable, with shocks that are largely unpredictable and often reported only at very short notice. Absenteeism is linked to smaller firm size, higher average wages, and especially large costs in sectors with strong co-worker complementarities. Our survey further highlights how firms attempt to manage these disruptions—most notably by hiring buffer workers and rotating staff across tasks.

We then develop a model of firm dynamics with absenteeism risk to interpret these findings and quantify their macroeconomic implications. The model shows that absenteeism reduces aggregate productivity beyond the mechanical loss of labor input, increases hiring costs, and amplifies misallocation by increasing the dispersion of the marginal product of labor across firms. Moreover, absenteeism introduces an idiosyncratic component to firm-level wages, providing a novel mechanism for wage inequality even among firms with identical technology. Finally, absenteeism dampens firm dynamism, reducing dispersion in employment growth, as firms anticipate the need to carry a buffer labor.

Our findings underscore that absenteeism is more than a human-resource issue: it represents a structural barrier to development. Goraya et al. (2025) illustrates that reducing worker absenteeism is a complicated challenge. Workers are absent not only due to agricultural work and illness—which would require moving workers out of agriculture and investing in public health—but also because of various social obligations that play important roles in sustaining network relationships and community ties. Addressing the latter driver of absenteeism would require society-level coordination—e.g., moving all social functions and religious festivals to the weekends—or reducing worker reliance on network benefits by making large investments into formal insurance systems and public infrastructure. A more promising potential strategy for development could be to find alternative ways of using labor, e.g., innovative work arrangements that can handle absence risk and provide schedule flexibility while still supporting skill formation and stable employment of workers.

References

Adhvaryu, Achyuta, Jean-François Gauthier, Anant Nyshadham, and Jorge Tamayo, "Absenteeism, Productivity, and Relational Contracts Inside the Firm," *Journal of the European Economic Association*, August 2024, *22* (4), 1628–1677.

Amodio, Francesco and Nicolás de Roux, "Measuring Labor Market Power in Developing Countries: Evidence from Colombian Plants," *Journal of Labor Economics*, October 2024, *42* (4), 949–977. Publisher: The University of Chicago Press.

_ , Pamela Medina, and Monica Morlacco, "Labor Market Power, Self-Employment, and Development," *SSRN Electronic Journal*, 2022.

Banerjee, Abhijit and Esther Duflo, "Addressing Absence," *Journal of Economic Perspectives*, February 2006, *20* (1), 117–132.

Barmby, Tim, John Sessions, and John Treble, "Absenteeism, Efficiency Wages and Shirking," *The Scandinavian Journal of Economics*, 1994, *96* (4), 561–566. Publisher: [Wiley, The Scandinavian Journal of Economics].

Bertrand, Marianne, Chang-Tai Hsieh, and Nick Tsivanidis, "Contract Labor and Firm Growth in India," 2024.

Besley, Timothy and Robin Burgess, "Can Labor Regulation Hinder Economic Performance? Evidence from India," *The Quarterly Journal of Economics*, 2004, *119* (1), 91–134. Publisher: Oxford University Press.

Bloom, Nicholas and John Van Reenen, "Measuring and Explaining Management Practices Across Firms and Countries*," *The Quarterly Journal of Economics*, November 2007, *122* (4), 1351–1408.

Breza, Emily and Supreet Kaur, "Labor Markets in Developing Countries," June 2025.

_ , _ , and Yogita Shamdasani, "Labor Rationing," *American Economic Review*, October 2021, *111* (10), 3184–3224.

Brooks, Wyatt J., Joseph P. Kaboski, Yao Amber Li, and Wei Qian, "Exploitation of labor? Classical monopsony power and labor's share," *Journal of Development Economics*, May 2021, *150*, 102627.

Brügemann, Björn, Pieter Gautier, and Guido Menzio, "Intra Firm Bargaining and Shapley Values," *The Review of Economic Studies*, March 2019, *86* (2), 564–592.

Bureau of Labor Statistics, "Absences from work of employed full-time wage and salary workers by occupation and industry," January 2025.

Callen, Michael, Saad Gulzar, Ali Hasanain, Muhammad Yasir Khan, and Arman Rezaee, "The political economy of public sector absence," *Journal of Public Economics*, February 2023, *218*, 104787.

Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, and F. Halsey Rogers, "Missing in Action: Teacher and Health Worker Absence in Developing Countries," *Journal of Economic Perspectives*, March 2006, *20* (1), 91–116.

Coles, Melvyn G. and John G. Treble, "The price of worker reliability," *Economics Letters*, January 1993, *41* (2), 149–155.

_ and _ , "Calculating the price of worker reliability," *Labour Economics*, September 1996, *3* (2), 169–188.

Cosar, A. Kerem, Nezih Guner, and James Tybout, "Firm Dynamics, Job Turnover, and Wage Distributions in an Open Economy," *The American Economic Review*, March 2016, *106* (3), 625–663. Num Pages: 39 Place: Nashville, United States Publisher: American Economic Association.

Cotton, Linda, Vijaya Ramachandran, Chad Leechor, Jean Michel Marchat, John Paton, Manju Kedia Shah, James Habyarimana, Michael Wong, William Kalema, and Charles Ntale, "Competing in the Global Economy: An Investment Climate Assessment for Uganda," Technical Report 14416, The World Bank Group 2004.

Donovan, Kevin, Will Jianyu Lu, and Todd Schoellman, "Labor Market Dynamics and Development," *The Quarterly Journal of Economics*, September 2023, *138* (4), 2287–2325.

Elsby, Michael W L and Axel Gottfries, "Firm Dynamics, On-the-Job Search, and Labor Market Fluctuations," *The Review of Economic Studies*, May 2022, *89* (3), 1370–1419.

Elsby, Michael W. L. and Ryan Michaels, "Marginal Jobs, Heterogeneous Firms, and Unemployment Flows," *American Economic Journal: Macroeconomics*, 2013, *5* (1), 1–48. Publisher: American Economic Association.

Engström, Per and Bertil Holmlund, "Worker Absenteeism in Search Equilibrium," *The Scandinavian Journal of Economics*, 2007, *109* (3), 439–467. Publisher: [Wiley, The Scandinavian Journal of Economics].

Goraya, Sampreet Singh, "How does caste affect entrepreneurship? birth versus worth," *Journal of Monetary Economics*, April 2023, *135*, 116–133.

_ , Suanna Oh, and Yogita Shamdasani, "Absent but Not Idle: the Social Roots of Demand for Flexible Work," *Working paper*, 2025.

Hsieh, Chang-Tai and Peter J. Klenow, "Misallocation and Manufacturing TFP in China and India*," *The Quarterly Journal of Economics*, November 2009, *124* (4), 1403–1448.

_ and _ , "The Life Cycle of Plants in India and Mexico *," *The Quarterly Journal of Economics*, August 2014, *129* (3), 1035–1084.

ILO, *More than 60 per cent of the world's employed population are in the informal economy*, International Labour Organization, April 2018.

James, Ralph C., "The Casual Labor Problem in Indian Manufacturing," *The Quarterly Journal of Economics*, 1960, *74* (1), 100–116. Publisher: Oxford University Press.

Jäger, Simon, Jörg Heining, and Nathan Lazarus, "How Substitutable Are Workers? Evidence from Worker Deaths," November 2022.

Kline, Patrick, Neviana Petkova, Heidi Williams, and Owen Zidar, "Who Profits from Patents? Rent-Sharing at Innovative Firms*," *The Quarterly Journal of Economics*, August 2019, *134* (3), 1343–1404.

Krishnaswamy, Nandita, "Missing and Fired: Worker Absence, Labor Regulation, and Firm Outcomes," 2019.

Martin, Leslie A., Shanthi Nataraj, and Ann E. Harrison, "In with the Big, Out with the Small: Removing Small-Scale Reservations in India," *American Economic Review*, February 2017, *107* (2), 354–386.

Midrigan, Virgiliu and Daniel Yi Xu, "Finance and Misallocation: Evidence from Plant-Level Data," *American Economic Review*, February 2014, *104* (2), 422–458.

Stole, Lars A. and Jeffrey Zwiebel, "Intra-firm Bargaining under Non-binding Contracts," *The Review of Economic Studies*, July 1996, *63* (3), 375–410.

Syverson, Chad, "What Determines Productivity?," *Journal of Economic Literature*, June 2011, *49* (2), 326–365.

Weiss, Andrew, "Absenteeism and wages," *Economics Letters*, January 1985, *19* (3), 277–279. Publisher: Elsevier BV.

World Bank, "The Africa competitiveness report 2017 - Addressing Africa's demographic dividend (English)," in "Africa competitiveness reports," Washington, D.C.: The World Bank Group, 2017.

Zane, Giulia, "Sharing a Workforce: the Effect of Agricultural Productivity Shocks on Industrial Performance," 2018.

Worker Absenteeism and Establishment Outcomes: Evidence from the Indian Manufacturing Sector

Sampreet S. Goraya, Nandita Krishnaswamy, Suanna Oh,

and Yogita Shamdasani

Online Appendix

A ASI Data

A.1 Data Construction and Cleaning

Data Sources. We use three components of the ASI:

1. **Part I** (annual establishment cross-section data, identifier: DSLNo)
2. **Part II** (monthly labor statistics, identifier: DSLNo)
3. **Panel** (annual establishment panel data, identifier: Factory ID)

The challenge is linking these three components (with different establishment identifiers) over time.

Merge Strategy. We merge Factory ID (panel data identifier) with the cross-section data through a two-stage observable matching: (1) *Full Match* using comprehensive variables (temporal, administrative, establishment characteristics, employment, financial, operational indicators), successfully merging most observations; (2) *Relaxed Match* excluding industry codes for remaining unmatched observations. Our approach follows Martin et al. (2017) and additionally uses temporal and operational variables that are better suited for the data post 2008.

Data Harmonization: We standardize variable types, convert industry codes to consistent 1998 NIC standard using concordance tables, and harmonize district codes to account for boundary changes. Duplicates are resolved by keeping open establishments over closed establishments.

Part II Integration: Monthly labor statistics are merged using DSLNo and survey year, creating monthly-level observations for labor analysis.

Data Cleaning. Some establishments that operate for 12 months report only quarterly data (March, June, September, December). We test whether quarterly values represent monthly or cumulative totals by comparing imputed annual totals against actual employment and testing for 2:3:4 cumulative patterns. For established monthly reporters, we propagate quarterly values backward within quarters. Establishments identified as cumulative reporters are excluded to prevent overestimation.

Quality Controls: We exclude several years of data from our analysis due to a combination of data quality issues and structural inconsistencies:

- 1999–2000 because it lacks establishment identifiers (`dslno`), which are essential for linking observations across datasets
- 2005–06 data as they are identical to the 2004–2005 data, with metadata suggesting they are, in fact, duplicates of the 2004–2005 records
- 2009–10 due to the complete absence of Part II data for that year
- 2008-09 and 2010-11 due to concerns about data quality. Both the Part I and Part II datasets report the number of man-days worked. In Part II, this figure reflects total man-days worked, whereas in the cross-section it refers specifically to man-days worked by direct workers. These two values should be closely aligned if the `dslno` consistently identifies the same establishment across datasets. Discrepancies between them may thus indicate issues with identifier consistency. We observed especially low correlation between man-days figures in the two datasets for 2008-09 and 2010-11
- 2011–12 due to a change in the reporting format—from a calendar year to a financial year—which introduced ambiguity in the interpretation of monthly absenteeism data

We winsorize continuous variables at 1st/99th percentiles to handle outliers common in administrative data. Observations with inconsistent accounting (scheduled $\neq$ worked + lost man-days) are removed.

Final Sample. The cleaned dataset spans 2000-2019 with close to 4.8 million establishment -month observations covering major manufacturing states. Cross-validation shows correlations greater than 0.85 between Part I and Part II employment variables for these years.

Table A.1: Sample Counts, by Year

Year	Observations	Percent
2000	284,403	5.94
2001	303,540	6.34
2002	316,132	6.61
2003	417,068	8.71
2004	352,778	7.37
2005	384,431	8.03
2007	173,870	3.63
2012	370,814	7.75
2013	361,029	7.54
2014	382,185	7.99
2016	372,918	7.79
2017	377,167	7.88
2018	356,054	7.44
2019	333,720	6.97
Total	4,786,109	100.00

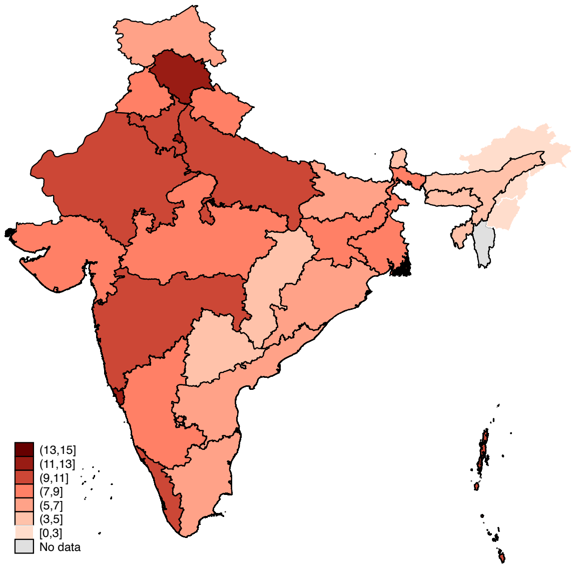
Table A.2: Absenteeism, by State

State	Mean	P5	P25	Median	P75	P90	N: establishment-year
Andaman & N. Island	9.997	0.405	3.550	6.579	14.560	23.026	947
Andhra Pradesh	5.539	0.000	2.793	4.444	7.589	10.153	261,778
Arunachal Pradesh	4.356	0.000	0.000	0.000	3.186	13.043	929
Assam	5.158	0.000	1.282	3.990	6.977	11.000	63,543
Bihar	7.052	0.000	2.564	5.000	8.800	14.286	34,757
Chandigarh	13.929	0.644	5.392	10.621	19.231	30.769	15,036
Chhattisgarh	5.116	0.694	2.222	3.846	6.000	9.597	69,557
Dadra & Nagar Haveli	8.353	1.600	4.567	7.343	10.096	15.296	42,305
Daman & Diu	8.276	1.753	4.916	7.529	10.000	14.000	45,278
Delhi	14.268	2.885	7.692	11.806	17.385	27.600	118,916
Goa	11.089	1.538	5.385	9.660	14.845	20.551	34,392
Gujarat	9.219	1.333	4.334	7.692	11.438	17.778	445,265
Haryana	10.525	1.991	4.972	8.379	13.360	20.192	254,051
Himachal Pradesh	12.520	1.989	6.891	11.111	16.086	22.324	55,076
Jammu & Kashmir	6.980	1.268	3.692	5.674	8.571	12.500	45,001
Jharkhand	7.709	0.746	3.254	6.154	9.959	14.835	53,651
Karnataka	8.237	0.993	4.000	6.667	10.064	15.278	316,414
Kerala	11.732	1.342	5.405	9.057	14.615	23.965	182,414
Madhya Pradesh	8.252	0.741	3.294	6.036	10.235	15.955	144,289
Maharashtra	9.788	1.282	4.724	8.333	12.555	18.343	585,519
Manipur	4.832	0.505	1.578	3.289	6.211	11.364	2,369
Meghalaya	4.434	0.477	1.442	2.693	5.128	9.798	4,630
Nagaland	1.452	0.000	0.000	0.278	1.923	4.000	6,744
Odisha	7.052	0.555	2.778	5.128	9.000	14.835	60,960
Pondicherry	6.044	0.800	3.134	4.912	7.418	11.012	40,194
Punjab	9.464	0.841	3.846	6.923	11.777	20.000	237,238
Rajasthan	10.439	0.503	3.551	7.733	13.636	22.692	172,836
Sikkim	3.988	0.352	1.280	2.772	5.002	8.854	4,194
Tamil Nadu	5.908	1.266	3.333	5.128	7.692	9.714	693,787
Telangana	3.760	0.000	1.538	3.143	4.597	7.933	82,907
Tripura	4.464	0.000	1.818	3.500	5.769	8.696	6,858
Uttar Pradesh	10.166	0.894	4.167	8.333	12.821	18.790	424,634
Uttarakhand	7.623	1.208	4.545	7.077	8.833	12.607	61,673
West Bengal	9.028	0.480	3.762	7.200	11.498	17.816	217,967

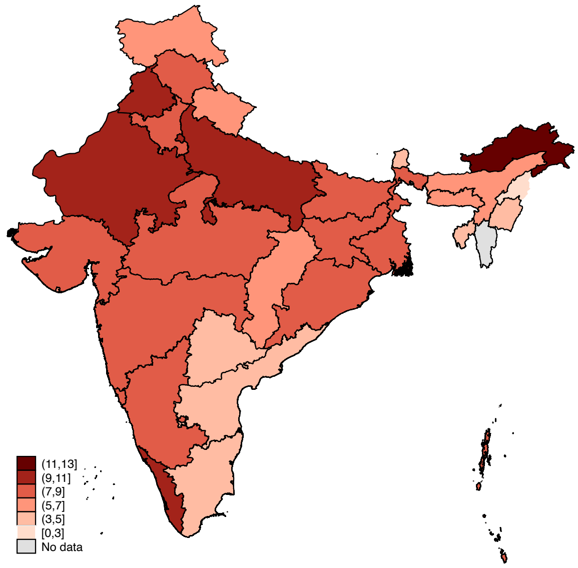
Notes: This table reports the average absenteeism rate, selected percentiles (P5, P25, P50/Median, P75, and P90), and the total number of establishment-year observations for each Indian state and union territory.

B Robustness of Stylized Facts

Figure B.1: Geography of Absenteeism



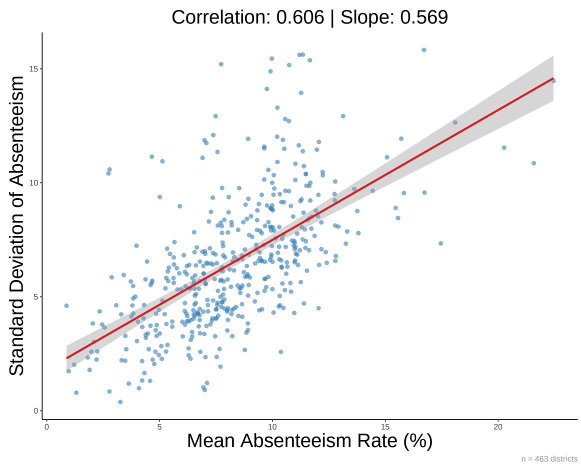
(a) By State (Mean)



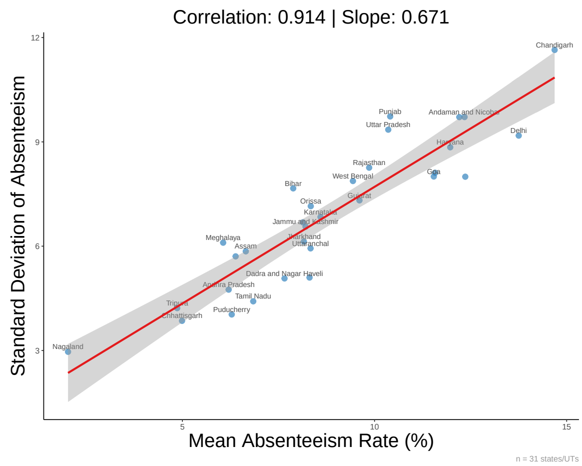
(b) By State (SD)

Note: Figure plots means and standard deviations of monthly absenteeism rates across states in India between 2000–2019.

Figure B.2: Absenteeism Correlation



(a) By Distrct



(b) By State

Note: Figure plots the linear fit relationship between the mean and standard deviation of monthly absenteeism by district (panel a, years 2000–2007) and by state (panel b, years 2000–2019).

Table B.1: Residual Variation in Absenteeism

	σ_e (1)	$\lvert e\rvert > 1$ (2)	$\lvert e\rvert > 2$ (3)	$\lvert e\rvert > 3$ (4)	$\lvert e\rvert > 5$ (5)	$\lvert e\rvert > 8.62$ (6)
No FE	7.337	0.863	0.729	0.610	0.366	0.106
Establishment FE	4.999	0.711	0.500	0.349	0.177	0.066
Establishment FE + Year FE	4.985	0.708	0.494	0.343	0.174	0.065
Establishment FE + Year×State FE	4.953	0.706	0.491	0.341	0.172	0.064
Establishment FE + Year×Sector FE	4.979	0.708	0.495	0.343	0.174	0.065

Note: This table summarizes residual variation in yearly absenteeism from a regression with fixed effects. Each row lists the relevant set of fixed effects included in each regression. Column 1 reports the standard deviation of the residuals (denoted as σ_e). Columns 2 to 6 report the proportion of observations that have a residual e with absolute values exceeding 1, 2, 3, 5 or 8.62% (mean monthly absenteeism rate).

Table B.2: Monthly Hiring and Separation Rates

	N:establishment-month	Mean	St. Dev	Pct(25)	Pct(75)
Hiring rate	1,670,087	2.016	6.338	0.000	0.000
Separation rate	1,673,480	1.744	5.175	0.000	0.074

Note: This table displays summary statistics of hiring and separation rates at the monthly level. Hiring rate is calculated as the number of accessions in a month divided by average number of workers on the first day of the month. Separation rate is calculated as the number of separations (excluding those due to death or retirement) divided by average number of workers on the first day of the month.

Table B.3: Absenteeism, Hiring and Separation (Yearly)

	Hiring rate		Separation rate	
	(1)	(2)	(3)	(4)
Absenteeism Rate	0.785	0.500	1.006	0.683
	(0.022)	(0.026)	(0.022)	(0.027)
	[0.000]	[0.000]	[0.000]	[0.000]
Year FE	Yes	Yes	Yes	Yes
Sector FE	Yes	No	Yes	No
State FE	Yes	No	Yes	No
Establishment FE	No	Yes	No	Yes
R^2	0.077	0.440	0.067	0.416
Dep Var Mean	28.862	28.245	27.822	27.409
N:establishment-year	417,244	345,888	417,244	345,888

Note: This table illustrates the relationship between absenteeism and labor churning. The dependent variables are as follows: hiring rate, calculated as the number of accessions in a year divided by number of workers on the first day of December in each year (columns 1 and 2), and separation rate, calculated as the number of separations (excluding those due to death or retirement) divided by the number of workers on the first day of December in each year (columns 3 and 4). We choose December as a representative month (as opposed to January) to account for firms that report quarterly values (in months 3, 6, 9, 12). Additionally, to account for firms that report quarterly stock values, we impute yearly hiring and separation rates by multiplying the values by 4. Standard errors are clustered at the establishment level and reported in parentheses. P-values reported in square brackets.

Table B.4: Absenteeism, Output and Co-worker Complementarities

	log(VA) (1)	log(VA) (2)	log(VA) (3)	log(VA) (4)	log(VA) (5)
log(Absenteeism Rate)	-0.016	-0.010	-0.015	-0.013	-0.010
	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)
	[0.000]	[0.006]	[0.000]	[0.000]	[0.009]
ONET Communication Index=1 × log(Absenteeism Rate)		-0.012			
		(0.005)			
		[0.016]			
ONET Contact Index=1 × log(Absenteeism Rate)			-0.002		
			(0.005)		
			[0.611]		
ONET Teamwork Index=1 × log(Absenteeism Rate)				-0.005	
				(0.005)	
				[0.289]	
ONET Composite Index=1 × log(Absenteeism Rate)					-0.011
					(0.005)
					[0.026]
Year FE	Yes	Yes	Yes	Yes	Yes
Establishment FE	Yes	Yes	Yes	Yes	Yes
Dep Var Mean	17.198	17.198	17.198	17.198	17.198
N:establishment-year	282,505	282,505	282,505	282,505	282,505

Note: This table illustrates the relationship between absenteeism, co-worker complementarities and establishment output. The dependent variable is the natural log of gross sales less total inputs. The independent variables is derived by standardizing O*NET scores from different dimensions (teamwork, communication, contact and impact). These indices are then transformed into binary variables, using the median as a cutoff. Standard errors are clustered at the establishment level and reported in parentheses. P-values are reported in square brackets.

Table B.5: Absenteeism and Output: Sales vs Value-added

	log(Sales) (1)	log(Sales) (2)	log(VA) (3)	log(VA) (4)
Absenteeism Rate	-0.021	-0.003	-0.023	-0.002
	(0.001)	(0.000)	(0.001)	(0.000)
	[0.000]	[0.000]	[0.000]	[0.000]
Year FE	Yes	Yes	Yes	Yes
Sector FE	Yes	No	Yes	No
State FE	Yes	No	Yes	No
Establishment FE	No	Yes	No	Yes
Dep Var Mean	18	18	17	17
N:establishment-year	347,672	287,000	347,672	287,000

Note: This table illustrates the relationship between absenteeism and establishment output.The dependent variables are log sales (columns 1 and 2) and log value added (columns 3 and 4). Standard errors are clustered at the establishment level and reported in parentheses. P-values are reported in square brackets.

C Establishment Survey

C.1 Sample Selection.

We construct our survey sample frame using the 2018–19 ASI.[16]

Industry Selection. We identify the largest industries in Odisha based on the number of registered establishments. We focus our survey sample on establishments in the top ten industries.[17]

District Selection. We select four districts in Odisha for the survey. Our initial criteria was distance from the state capital, Bhubaneswar, where the field team was based, so we started in two nearby districts, Khurda and Cuttack. Subsequently, we expanded to Ganjam and Kalahandi, approximately 140 km and 390 km away from Bhubaneswar respectively.

Screening. We focus on establishments that have more than 4 unskilled workers engaged in the main production activity. We target establishments that use unskilled workers so that we can "standardize" worker quality across establishments. This allows us to hold fixed worker quality.

Replacement Strategy. We allow for replacement establishments when there is an incomplete survey in the main survey sample roster. This arises when establishments are reluctant to participate (as is common in the case of hand-collected establishment surveys due to concerns over confidentiality etc.) and when establishments have shut down or are temporarily closed.[18] Replacement establishments are similar to establishments in our main list, satisfying three criteria: 1) They are in the same district as our main sample establishment, 2) They are within a 15 kilometer radius of our main sample establishment, 3) They belong to the same industries as the main sample

[16] Names and addresses of establishments were downloaded from the ASI website. The file was subsequently removed from the site at a later time.

[17] The top three industries in Odisha are: (1) manufacture of grain mill products, (2) cutting, shaping, and finishing of stone, and (3) manufacture of basic iron and steel.

[18] This is particularly acute in our survey as the COVID-19 pandemic took place after the ASI sampling frame (2018–19) was canvassed and before our survey was implemented (2023).

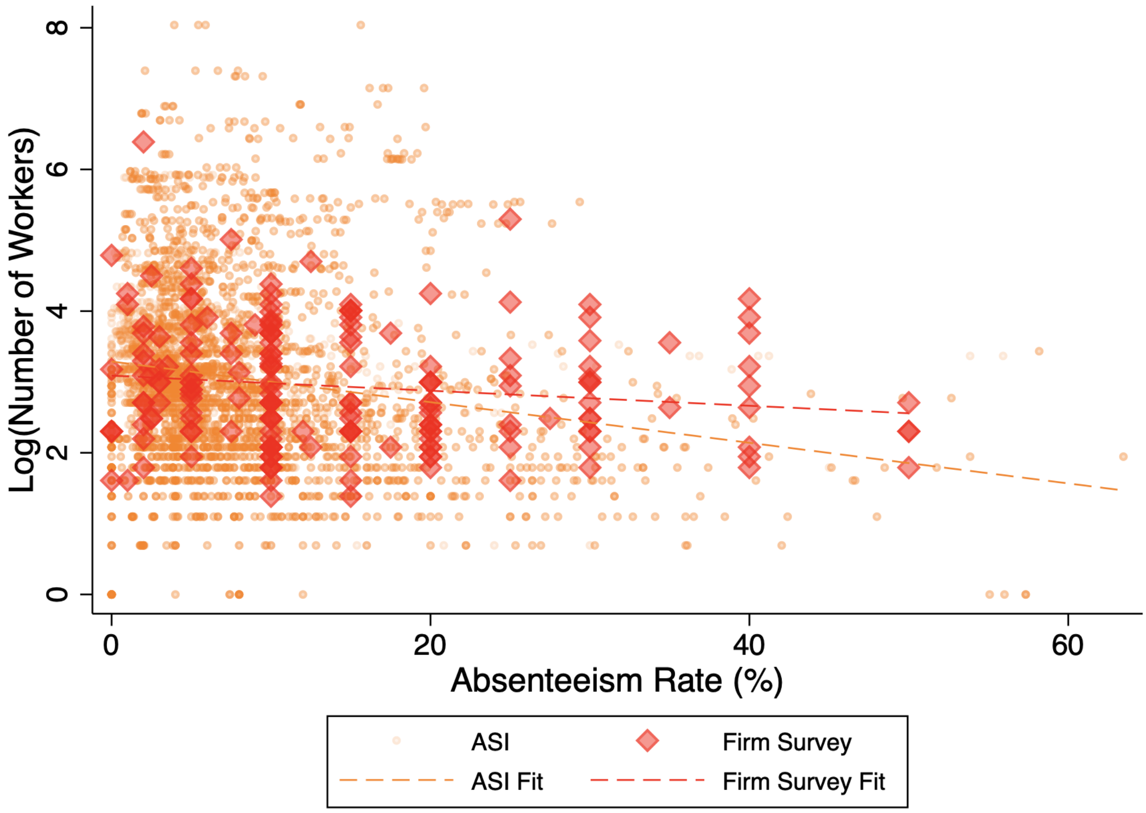
establishment.

Respondents. We target respondents who possess adequate knowledge of day-to-day production management within the establishment, such as managers, owners, clerks, supervisors, and accountants.

Survey Completion. A total of 260 surveys were initiated, and 206 were completed. Among the 54 incomplete surveys, 50 were due to non-consent, missing responses, or ineligibility determined during the screening process. In 2 cases, the establishment owner began the survey but stopped midway.

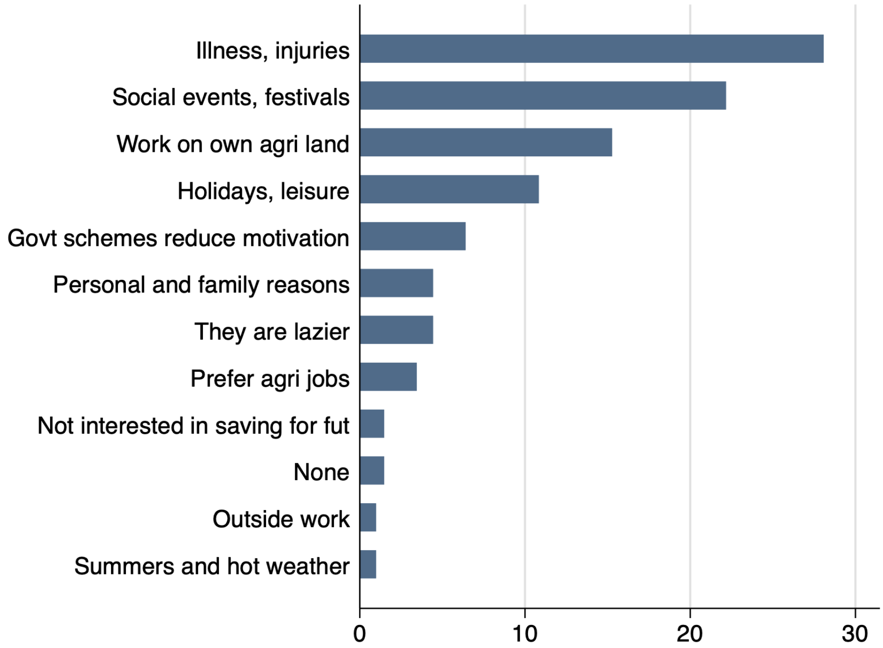
An additional dimension of interest is the caste identity of the respondent. Consistent with the findings in Goraya (2023), a majority of establishments are owned by individuals from the General caste category (64%), followed by Other Backward Castes (28%) and Scheduled Castes/Scheduled Tribes (7%).

Figure C.1: Correlation between Absenteeism and Number of Workers

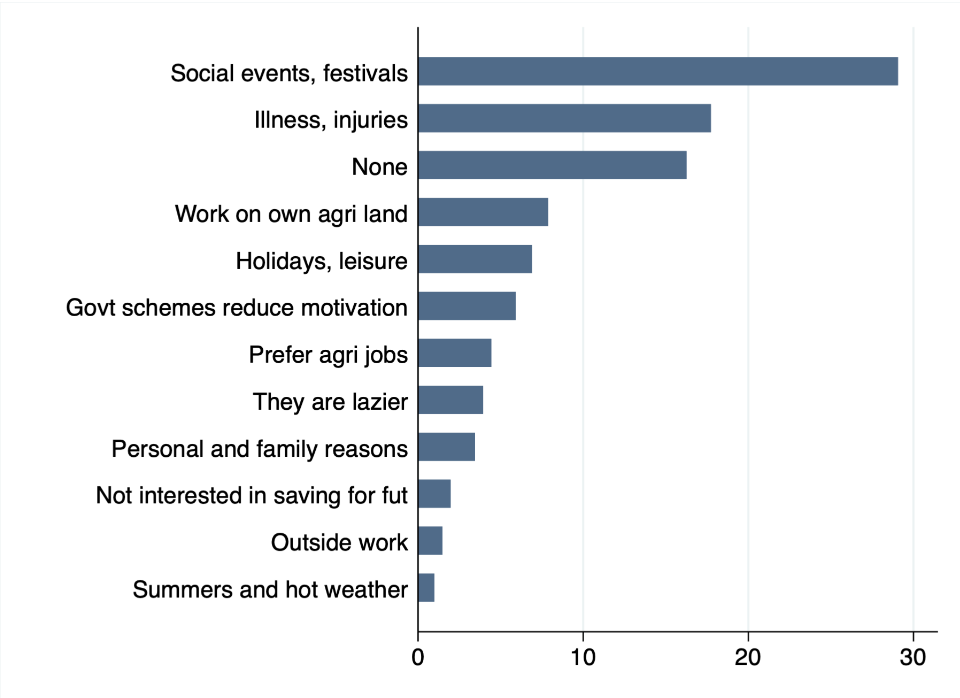


Note: This scatter plot shows the relationship between firm size (number of workers) and absenteeism rates. Red diamond points use data from the firm survey, while the smaller circle points use ASI data. For the latter, we restrict the sample to the four districts within Odisha that are covered by the firm Survey. Each point represents a firm, with fitted trend lines showing the correlation between firm size and absenteeism.

Figure C.2: Reasons for Worker Absenteeism



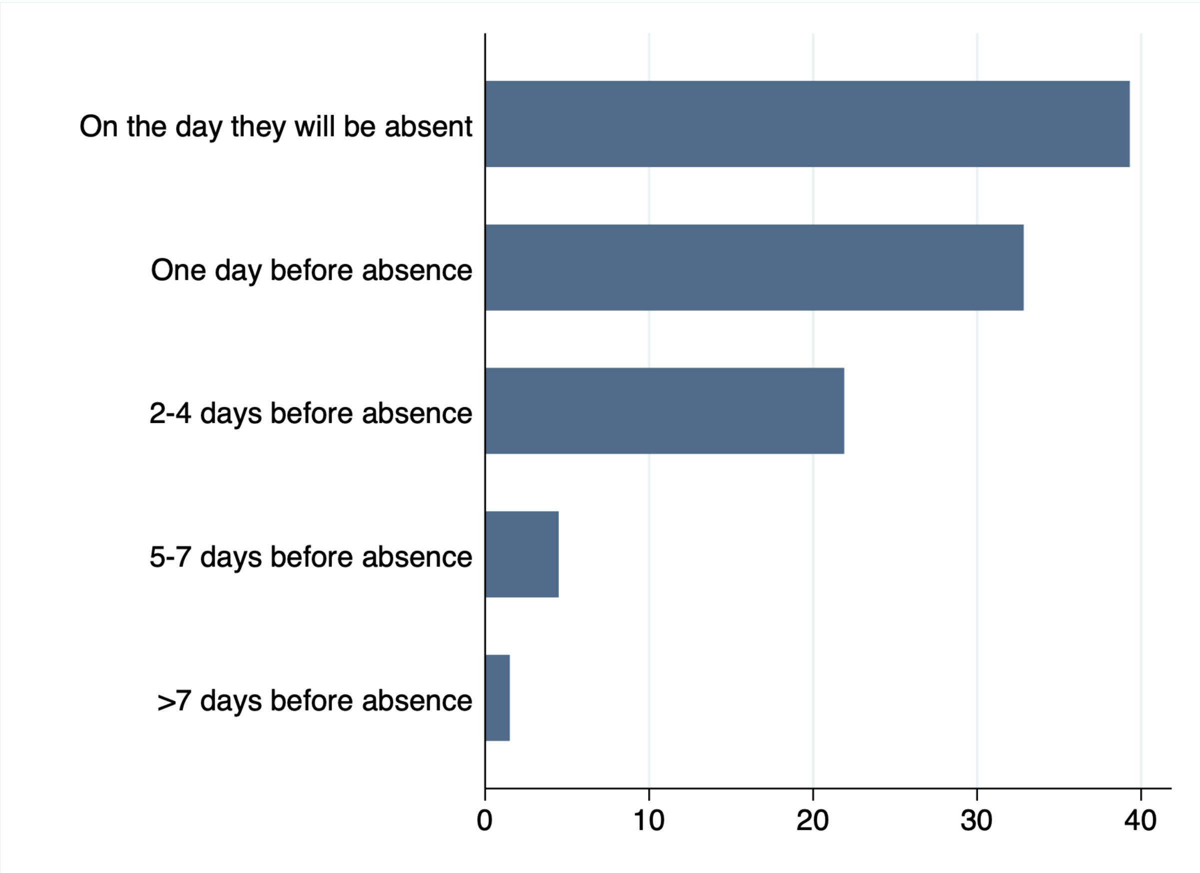
(a) Primary Reason



(b) Secondary Reason

Note: This figure shows the distribution of primary and secondary reasons for worker absenteeism as reported by firm managers in the survey. Percentages represent the proportion of firms citing each reason.

Figure C.3: When do Workers Report Absences?



Note: This figure shows the distribution of when unskilled workers are likely to inform employers about their absence from work, according to firm managers in the survey.