The impact of temperature on worker absenteeism in the Indian services sector.

Ridhima Gupta* E. Somanathan†

Abstract

Are salaried workers more likely to be absent on hotter days? Do cooling technologies reduce heat stress and lower worker absenteeism? In this paper, we answer these questions by analysing daily data on 274 employees across 86 locations. We find that higher temperatures lead to more absenteeism but only for workers without access to climate control technologies at work. Cooling technologies are therefore adaptive. We also find that higher temperatures decrease the probability of missing work for workers with climate control. Given that absenteeism takes worker output to zero, our findings imply that firms can minimise output losses by investing in technologies that produce thermal comfort.

*Faculty of Economics, South Asian University, Delhi. ridhima@sau.int.
†Economics and Planning Unit, Indian Statistical Institute, Delhi. som@isid.ac.in

1 Introduction

In India, the formal organised sector contributed almost 50% to Gross Value Added (GVA) in 2017-18 (Murthy, 2019). Although the the share of salaried workers in total employment in India stood at 22.8% in 2017-18, there has been a steady increase in formal sector employment in India overtime due to economic growth (Mehrotra et al., 2019). The share of regular wage workers increased from 14.3 per cent in 2004-5 to 22.8 percent in 2017-18 (Mehrotra et al., 2019). India is projected to continue to grow and as it does more and more workers will be employed by the formal sector in the future. Most climate models also predict an increase in the number of hot days at all land locations in this century (Lehner et al., 2018; Sillmann et al., 2013; Seneviratne et al., 2014; Trenberth, 2015; Perkins et al., 2012). These facts suggest that it is important to understand the impact of temperature on worker productivity in the formal sector that employs salaried workers.

Workers can respond to thermal discomfort in two ways, they are less productive at work and they miss work more often. The former channel has been extensively studied particularly in the physiology literature that concluded that higher temperatures reduced worker output. Prior evidence of temperature impacts on worker absenteeism in developing countries is much more limited. In this study, we focus on worker absenteeism because that is largely the only observable measure of worker productivity in the formal sector.

For developing countries, Somanathan et al. (2021); Adhvaryu et al. (2020) have analysed the impact of high temperatures on worker output and estimated temperature impacts on worker absenteeism. Given their primary focus is analysing worker output that is likely to be more impacted by higher daytime temperatures at the workplace, these studies do not examine the impact of mean temperature on worker absenteeism. Somanathan et al. (2021) use temperature bins of maximum temperature to model a non-linear relationship between worker absenteeism and temperature whereas Adhvaryu et al. (2020) use reanalysis data to calculate wet bulb temperature and its impact on absenteeism in two specifications in their paper but do not analyse the impact of climate control technologies on worker absenteeism. Recently, Heyes and Saberian (2022) use Indian survey data to identify temperature impacts on ability to work. Their measure of absenteeism is self reported inability to work or carry out normal duties in the last 30 days. Spencer and Urquhart (2021) too use self reported survey data on absenteeism and link it to extreme climatic events in Jamaica. The authors' do not find any statistically significant impact of heat on the odds of being absent from work.

There are a few papers that study heat related effects of worker absenteeism in high income countries. Zander et al. (2015) use self-reported estimates of work absenteeism due to heat to quantify the cost of productivity loss resulting from work related heat stress in Australia. The impacts of temperature on time allocation in US have been estimated by Graff Zivin and Neidell (2014). Since the authors focus on time allocation they do not distinguish between not going to work for one eight-hour day and working one hour less for eight days. Markham and Markham (2005) use data from piece-rate workers and a single time series to analyse effects of weather on the absenteeism rate in a plant in North Carolina. However, the temperature distribution in North Carolina will be lower than in India.

In this paper, we examine the impact of daily temperature on worker absenteeism in the formal sector in India. We measure temperature using daily maximum temperature and show the results with daily mean temperature in an appendix. Our preferred measure of temperature is maximum temperature because we only have data on the type of climate control at the workplace. But absenteeism can be influenced by high night time temperatures and cause thermal discomfort in absence of climate control technologies at home. We attempt to find indirect evidence for this channel by using mean temperature to measure temperature.

We estimate separate models for workers with access to climate control technologies and workers without climate control. The data on worker absenteeism is from a large welding company in India. We obtain daily data on worker attendance that spans 86 locations (see Figure 1) across India. More than 200 workers in these locations are followed over a period of almost 3 years from 2016-2018. The workers employed by this company comprise mainly of engineers. They commute on a daily basis to offer assistance with the installation and use of welding equipment to their clients. This allows us to identify outdoor workers who are hit the hardest by high temperatures.

Our results imply that higher maximum temperatures increase the probability of missing work for workers without access to climate control at work. We explore both non-linear and lagged temperature impacts and our findings remain intact for this sample of workers. Maximum temperatures at the top of the temperature distribution decrease the probability of missing work for workers that work in climate controlled environments. For workers with climate control, we estimate contemporaneous impacts of about 4% per degree relative to mean absenteeism and lagged impacts of about 5% per degree.

The findings above are based on regression models that control for worker fixed effects and time fixed effects i.e. year-month and day of the week fixed effects. All the models are estimated for two sub-samples in the data, workers with and without access to climate control technologies. The source of the identifying variation is the day-to-day variation in worker attendance and temperature that remains after we have removed the variation due to unchanging worker characteristics and seasonality due to day of the week and month and year.

The primary contribution of this study is to the literature on the relationship between worker absenteeism and temperature in the formal sector that employs salaried workers. Prior two studies (Somanathan et al., 2021; Adhvaryu et al., 2020) in this area have largely focused on the impact of temperature on worker output in industries that record output. In many settings though worker output is not readily observable. The only observable measure of worker productivity is worker absenteeism. Estimates of the magnitude of the impact of higher temperature on worker productivity measured by worker absenteeism are important for two reasons. First, the formal sector employs a substantial number of people even in developing countries. About 24% of India's workforce was made of regular salaried employees in 2018 (Initiative, 2020). Second, salaried employees have higher productivity than piece-rate employees (Bryson et al., 2011) and therefore their absenteeism leads to a larger decline in output.

2 Data

The following section describes the data sources.

2.1 Worker Data

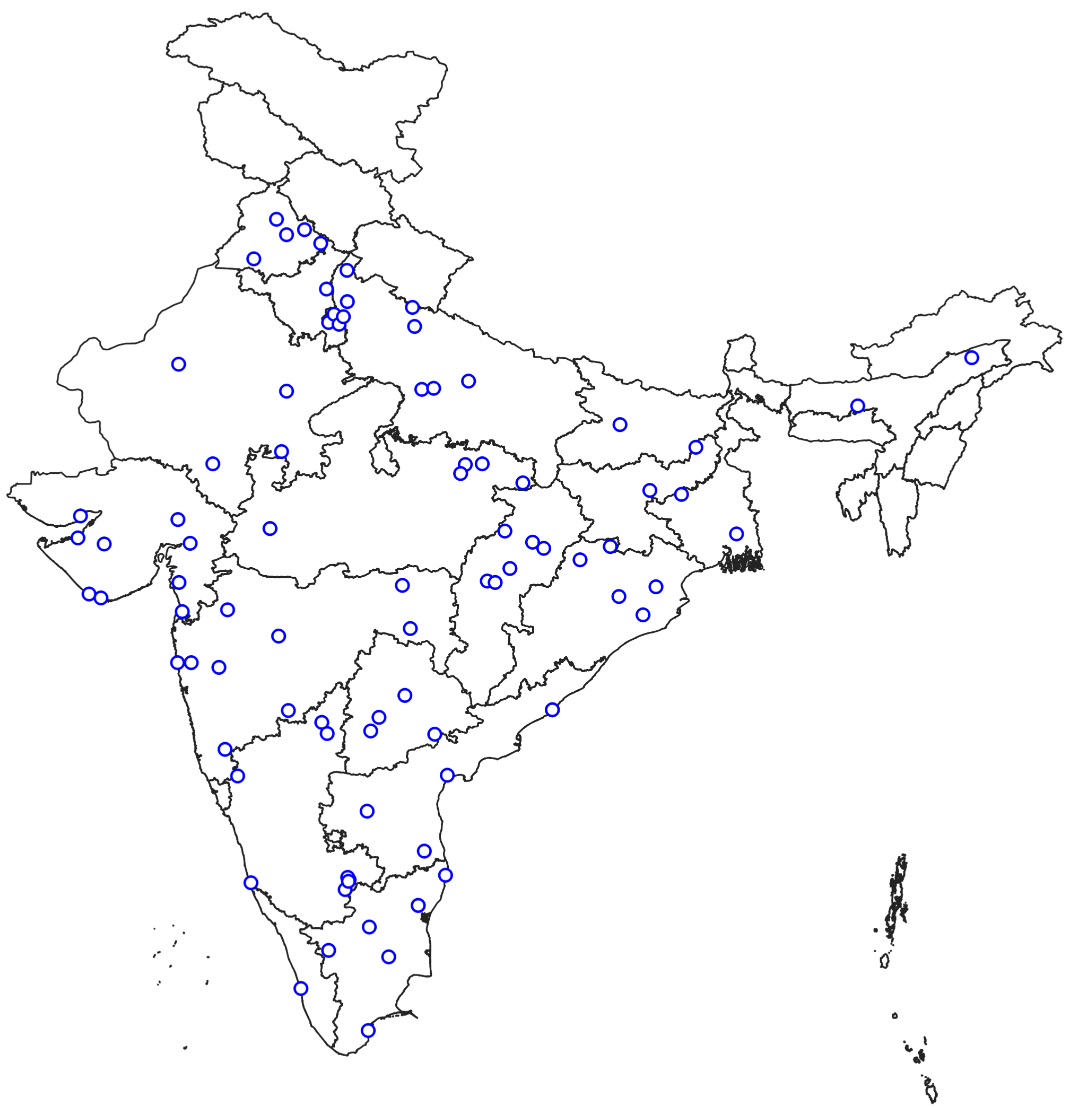
Daily data on worker attendance is taken from a large welding company in India. The company is one of the largest provider of high-quality welding equipment, consumables, automation solutions and training. A majority of their workers are sales workers that are engineers. Their primary job is to offer assistance to their clients with the use of welding equipment and promote sales. The workers are required to make visits to sites where the machinery is installed. Their main mode of transport is two-wheelers. The non-sales workers are stationed in offices and their task ranges from managing the company accounts to dealing with human resources. All employees are salaried and are entitled to a fixed monthly wage.

The company follows a six day working week with Sunday being a holiday. Employees can take upto 44 days of paid leave every year. This includes seven days of casual leave, seven days of sick leave and 31 days of privilege leave. We measure worker absenteeism by a dummy variable that takes the value1 if a worker was absent on a working day and zero otherwise. We, therefore, exclude holidays and Sundays from the sample.

The sales workers do not have access to any climate control technologies when they are on site visits. On days that they are not performing site visits they work from their climate-controlled offices. The non-sales workers that comprise of accountants, human resource personnel etc., on the other hand, work in climate controlled environments. The data therefore has information on the type of climate control technology that is available to each worker.

The dataset includes daily data on more than 200 workers that are spread across 86 locations in India. The study covers the period from 2016-2018. The use of climate control technologies by a worker affects his heat exposure and in turn his productivity. Temperature impacts on absenteeism should be higher where climate control and cooling is likely to be limited. Hence, we split workers into two sub-samples depending on whether a worker had access to climate control technology. One worker was absent for more than 200 days in 2016 and 2017. He was treated as an outlier and dropped from the sample.

Figure 1: Workplace Locations



2.2 Weather Data

Daily data on temperature and rainfall were obtained from the Global Surface Summary of the Day (GSOD) data from the National Climatic Data Center, NESDIS, NOAA, U.S. Department of Commerce (NOAA, 2015). This database contains global daily station-level data on weather variables such as temperature and rainfall. Data from the closest weather station to a factory location is assigned to all the workers in that location. This data allows us to estimate impacts of temperature on worker absenteeism for a larger sample as it contains data from weather stations all over India. Station-level data provides measurements of actual weather in a location that can be used to accurately pin down weather.

Table 1 and Figure 2 report summary statistics of the key variables of interest. The dependent variable absenteeism is defined as the number of days a worker is absent per 1000 days. It is binary and takes the value 1000 if a worker was absent on a working day and 0 otherwise. Figure 2 shows that worker absenteeism is highest during May, the hottest month during which schools are typically shut, followed by the month of December when business activity is low all over the world. The summary statistics in Table 1 are shown for the 2 sub-samples in the study i.e. workers with and without access to climate control technologies.

Figure 2: Worker absence and Mean Temperature by month

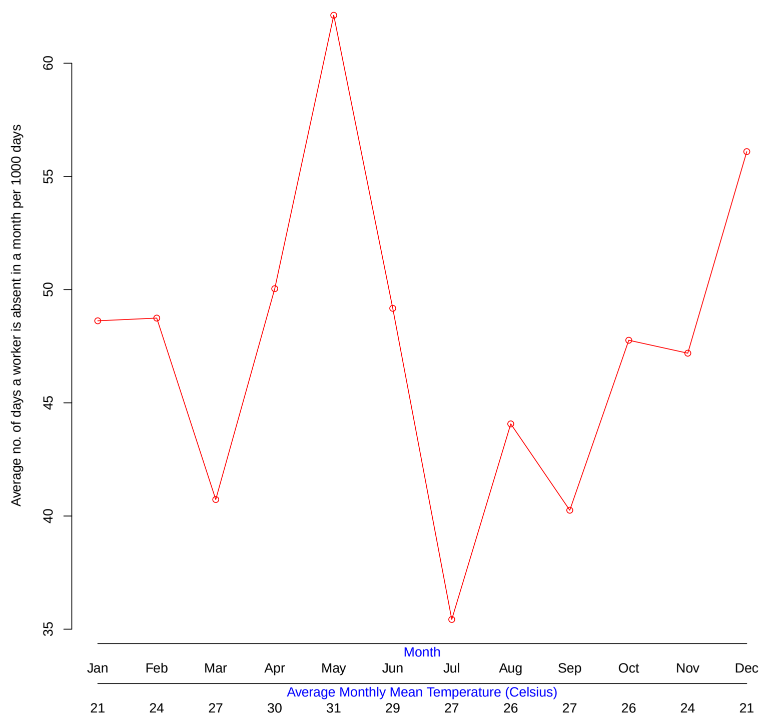


Table 1: Summary Statistics

Sample of Workers without Climate Control					
	Mean	Standard Deviation	Min	Max	Count
Absent (No. of days a worker is absent per 1000 days)	44.11	205.34	0	1000	88783
MaxT (Celsius)	32.40	4.49	11	50	69472
$Weekly_{MaxT}$	32.43	4.28	13.20	47.43	69127
B2$_{MaxT}$ (Day MaxT was in (28, 31] in the week)	0.27	0.44	0	1	66499
B3$_{MaxT}$ (Day MaxT was in (31, 34] in the week)	0.27	0.45	0	1	66499
B4$_{MaxT}$ (Day MaxT was in (34, 37] in the week)	0.18	0.38	0	1	66499
B5$_{MaxT}$ (Day MaxT was in (37, 50] in the week)	0.13	0.34	0	1	66499
L2$_{MaxT}$ (No. of days MaxT was in (28, 31] in the week)	1.90	2.25	0	7	66499
L3$_{MaxT}$ (No. of days MaxT was in (31, 34] in the week)	1.90	2.24	0	7	66499
L4$_{MaxT}$ (No. of days MaxT was in (34, 37] in the week)	1.22	1.98	0	7	66499
L5$_{MaxT}$ (No. of days MaxT was in (37, 50] in the week)	0.93	2.12	0	7	66499
Rainfall (mm)	3.12	12.75	0	361.95	66748
Sample of Workers with Climate Control					
Absent (No. of days a worker is absent per 1000 days)	56.69	231.24	0	1000	92511
MaxT (Celsius)	31.86	4.29	13.22	47	90756
$Weekly_{MaxT}$	31.88	4.09	15.86	45.56	90289
B2$_{MaxT}$ (Day MaxT was in (28, 31] in the week)	0.32	0.47	0	1	88852
B3$_{MaxT}$ (Day MaxT was in (31, 34] in the week)	0.25	0.43	0	1	88852
B4$_{MaxT}$ (Day MaxT was in (34, 37] in the week)	0.17	0.38	0	1	88852
B5$_{MaxT}$ (Day MaxT was in (37, 50] in the week)	0.09	0.29	0	1	88852
L2$_{MaxT}$ (No. of days MaxT was in (28, 31] in the week)	2.27	2.40	0	7	88852
L3$_{MaxT}$ (No. of days MaxT was in (31, 34] in the week)	1.71	2.15	0	7	88852
L4$_{MaxT}$ (No. of days MaxT was in (34, 37] in the week)	1.21	2.01	0	7	88852
L5$_{MaxT}$ (No. of days MaxT was in (37, 50] in the week)	0.65	1.76	0	7	88852
Rainfall (mm)	3.23	13.86	0	361.95	86523

3 Temperature Effects on Absenteeism

3.1 Linear Modelling of Temperature

Daily absenteeism was regressed on daily maximum temperature, daily minimum temperature, rainfall and a bunch of fixed effects. We estimate

$$A_{id} = w_i + \alpha_D + \nu_M + \gamma_Y + \beta_1 \textit{MaxT}_{id} + \delta \textit{Rain}_{id} + u_{id} \tag{1}$$

where the subscripts i and d refer to worker and day respectively, $MaxT$ is the average daily maximum temperature, $Rain$ is the daily rainfall and u is the error term. w_i is a worker-specific intercept that controls for all worker-specific time-invariant factors. The coefficient on α_D accounts for seasonality due to the day of the week, ν_M for month and γ_Y controls for year specific seasonality. The β coefficients, therefore, capture the effect of deviations from mean maximum temperature on deviations from mean absenteeism after removing the variation due to seasonality and fixed characteristics of workers and rainfall.

Residuals in these regressions could also be spatially correlated across workers in the same location and serially correlated over days. We addressed this issue by using Driscoll-Kraay standard errors (Driscoll and Kraay, 1998; Hoechle, 2007) that are robust to both cross-sectional dependence and temporal dependence when the time dimension becomes large. Since we have daily data on workers for almost 3 years, the time dimension is large.

Sustained high temperatures may lead to fatigue or illness. Following prior literature, we also estimate the impact of lagged temperature on absenteeism. Lagged temperature is measured by the average temperature up to 6 working days prior and on the day the worker reported to work. We modify Equation 1 and replace concurrent temperature with the average of lagged and current temperature.

The model is of the form

$$A_{id} = w_i + \alpha_D + \gamma_Y + \nu_M + \beta_1 \textit{Weekly}_{MaxT} + \delta \textit{Rain}_{id} + u_{id} \tag{2}$$

The new temperature measure in Equation 2 is $Weekly_{MaxT}$. This is the average of lagged and current temperature with lags upto the last 6 days.

4 Non-Linear Modeling of Temperature

We model temperature with indicator variables to capture potential nonlinear and lagged effects on absenteeism. We follow the approach of Schlenker and Roberts (2009); Somanathan et al. (2021) to create the indicator variables. We use temperature bins defined as (10,28], (28,31], (31,34], (34,37], (37,50]. Taken together, these bins do not overlap and span the observed range of temperatures in the data, so that any given day is assigned to exactly one bin. The first bin is the omitted category. We estimate

$$A_{id} = w_i + \alpha_D + \gamma_Y + \nu_M + \sum_{j=2}^{5} \omega_j B^j_{MaxTid} + \delta Rain_{id} + u_{id} \tag{3}$$

The temperature control in Equation 3 is B^j, an indicator for the day falling in temperature bin j. To calculate lagged impacts, we estimate 4 below. Here, L^j is a count of the number of days falling in bin j in the six days preceding day d and on day d.

equation

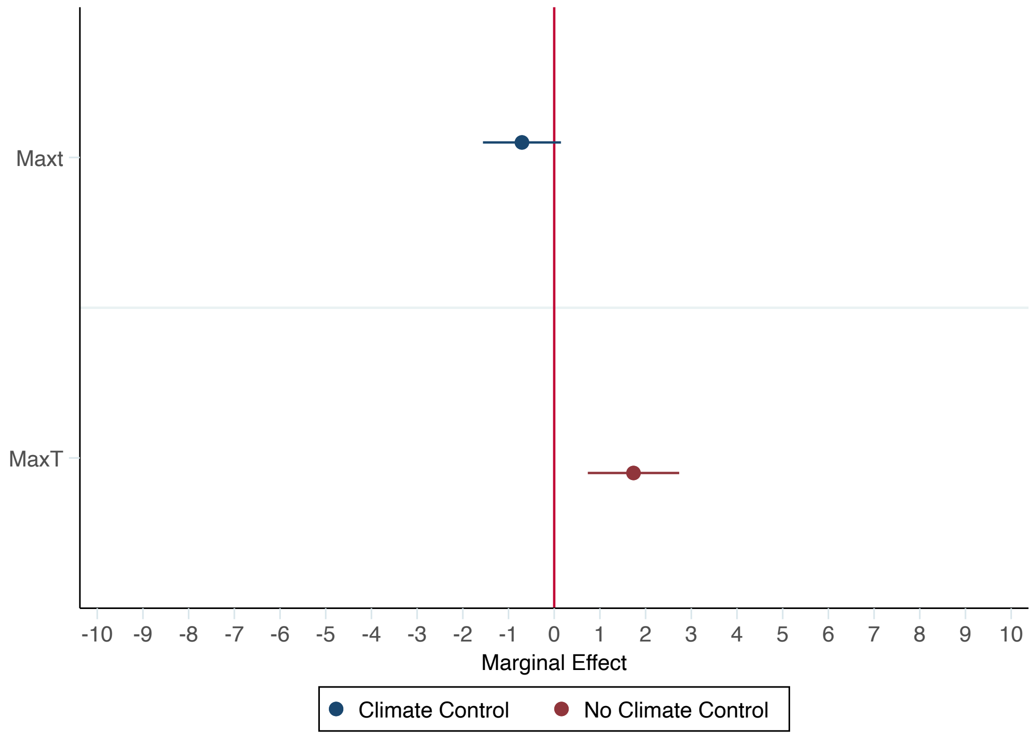
$$A_{id} = w_i + \alpha_D + \gamma_Y + \nu_M + \sum_{j=2}^{5} \omega_j L^j_{MaxTid} + \delta Rain_{id} + u_{id} \tag{4}$$

equation

4.1 Results

Coefficient estimates of Equation 1 are shown in Figure 3. We find that daily maximum temperature increases absenteeism for workers without climate control at work. Results from column (1) imply about a 3.7% increase in the probability of missing work for a 1°C increase in maximum temperature relative to mean absenteeism in this sample. For workers with climate control, we do not find any impact of maximum temperature on absenteeism.

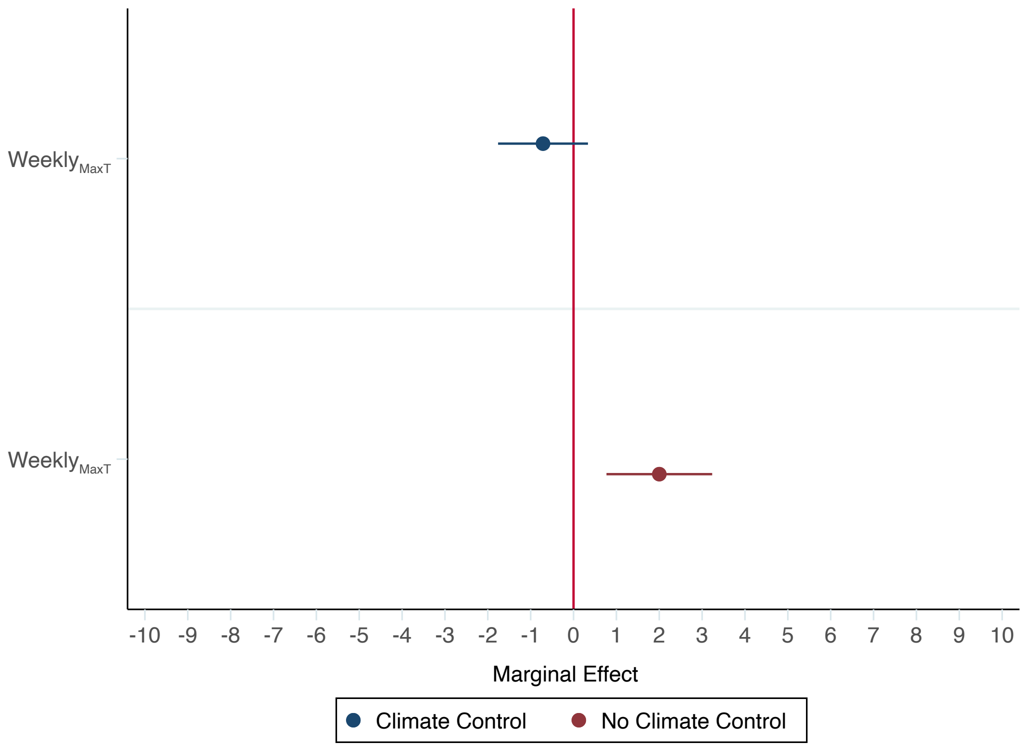
Figure 3: Linear Impact of Temperature on Absenteeism



Notes: The dependent variable is absent per 1000 days. Plots depict the β coefficients from Equation 1 with 95% confidence intervals computed using Driscoll-Kraay standard errors that are robust to cross-sectional and temporal dependence. Regression estimates are shown for both the sub-samples in the study i.e. workers with and without climate control. The regression uses 86,499 observations on 143 employees that have access to climate control technologies and 66,746 observations on 130 employees that did not have access to climate control technologies over the three year time span of the study. Year, month, and day-of-the-week fixed effects are included in the regression (Equation 1).

The impact of higher temperatures during the week on worker attendance i.e. estimates of Equation 2 are shown in Figure 4. The results for the weekly temperatures are similar in direction and slightly higher in magnitude to the results from the model shown in Figure 3 that accounts for contemporaneous temperature. The probability of missing work due to a 1°C increase in average maximum temperature in the week increases to about 1.93 per thousand in absence of climate control at work. This corresponds to about 4.3% of the mean absenteeism in this sample.

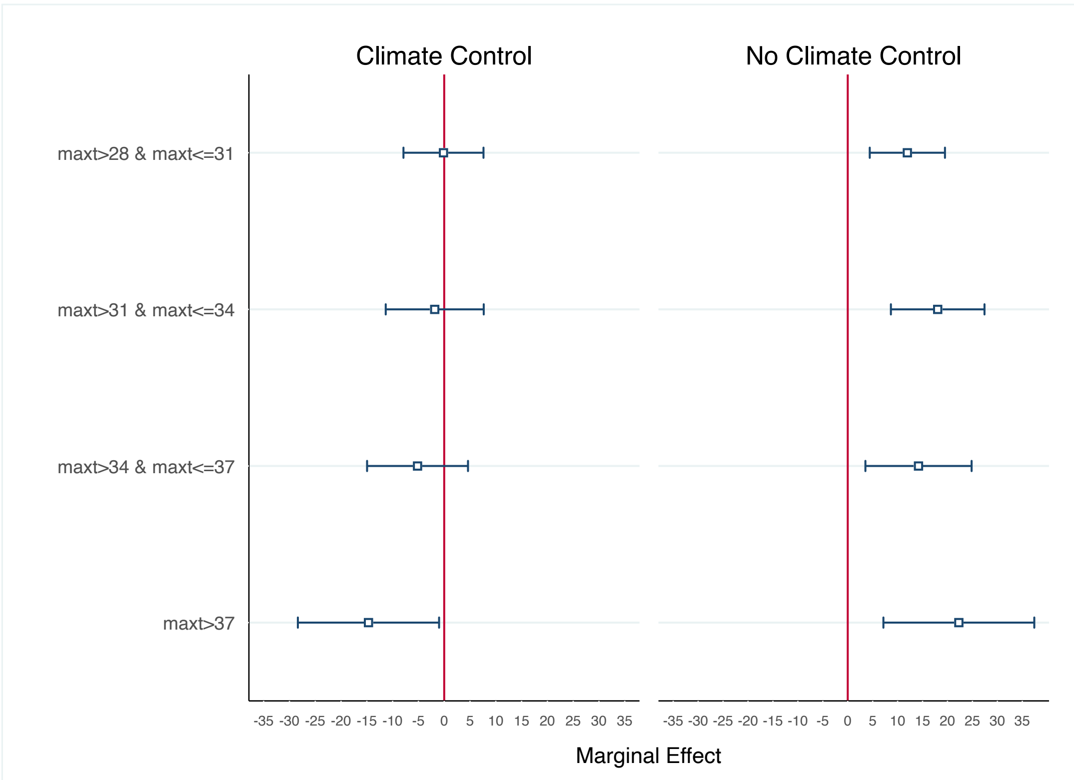
Figure 4: Impact of Average Weekly Temperature on Absenteeism



Notes: The dependent variable is absent per 1000 days. Plots depict the β coefficients from Equation 2 with 95% confidence intervals computed using Driscoll-Kraay standard errors that are robust to cross-sectional and temporal dependence. Regression estimates are shown for both the sub-samples in the study i.e. workers with and without climate control. The regression uses 85,653 observations on 143 employees that have access to climate control technologies and 65,756 observations on 127 employees that did not have access to climate control technologies over the three year time span of the study. Year, month, and day-of-the-week fixed effects are included in the regression (Equation 2).

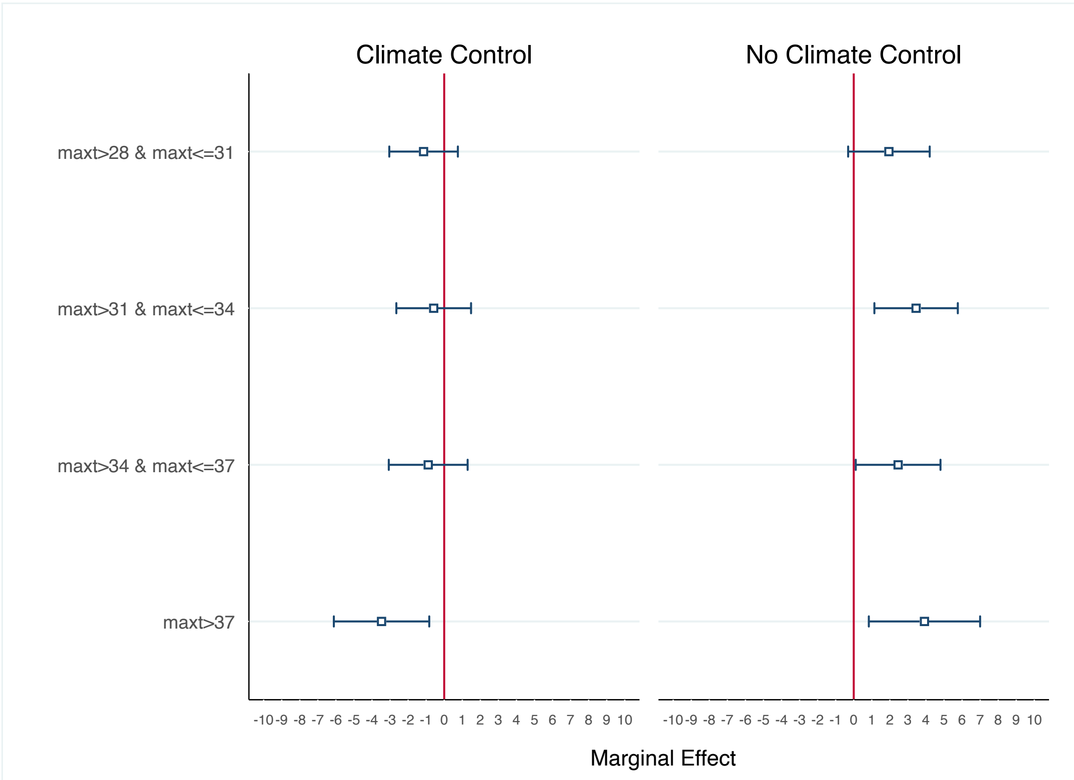
Figure 5 plots non-linear impacts of contemporaneous temperature on absenteeism. If there is no climate control at work then an additional day that maximum temperature is above 37°C relative to a day with maximum temperature less than 28°C causes an increase by 22.28 per thousand in the probability of missing work. This estimate is negative and decreases by about 66% for the sample of workers that have access to climate control. Effects of exposure to sustained high temperatures are shown in Figure 6. The direction of the impact is the same compared to the estimates in 6 but the magnitude decreases by one order of magnitude due to climate control.

Figure 5: Impact of High Temperatures on Absenteeism



Notes: The dependent variable is absent per 1000 days. Plots depict the ω_j and ζ_j coefficients from Equation 3 with 95% confidence intervals computed using Driscoll-Kraay standard errors that are robust to cross-sectional and temporal dependence. Regression estimates are shown for both the sub-samples in the study i.e. workers with and without climate control. The regression uses 84,738 observations on 143 employees that have access to climate control technologies and 63.923 observations on 126 employees that did not have access to climate control technologies over the three year time span of the study. Year, month, and day-of-the-week fixed effects are included in the regression (Equation 3).

Figure 6: Impact of Sustained High Temperatures on Absenteeism



Notes: The dependent variable is absent per 1000 days. Plots depict the ω_j and ζ_j coefficients from Equation 3 with 95% confidence intervals computed using Driscoll-Kraay standard errors that are robust to cross-sectional and temporal dependence. Regression estimates are shown for both the sub-samples in the study i.e. workers with and without climate control. The regression uses 84,738 observations on 143 employees that have access to climate control technologies and 63.923 observations on 126 employees that did not have access to climate control technologies over the three year time span of the study. Year, month, and day-of-the-week fixed effects are included in the regression (Equation 3).

To sum-up the results from our models indicate that higher temperatures (both contemporaneous and lagged) impact worker absenteeism. We find that for workers without access to climate control higher day-time temperatures make it more likely that they will miss work. This finding is robust across specifications. We also find that climate control reduces the probability of missing work on days with extreme temperatures.

5 Discussion and Conclusions

In this paper we analyse the impact of daily maximum temperature on worker absenteeism by analysing data from multiple locations across India. The sample consists of salaried workers that will in the near future constitute a larger and larger proportion of the Indian workforce due to economic growth. We estimate separate regressions for workers with climate control and for workers without climate control to determine whether climate control technologies are adaptive.

The results imply that higher day-time temperatures increase worker absenteeism for workers without climate control at work. We find evidence of both concurrent and lagged impacts of maximum temperature on the probability of missing work for these workers. The estimated impacts of a hot week on absenteeism are about 4.7% per degree relative to the mean absenteeism in this sample. Another robust finding is workers that work in climate controlled environments are less likely to miss work due to exposure to temperatures that exceed 37°C.

Our findings conform to the economics literature (Somanathan et al., 2021; Adhvaryu et al., 2020) on estimates of temperature on worker absenteeism in developing countries. We are unable to compare our estimates with the estimates in (Somanathan et al., 2021; Adhvaryu et al., 2020) because of differing methodologies. (Somanathan et al., 2021) use two temperature bins in daily maximum temperature to estimate contemporaneous impacts and counts of the number of days maximum temperature was in each bin the preceding six days to estimate lagged impacts. Our preferred measure is also daily maximum temperature but the contemporaneous and lagged impacts of both these temperature variables in our paper are subsumed in one variable that is a count of the number of days temperature was in a bin on the same day and the preceding six days.

Knowledge of temperature impacts of worker absenteeism in the formal sector is important particularly for developing countries that already employ a large number of people in it. Further, as the data shows this number is likely to increase due to economic growth. If we assume that all workers are equally productive and a proportionate increase in absenteeism implies a proportionate reduction in worker output, then our estimates imply a 3.7% per degree decline in output due to worker absenteeism for workers without climate control at work due to a 1°C increase in maximum temperature. These workers are more likely to be impacted by global warming.

References

Achyuta Adhvaryu, Namrata Kala, and Anant Nyshadham. The Light and the Heat: Productivity Co-benefits of Energy-Saving Technology. *Review of Economics and Statistics*, 102(4):779–792, 2020.

Alex Bryson, Babatunde Buraimo, and Rob Simmons. Do Salaries Improve Worker Performance? *Labour Economics*, 18(4):424–433, 2011.

John C Driscoll and Aart C Kraay. Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data. *Review of Economics and Statistics*, 80(4):549–560, 1998.

Joshua Graff Zivin and Matthew Neidell. Temperature and the allocation of time: Implications for climate change. *Journal of Labor Economics*, 32(1):1–26, 2014.

Anthony Heyes and Soodeh Saberian. Hot days, the ability to work and climate resilience: Evidence from a representative sample of 42,152 indian households. *Journal of Development Economics*, 155:102786, 2022.

Daniel Hoechle. Robust Standard Errors for Panel Regressions with Cross-Sectional Dependence. *Stata Journal*, 7(3):281, 2007.

Accountability Initiative. A Comprehensive Overview: Defining and Measuring Informality in India's Labour Market. 2020.

Flavio Lehner, Clara Deser, and Benjamin M Sanderson. Future Risk of Record-Breaking Summer Temperatures and its Mitigation. *Climatic change*, 146(3-4):363–375, 2018.

Steven E Markham and Ina S Markham. Biometeorological effects on worker absenteeism. *International journal of biometeorology*, 49(5):317–324, 2005.

Santosh Mehrotra et al. Informal employment trends in the indian economy: Persistent informality, but growing positive development. *ILO (working paper)*, 2019.

SV Ramana Murthy. Measuring informal economy in india: Indian experience. In *Seventh IMF Statistical Forum, Washington, DC*, 2019.

NESDIS NOAA. Global surface summary of the day. *U.S. Department of Commerce*, 2015.

SE Perkins, LV Alexander, and JR Nairn. Increasing Frequency, Intensity and Duration of Observed Global Heatwaves and Warm spells. *Geophysical Research Letters*, 39(20), 2012.

Wolfram Schlenker and Michael J Roberts. Nonlinear Temperature Effects Indicate Severe Damages to US Crop Yields under Climate Change. *Proceedings of the National Academy of Sciences*, 106(37):15594–15598, 2009.

Sonia I Seneviratne, Markus G Donat, Brigitte Mueller, and Lisa V Alexander. No Pause in the Increase of Hot Temperature Extremes. *Nature Climate Change*, 4(3):161, 2014.

Jana Sillmann, Viatcheslav V Kharin, FW Zwiers, Xuebin Zhang, and D Bronaugh. Climate Extremes Indices in the CMIP5 Multimodel Ensemble: Part 2. Future Climate Projections. *Journal of Geophysical Research: Atmospheres*, 118(6):2473–2493, 2013.

Eswaran Somanathan, Rohini Somanathan, Anant Sudarshan, and Meenu Tewari. The Impact of Temperature on Productivity and Labor Supply: Evidence from Indian Manufacturing. *Journal of Political Economy*, 129(6):1797–1827, 2021.

Nekeisha Spencer and Mikhail-Ann Urquhart. Extreme climate and absence from work: Evidence from jamaica. *International Journal of Disaster Risk Science*, 12(2):232–239, 2021.

Kevin E Trenberth. Has There Been a Hiatus? *Science*, 349(6249):691–692, 2015.

Kerstin K Zander, Wouter JW Botzen, Elspeth Oppermann, Tord Kjellstrom, and Stephen T Garnett. Heat stress causes substantial labour productivity loss in australia. *Nature climate change*, 5(7):647–651, 2015.