Essays on environmental, labor, and development economics

by

Daniel Minoru Higa Reyes

M.A., University of California Santa Barbara, 2015
M.Sc., Tilburg University, 2011
B.A., Pontificia Universidad Catolica del Peru, 2008

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Department of Economics
Faculty of Arts and Social Sciences

SIMON FRASER UNIVERSITY
Summer 2023

Declaration of Committee

Name: **Daniel Minoru Higa Reyes**

Degree: **Doctor of Philosophy**

Thesis title: **Essays on environmental, labor, and development economics**

Committee:

Chair: Bertille Antoine
Professor, Economics

Fernando Aragon
Co-supervisor
Professor, Economics

Hendrik Wolff
Co-supervisor
Professor, Economics

Kevin Schnepel
Committee Member
Associate Professor, Economics

Alexander Karaivanov
Examiner
Professor, Economics

Sumeet Gulati
External Examiner
Professor
Department of Land and Food Systems
University of British Columbia

Abstract

This thesis is composed of three essays on environmental, labor, and development economics.

The first chapter investigates whether traffic congestion affects time allocation. I use highly granular smartphone data from Mexico City to empirically study how traffic congestion affects work-time allocation. I find that traffic increases hours worked. The effect is driven by workers leaving work later, rather than by changes in arrival time. I show modest evidence that labor income does not increase despite the increase in total hours worked. These results highlight an avoidance mechanism (consistent with bottleneck models) that has been previously overlooked when estimating the costs of congestion.

The second chapter is co-authored with Jerico Fiestas-Flores and Javier Montoya-Zumaeta. It investigates how pandemics affect nature. We explore the effect of COVID-19 on deforestation in the Amazon rainforest in Peru. Using an event study design and a difference-in-differences approach, we find that COVID-19 increased deforestation by 35%. This increased CO2 emissions by more than 17 million tons, representing a social cost equivalent to 3 times the national budget for forest management. The main mechanisms behind these findings are the reduction in monitoring efforts combined with an increase in illegal activities related to coca production and mining.

The third chapter studies whether raising temperatures due to climate change affects labor markets. This paper studies the effect of temperature on hours worked using panel data for Peru from 2007-2015. I combine hours worked from household surveys with reanalysis and satellite weather data. I find evidence that hours worked are negatively affected by hot temperatures. This effect is driven by informal jobs instead of jobs in industries highly exposed to the weather. These results suggest that labor market segmentation may play a role in the impacts of climate change on labor market outcomes in developing countries.

Keywords: traffic congestion; labor supply; temperature; deforestation; big data; remote sensing

Dedication

To my beloved parents, who passed away while I pursued my doctoral degree, I offer my heartfelt dedication. Although you are no longer physically present, your spirit continues to guide me. I am forever grateful for the sacrifices you made, and I carry your memory with me as I embark on this intellectual pursuit.

To the shining light in my life, Giuliana. Your support, understanding, and belief in me have been a constant source of motivation. You have embraced the challenges and sacrifices that came with my pursuit of knowledge, and your love has provided clarity in moments of doubt. Keep shining.

To my dear son, Benjamin. Your innocent laughter and boundless curiosity remind me of the beauty in this world and to keep fighting for a better environment. You have unknowingly taught me the importance of balance, reminding me of what truly matters in life.

Finally, this dedication extends to all those who are fighting for a second chance. To those who have faced adversity, setbacks, or missed opportunities, I celebrate your power of perseverance and your potential for transformation.

Acknowledgements

I would like to express my deepest gratitude to my supervisory committee Fernando Aragon, Hendrik Wolff, and Kevin Schnepel. I am truly fortunate to have had Fernando as my supervisor, and I am indebted to him for his generosity with his time and his dedication to my academic growth. Fernando not only believed in me but also provided invaluable guidance and support. His profound knowledge, genuine enthusiasm, and accurate insights have been instrumental in shaping the trajectory of this thesis. I am immensely grateful for the opportunity to learn from him. I also offer my heartfelt gratitude to Hendrik and Kevin. Their insightful feedback, constructive criticism, and support have contributed significantly to the refinement of my ideas and the development of this thesis. Their mentorship has been invaluable, and I am grateful for their willingness to share their knowledge and expertise. To Fernando, Hendrik, and Kevin, your guidance, mentorship, and support have been instrumental in the successful completion of this thesis. Without your collective wisdom and belief in my abilities, this accomplishment would not have been possible. Thank you for being such exceptional mentors and for shaping my academic journey in profound ways.

It takes a village to raise a Ph.D. student. I am truly humbled and fortunate to have received support and guidance from caring individuals who have contributed to my journey and the completion of this doctoral thesis. I extend my deepest appreciation to the graduate chairs in the Department of Economics for playing pivotal roles in my academic journey: Simon Woodcock, Alex Karaivanov, and Bertille Antoine for granting me the opportunity to enter the graduate program, for allowing me to continue my studies, and for providing support as I prepare to exit the program. I would also like to express my sincere thanks to the amazing staff members who have gone above and beyond to support me: Kathleen Riva, Carmen Chin, Mike Perry, and especially Lisa Agosti, the heart of the department. To Dat's Why for sharing their traffic data, which has been instrumental in my research. Additionally, I would like to thank Harjap and Laura from GSS, Sarah Zhang and Julia Lane at SFU Library, and members of the SFU Big Data Lab for their assistance in accessing crucial resources. I am indebted to my letter writers, Ofer Malamud, Matteo Picchio, and Doug Steigerwald, for their belief in my potential and their generous support in recommending me for various opportunities. My heartfelt gratitude to Rosa Morales and Jose Rodriguez, who provided me with my first job as a research assistant and introduced me to the world of research. To my wonderful coauthors, and to my fellow graduate students (Chimaobi, Esmaeil, Garam, Farouk, and

Hanif), thank you for your collaboration, support, and camaraderie throughout this journey. My sincere appreciation to all the faculty members at SFU, colleagues at conferences, and seminar speakers who have generously shared their expertise and provided valuable feedback on my work. Finally, I am truly grateful for the love and support from my close friends, family, and former bosses. To everyone who has played a part, whether big or small, in shaping my academic and personal growth, thank you from the bottom of my heart.

Table of Contents

List of Tables

List of Figures

Chapter 1

Traffic Congestion and Labor Supply: Evidence from Smartphone data

1.1 Introduction

Traffic congestion has become one of the "plagues of modern life" in most cities worldwide, depleting the benefits that cities offer (Arnott and Small, 1994). Congestion contributes to air pollution, increases crime, and wastes valuable leisure time spent seated in traffic. However, traffic congestion may also distort work-time allocation decisions, which would have important welfare implications if (for example) changes in hours worked are not compensated by changes in income. Likewise, this may lead to reconsidering the way we measure congestion externalities, a crucial factor in the calculus of congestion pricing and assessing the benefits of different urban transit policies.

This paper examines the effect of traffic congestion on work-time allocation. Because the relationship between these two variables is theoretically ambiguous, researchers have sought empirical methods to identify the causal effects. However, the lack of data directly measuring work time and traffic has prevented researchers from doing so up to this point, despite the high degree of policy relevance. The current debates over automobile use in the developed world (e.g., investment in electric vehicles or autonomous vehicles) and trends in developing countries (e.g., rapid growth in urban population and private vehicle ownership) indicate that congestion will likely increase over time.

The main contribution of the paper is to identify and quantify an unintended externality of traffic congestion previously overlooked when estimating the costs of congestion. Existing estimates of welfare loss from traffic congestion only consider the time lost on congested roads (Akbar et al., 2020, Kim, 2019). However, this may underestimate the real costs of congestion in two ways: (i) by missing the costly avoidance behavior of staying at work longer without receiving additional compensation, and (ii) as these measures of time lost include commuting times already reduced by the avoidance behavior.

To answer this question, I use data for one of the most congested cities in the world, Mexico City (Akbar et al., 2020, INRIX, 2019). I build a unique longitudinal dataset with individual daily hours spent at work and daily exposure to district-level traffic congestion for 2019. The smartphone data allows me to track the daily work-time allocation of individuals and to identify where individuals work (once combined with geocoded establishment-level data). I exploit the richness of the smartphone data to recover each individual's arrival and departure time for work. Traffic congestion is measured using GPS sensors installed in vehicles circulating around the city, and proxied by the inverse of the average speed. My identification strategy exploits within-district daily variation in traffic congestion. I complement this approach using road accidents as an exogenous shifter of traffic congestion.

I find that traffic congestion increases the time workers spend at work. The magnitude is economically relevant. In a single day, doubled traffic congestion lengthens the workday by one hour. This effect is driven by congestion during the afternoon rush hour. I also find that workers adapt to traffic congestion in the sense that individuals working in more congested areas are less affected than individuals working in less congested areas.

The positive effect of traffic congestion on hours at work is robust. Replacing hours at work using smartphone data with self-reported hours worked from household surveys does not affect the results. This finding is also robust to using road accidents as a measure of exogenous variation in traffic congestion.

I find that individuals stay longer at work primarily because they delay their departure time. These results are consistent with the bottleneck model where one may choose when to start their commute in response to congestion. These results may also suggest the presence of labor market frictions that prevent workers from arriving late to work or departing earlier. Hence, a potential mechanism for this effect is that workers respond to traffic congestion by departing later from work, despite starting at the same time or earlier.

Even though workers stay longer at work, labor income does not seem to increase. I find suggestive evidence that workers are not paid more. One potential explanation may be that workers stay longer one day at work in response to congestion, but they compensate by leaving earlier another day, therefore creating minimal change to the total hours of work over a given week or month. However, I do not find evidence of that compensating behavior in the short run. Alternatively, workers may be rewarded in the future for their longer hours today. However, this cannot be explored in this study due to data limitations.

This study contributes to a broader literature analyzing the effects of commuting costs on labor supply. In these studies, commuting costs are usually measured by changes in distance (Fu and Viard, 2019, Gutiérrez-i-Puigarnau and van Ommeren, 2010) or changes in commuting time (Black et al., 2014, Gutiérrez-i-Puigarnau and van Ommeren, 2015). However, we cannot attribute results from those studies to changes in traffic congestion. This study also contributes to the literature on environmental outcomes and labor supply. Previous literature indirectly addresses the relationship between traffic congestion and work-time allocation, investigating how driving restriction policies affect leisure time (Viard and Fu, 2015). One contribution of the present study is to add traffic congestion as a new variable of interest. Second, this paper uses novel “big data” from smartphones to track individuals’ daily time allocation, particularly, the number of hours at work, and work arrival and departure times. With these new sources of data, I can directly study the relationship between traffic congestion and work-time allocation. Finally, this study is related to the literature about the value of time (Becker, 1965, Bento et al., 2020, Wolff, 2014).

The remainder of the paper is organized as follows. Section 1.2 describes the conceptual framework. Section 1.3 describes the data used to measure hours spent at work and traffic congestion. Section 1.4 discusses the empirical approach and identification concerns. Section 1.5 describes the results. Section 1.6 presents the discussion. Finally, section 1.7 concludes.

1.2 Conceptual Framework

There are two main models to understand the relationship between traffic congestion and labor supply: (i) the bottleneck model, and (ii) the standard neoclassical model of labor-leisure choice with commuting costs.

The bottleneck model (Arnott et al., 1990, 1993, Noland and Small, 1995, Small, 1982, Vickrey, 1969) allows individuals to choose when to start their commute to respond to congestion. Hence, individuals may choose to leave earlier from home to avoid the morning rush hour or delay their departure time from work to avoid the afternoon rush hour. Consequently, congestion may change the number of hours allocated to work. However, this model has not been yet used to study the effect of congestion on labor supply. It is focused mainly on the morning commute and on the “schedule delay” which is the difference between arrival time to work and some ideal time that usually coincides with the time work starts (*e.g.* 9 am). The model uses the schedule delay to measure the social welfare loss due to congestion (Kim, 2019).

On the other hand, traffic congestion can be seen as a shifter of commuting costs. Black et al. (2014) introduce commuting time costs in the labor supply model. In this model, traffic congestion increases commuting time costs, and this increases the value of leisure relative to the value of working. This effect may push some individuals to work fewer hours or to exit the labor force. However, in a two-person household, if the labor supply of one of the members is negatively affected by the increase in commuting costs, the household will face a negative income shock. Then the other member increases the time allocated given that leisure is assumed to be a normal good. The effect on the overall labor supply is ambiguous, but the negative effect on labor force participation is unambiguous. In a similar fashion, Gutiérrez-i-Puigarnau and van Ommeren (2010) develop a labor supply model with both time and monetary commuting costs. These are variable costs when choosing workdays, but fixed costs when deciding the number of work hours within a day. Given an increase in commuting costs, workers may respond by working fewer days to avoid extra commuting costs but may increase the number of hours worked per day to mitigate a reduction in income. It is again ambiguous which of these effects dominates.

1.3 Data

This paper aims to estimate the effect of traffic congestion on work-time allocation. This requires longitudinal information that links individual hours worked with traffic congestion on a daily basis. Ideally, traffic congestion should be measured on the individual's commuting route considering their preferred mode of transport. Data with such granularity is not available yet. I, therefore, construct a novel longitudinal dataset combining smartphone data that allows me to track the time allocation of individuals with daily traffic congestion that comes from GPS sensors installed in vehicles. The unit of observation is the owner of the smartphone device. I restrict the sample to manufacturing and office workers to reduce measurement error in the outcome variable. In several economic sectors such as retail and services (e.g., leisure, health, and education), I am unable to distinguish between workers and clients, whose labor choices and outside options differ considerably. My final dataset consists of an unbalanced panel of 6,709 observations, representing 1,262 devices for all sixteen districts in Mexico City (CDMX) in 2019.

Tables 1.1 shows the description of the main variables. I approximate hours worked with the number of hours spent at work when using the smartphone data. Table 1.2 shows summary statistics. The number of hours spent at work using the smartphone data is higher than self-reported hours worked using household surveys, on average. The average worker arrives to work around 9 am and departs from work at approximately 7 pm. This pattern occurs either in high- or low-congested districts.

Table 1.1: Variable description and data source

Variable	Description	Data Source
Hours worked	Number of hours spent at workplace	Quadrant
Hours worked	Number of self-reported hours worked	ENOE (INEGI)
Arrival time	Device's first time at work (in 24h format)	Quadrant
Departure time	Device's last time at work (in 24h format)	Quadrant
Traffic congestion	Inverse of average speed (h/km)	Dat's why
Accidents	Number of confirmed road incidents by CDMX 911	Gobierno CDMX
Temperature	Average temperature (in Celsius)	CONAGUA
Precipitation	Rain (in mm)	CONAGUA
Humidity	Relative humidity (in %)	CONAGUA
Daylight hours	Difference between sunshine and sunset times.	CONAGUA

Notes: This table presents the description of the main variables and their corresponding source. All variables are available from January-December 2019.

Table 1.2: Summary statistics

	All	High-congestion districts	Low-congestion districts
	(1)	(2)	(3)
Panel A. Labor outcomes			
Hours worked (daily, mobile data)	10.47	10.45	10.50
Hour worked (daily, survey data)	8.36	8.36	8.37
Arrival time	8.84	8.89	8.79
Departure time	19.32	19.34	19.29
Panel B. Traffic congestion			
Inverse of avg. speed (h/km)	.042	.044	.039
Panel C. Weather			
Temperature (C)	19.96	19.97	19.94
Humidity (%)	63.02	63.06	62.98
Rain (mm)	0.02	0.02	0.02
Daylight (hours)	12.07	12.05	12.10
No. of Smartphones	1,262	671	591
No. of Observations	6,709	3,723	2,986

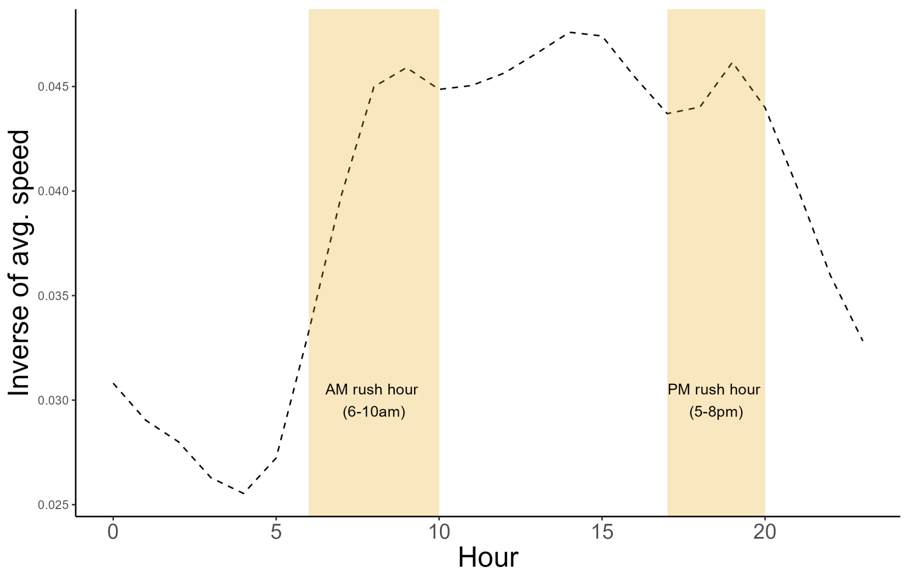
Notes: This table presents mean values for the main variables. Arrival and departure time are in 24 hours format.

Smartphone data. This data is provided by Quadrant, a private organization specializing in high-quality mobile location-based data. The raw data consists of pings (i.e. the time and location of a given smartphone) collected from applications installed in deidentified smartphones. A ping is recorded every time the location of the smartphone is requested by the applications installed on the device. This data provides representative information on the population in Mexico City. Figure A.1 shows that the total population at the district level according to the Census 2020 is correlated with the total number of smartphone devices with an R^2 of 0.62.

I use this data to build a panel of individuals and identify where they work and live, and the number of hours they stay at work. I follow individuals for several days within a week, during all weeks in 2019 except for the first and the last weeks of the year given that patterns in working hours and congestion may be particularly unusual in these two weeks. I combine this data with geocoded establishment-level information to identify workplaces, and with residential areas from census data to identify homes. See the data appendix for details about the algorithm used for this purpose. The richness of this data allows me to know both the time individuals arrive and depart from work. Using this information, I estimate the number of hours individuals stay at work which I use as a proxy of hours worked.

This dataset has three main limitations: (i) It becomes sparse very quickly. The raw semi-unstructured data contains billions of pings per month. However, most of the devices are observed one single day making it difficult to follow them across days. For instance, imposing the structure described in the data appendix to identify workplaces reduces drastically the number of observations. Hence, there is a trade-off between the number of observations and the reliability of the statistics. (ii) This data does not provide information regarding the demographics of the owners of the devices such as gender, age, etc. Socio-economic characteristics can be inferred from the neighborhood of residence or points of interest (POI) visited regularly. (iii) This data does not provide work and home locations, which then need to be inferred using supplemental data.

Figure 1.1: Distribution of traffic congestion



Notes: The figure depicts the distribution of traffic congestion per hour using data from Dat's Why for Mexico City in 2019. Morning (6 am-10 am) and afternoon (5 pm-8 pm) rush hours are highlighted in yellow.

Traffic congestion data. This data is provided by Dat's Why, a private company with the largest real-time Big Data network of smartphones, vehicles, and sensors in Mexico to monitor traffic congestion. The raw data consists of hourly average speed measures at the street segment level in Mexico City for every day of 2019. I use this data to build a district-level panel of daily average speed.

I use the inverse of average speed as a proxy of traffic congestion as in Hanna et al. (2017). In addition to the daily average traffic congestion, I use this data to calculate the traffic congestion during the morning (6 am-10 am) and afternoon (5 pm-8 pm) rush hours for Mexico City. Figure 1.1 shows the distribution of congestion per hour using data from Dat's Why.

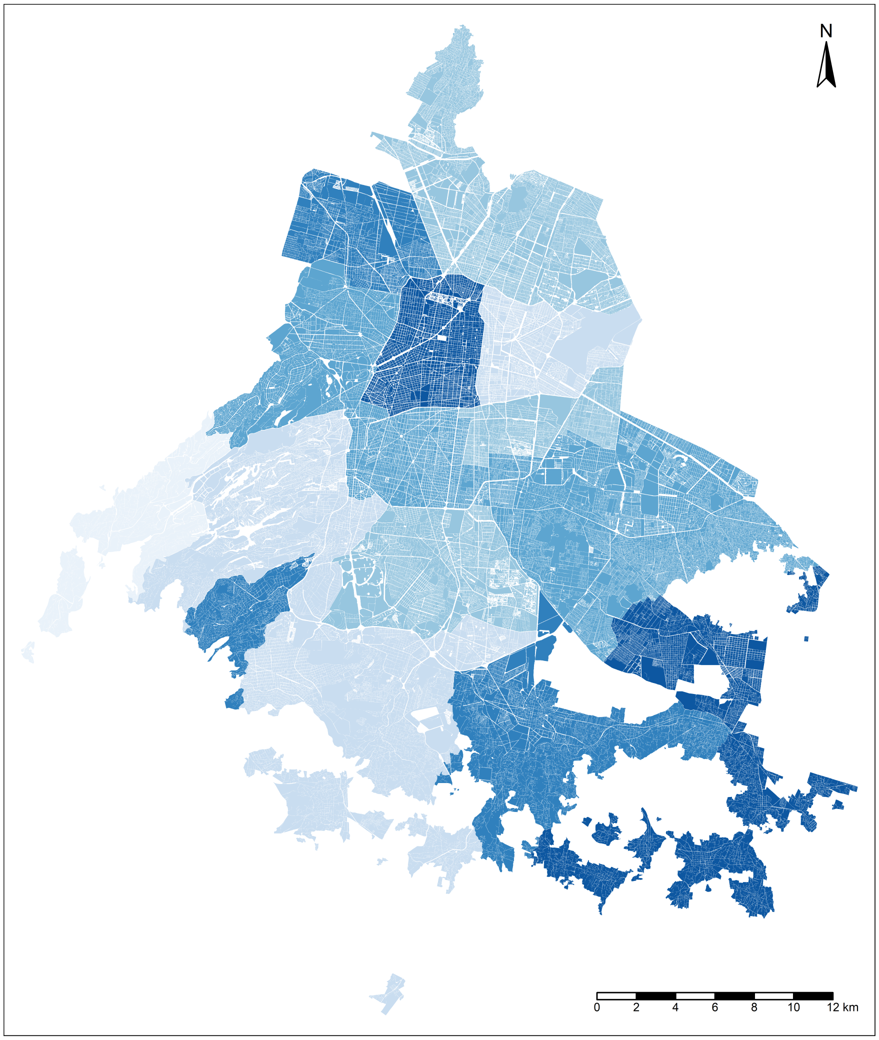
This data is representative of the traffic congestion in Mexico City. Figure A.3 shows the distribution of congestion per hour using aggregated data from Waze reported in Calatayud et al. (2021) and provided by the corresponding authors. We can observe that both distributions in Figure 1.1 and Figure A.3 show a similar pattern of traffic congestion in the city. Both distributions capture the morning and afternoon rush hours for similar hours during the day. The correlation between them is 0.93.

Supplemental data. I complement the smartphone and traffic congestion data with information for the year 2019 about establishments, residential venues, self-reported income and hours worked, weather, daylight hours, and road accidents. I use the National Statistical Directory of Economic Units (DENUE) to obtain the latitude and longitude coordinates for the location of the establishments, the size of the firm, and the economic sector. See the distribution of establishments with 50 workers or more in Figure A.5 in the appendix. I use information from the 2020 Census and the National Geostatistical Framework (MGN) to identify residential venues in the city. To address identification concerns regarding omitted variables related to weather I use information about temperature, precipitation, and relative humidity from monitoring stations. Likewise, I include daylight hours calculated by taking the difference between sunset and sunrise times. I use self-reported hours worked from household surveys (ENOE) as an alternative outcome variable to the hours worked built using the smartphone data. Finally, I use road accidents from administrative records as an exogenous source of variation in traffic congestion to address different identification concerns. The information on accidents is used to build a district-level panel of daily accidents, and accidents that occur close to the border of different districts do not receive a particular treatment. See the appendix for more details about these data sources, and about Dat's Why and Quadrant.

Location This study uses information from Mexico City (CDMX) in 2019. Mexico City is one of the most congested cities in the world. It is more congested than cities such as Mumbai and Delhi in India, and New York in the US (Akbar et al., 2020). For instance, it was the third most congested city in the world in 2019 (INRIX, 2019). Also in 2019, residents lost more than 600 million hours due to congestion representing a cost of more than twice the budget assigned for education in the city (Calatayud et al., 2021). Figure 1.2 displays a map of Mexico City with the average congestion per district.

Mexico City is an ideal setting to study the impacts of traffic congestion on our well-being. It is ideal not only because it is one of the most congested cities in the world (Akbar et al., 2020, INRIX, 2019). Studying the context of a city in a developing country is relevant given current trends in urban population and motorization rates (Akbar et al., 2020, Calatayud et al., 2021, Kreindler, 2022). First, the urban population is growing rapidly. By 2050, approximately 2.5 billion people will migrate to cities in developing countries. This may pressure cities in the developing world

Figure 1.2: Spatial distribution of congestion in Mexico City



Notes: Figure depicts a map of Mexico City with the average annual traffic congestion per district in 2019. The darker the more congested is the district.

where the transportation infrastructure is already outdated to the current population size. Second, private vehicle ownership is also growing rapidly. This is because of increasing motorization rates due to economic growth.

1.4 Empirical Approach

To study the effect of traffic congestion on work-time allocation, ideally, we would like to observe how many hours a person works where there is and there is no traffic congestion on a given day. However, we cannot observe the counterfactual for each person. We can only observe the hours worked either when there is or there is no traffic congestion, but not the hours worked under both scenarios. Alternatively, we can design a randomized controlled trial where, ceteris paribus, we randomly assign traffic congestion to a group of workers (treated group) and no traffic congestion to another group of workers (control group) on a given day. We can then compare the average hours worked between groups to find the average treatment effect. However, traffic congestion cannot be randomly assigned.

Baseline regression To explore the effect of traffic congestion on work-time allocation, I estimate the following regression model:

$$y_{ijt} = \delta_t + W_{jt} + \beta \times \ln(\text{Traffic Congestion})_{jt} + \epsilon_{ijt}, \tag{1.1}$$

where the unit of observation is individual i working in district j in day t. y_{ijt} represents the labor outcome variables such as hours worked, and arrival and departure times from work. Traffic congestion is proxied by the inverse of the average speed as in Hanna et al. (2017). W_{jt} is a set of weather variables that include temperature, precipitation, humidity, and daylight hours. δ_t includes day of the week and month fixed effects. Once we divide it by 100, β can be interpreted as the unit change in the outcome variable when traffic congestion increases by 1%. I estimate the model using OLS and clustering standard errors at the week-district level. Identification comes from assuming that within-district daily variation in traffic congestion is exogenous conditional on weather and fixed effects or from quasi-random (temporal) variation in (demean) traffic congestion across days.

Identification concerns I include weather controls and time fixed effects in the baseline panel regression to reduce concerns regarding omitted variables bias. Traffic Congestion is not randomly assigned and confounders elements in ϵ_{ijt} may be correlated with both traffic congestion and our outcome y_{ijt}. For example, rainy days may be positively correlated with both congestion and hours worked, or darkness of the day may be positively correlated with congestion, and negatively correlated with hours worked. Alternatively, a higher temperature may be negatively correlated with congestion and hours worked. Hence, I control for temperature, precipitation, humidity, and daylight hours in W_{jt}. Likewise, Fridays may be positively correlated with congestion, but negatively correlated with hours worked, or a particular month may experience a decline in business activity that both affect congestion and work-time allocation. Thus, I control for day of the week and month fixed effects. I also address individual time-invariant unobservables by individual fixed effects as part of the robustness checks.

It is likely that my measures of hours worked and traffic congestion contain measurement error. As described in the data section, my measure of traffic congestion seems to represent the patterns regularly observed on the streets of Mexico City. Regarding hours worked, I use self-reported hours worked from household surveys (ENOE) as an alternative outcome variable.

Reverse causality is unlikely in a context where I only follow individuals during a week and estimate short-run effects. Changes in traffic congestion patterns may affect the spatial distribution of economic activities. In response to these changes, residents may re-optimize their decision of where to live, work or consume (i.e. sorting). However, changes in the spatial distribution of economic activities may affect patterns in traffic congestion (**?**). This can be a problem for the long-run effects of congestion on work-time allocation, but not for the short run. Given the focus on the short run, all the analysis is conditional on sorting (i.e. sorting already took place). It is unlikely that we see people shifting residential areas or workplaces across days during the short period of analysis.

Instrumental variable I complement the baseline identification strategy with an instrumental variable approach to address concerns regarding additional potential omitted variables, measurement error in traffic congestion, and reverse causality that may persist. For example, there may be time-variant unobservables that are correlated with congestion and that also affect hours worked. Hence, I complement the identification strategy in the baseline regression with an instrumental variable design. I use road accidents as my instrumental variable as in Beland and Brent (2018). This instrument is relevant and as good as random. Regarding the exclusion restriction, it is likely that accidents are only affecting hours worked via changes in traffic congestion. In this context, accidents introduce exogenous variation in traffic congestion to lessen concerns regarding omitted variable bias, measurement error, and reverse causality.

1.5 Results

1.5.1 Main results

Table 1.3 presents the main results. Column (1) shows the results for the baseline model in equation 1.1 estimated using OLS. The outcome variable is hours worked, approximated by the daily number of hours spent at work constructed using smartphone data. Traffic congestion is measured as the inverse of the daily average speed in the district where individuals work. Column (1) indicates that a ten percent increase in traffic congestion increases time at work by 0.13 hours. The estimated coefficient is statistically significant at the five percent level. Identification in equation 1.1 may be affected by individual time-invariant unobservables. For this reason, I present a specification demonstrating robustness to including individual fixed effects (column (2)). The estimated coefficient is similar in magnitude and statistically significant at the five percent level. The estimates in columns (1) and (2) are similarly positive, but the larger coefficient in column (2) indicates that omitting individual fixed effects generates a negative bias on the estimated coefficient. Note, however, that by including individual fixed effects, the estimates in column (2) exploit the variation of an unbalanced panel where the number of observations per device is skewed to the left.

Column (3) shows results using a 2SLS approach where road accidents act as an instrumental variable for traffic congestion. These results reduce concerns regarding omitted variable bias, mea-

Table 1.3: The effect of traffic congestion on hours worked

	Dependent variable: Hours worked		
	(1)	(2)	(3)
Traffic Congestion (log)	1.277**	1.744**	1.769**
	(0.518)	(0.778)	(0.894)
Method	OLS	OLS	2SLS
Individual FE	No	Yes	No
Observations	6,333	6,271	6,307
R-squared	0.087	0.763	0.004

Notes: Standard errors clustered at the week-district level in parentheses. Statistical significance at the one, five, and ten percent levels is indicated by ***, **, and *, respectively. Traffic congestion is measured as the inverse of the daily average speed at the district level. All regressions use smartphone data and include control variables for weather (average daily temperature, precipitation, and humidity) and number of daylight hours, and day-of-week and month fixed-effects. The sample considers manufacturing and office workers only. In column (3), the Kleibergen-Paap rk Wald F-statistic is 157.96, and the first-stage coefficient is 0.01 and statistically significant at the one percent level.

surement error in traffic congestion, and reverse causality. The instrumental variable satisfies the relevance condition. The first-stage estimated coefficient is 0.01 and is statistically significant at the one percent level. The exclusion restriction is also plausible, as it is unlikely that individuals change the number of hours worked in response to road accidents for reasons other than to avoid related traffic congestion. Column (3) indicates that a ten percent increase in traffic congestion increases time at work by 0.18 hours. The estimated coefficient is statistically significant at the five percent level. This is higher than the estimated coefficient in column (1), most likely because the instrumental variable is correcting for measurement error. I obtain a qualitatively similar result when including individual fixed effects. While the sign and magnitude of the coefficient stay the same, the estimates with individual fixed effects do lose a lot of precision (see column (2) in Table A.1).

Previous studies have not always found a positive effect of commuting costs on labor supply. Commuting costs are usually measured by changes in distance or changes in commuting time. One study found that increasing commuting distance (Fu and Viard, 2019) reduced labor supply in China, while another found a negative effect of commuting time (Black et al., 2014) on female labor force participation in the US. Data from Germany and the UK has been used to show that increasing commuting distance (Gutiérrez-i-Puigarnau and van Ommeren, 2010) or commuting time (Gutiérrez-i-Puigarnau and van Ommeren, 2015) increases the number of hours worked. Another study indirectly addresses the relationship between traffic congestion and labor supply by investigating how changes in driving restriction policies affect hours of leisure time (Viard and Fu, 2015).

Those authors find a positive effect of driving restrictions on leisure time for individuals who are self-employed and a negative effect for workers making hourly wages.

The magnitudes of the estimated coefficients in Table 1.3 are not small. Gutiérrez-i-Puigarnau and van Ommeren (2010) find that doubling commuting distance increases labor supply by approximately 15 minutes per week, which is equivalent to 13 hours per year. However, I find that doubling traffic congestion increases hours worked by one hour *per day*, which is equivalent to five hours per week or 260 hours per year. Further, we cannot attribute results from those previous studies to changes in congestion (since changes in commuting time can be a result of a less direct route or changes in the commuting distance without changes in traffic congestion), whereas I am able to do so in this paper.

Table 1.4: Robustness checks

	Dependent variable: Hours worked		
	(1)	(2)	(3)
Traffic Congestion (log)	1.375***	0.120**	1.238***
	(0.487)	(0.049)	(0.425)
Change in specification	Week FE	Outcome in log	All sectors
Observations	6,333	6,333	15,870
R-squared	0.110	0.076	0.082

Notes: Standard errors clustered at week-district level in parentheses. Statistical significance at the one, five and ten percent levels is indicated by ***, ** and *, respectively. All regressions use the baseline model and smartphone data.

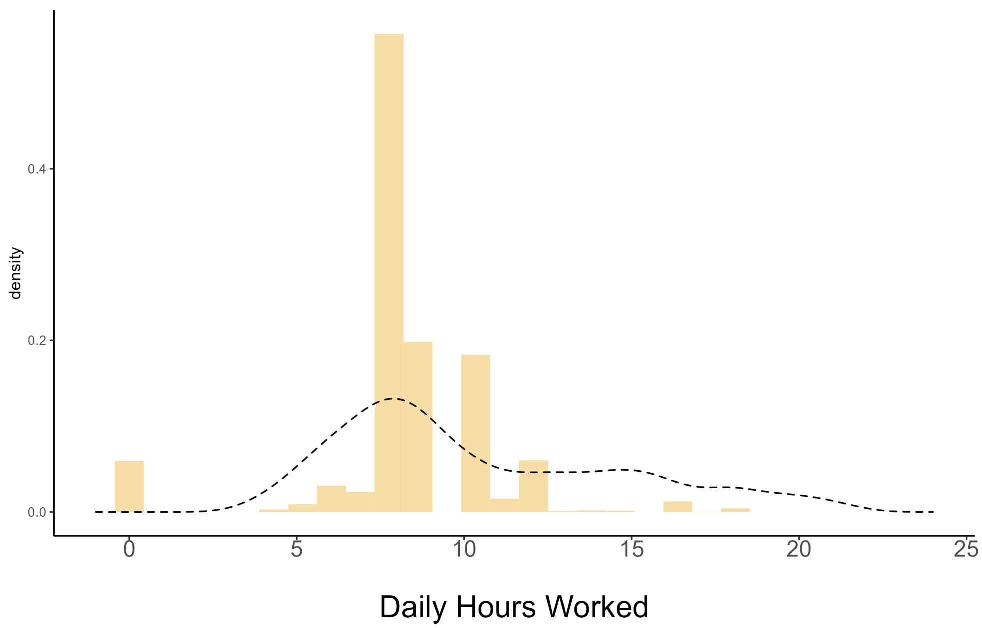
Robustness checks Table 1.4 displays a robustness analysis of the main results to alternative specifications. Column (1) replaces month fixed effects with week fixed effects in the baseline specification. Controlling for seasonality at a finer level in this way yields similar results. Column (2) replaces the outcome variable (hours worked) in levels from equation 1.1 with a log transformation, meaning that the coefficient in column (2) represents an elasticity. A one percent increase in congestion increases hours worked by 0.12 percent. If congestion doubles in a day, then hours worked increase by 12 percent. This is equivalent to a one-hour increase, considering that the average individual works around ten hours (as reported in table 1.2). Column (3) considers all sectors in the economy as opposed to only manufacturing and office workers. I restrict the sample to manufacturing and office workers in the baseline model to reduce measurement error in the outcome variable. In several economic sectors such as retail and services (e.g., leisure, health, and education), I am unable to distinguish between workers and clients, whose labor choices and outside options differ considerably. However, column (3) shows that results are robust including all economic sectors.

Additional robustness analyses are reported in appendix table A.1.

1.5.2 Using labor household surveys

This section presents results using self-reported hours worked from household labor surveys (ENOE) as the outcome variable of interest. ENOE is the main labor market household survey in Mexico and provides both monthly and quarterly data. The National Statistics Office (INEGI) collects information on individuals aged 15 and above on a continuous basis throughout the year. ENOE has a rotating panel design where one household can be followed for five consecutive quarters. The quarterly sample size is around 126,000 housing units. ENOE is representative of the country and cities such as Mexico City.

Figure 1.3: Distribution of daily hours worked: ENOE vs. Smartphone data



Notes: The figure depicts self-reported daily hours worked from the labor household survey ENOE (in yellow) and daily hours spent at the workplace from the smartphone data (dashed line). Data are limited to individuals working in manufacturing or as office workers in firms with more than 50 employees. The first week of January and the last week of December are excluded, as well as all Saturdays and Sundays.

Figure 1.3 compares the distribution of hours worked between the official household labor survey (ENOE) and the smartphone data. The sample consists of manufacturing or office workers in firms with more than 50 employees. Data from the first week of January and the last week of December are excluded, as well as for Saturdays and Sundays. Both distributions visually represent the same overarching patterns of work time. However, using a Kolmogorov-Smirnov test I am able to reject the null of equality of distributions. Hours worked from ENOE are relatively highly concentrated around 8 hours compared to the distribution from the smartphone data. ENOE also reports zero hours worked, which contrasts with the smartphone data where all individuals work a positive number of hours by design. These differences may suggest the presence of measurement error in hours worked from the smartphone data. Hence, I present the results using self-reported hours worked from ENOE as a robustness test.

Table 1.5 shows that results are similar after replacing the smartphone data with self-reported hours worked from household labor surveys (ENOE). Column (1) contains the baseline results from table 1.3. Column (2) shows estimated coefficients using daily hours worked reported in ENOE as the dependent variable of interest. In column (1), traffic congestion is measured in the district where the individual works. In column (2), it is measured instead at the district where they live. This is relevant since 41.6% of individuals in Mexico City work in a district other than where they reside according to the 2015 intercensal survey. The estimated coefficients in both columns are similar in terms of magnitude and statistical significance, allaying concerns over measurement error in hours worked using the smartphone data.

Table 1.5: The effect of traffic congestion on hours worked using smartphone and labor survey data

	Dependent variable: Hours worked	
	(1)	(2)
Traffic Congestion (log)	1.277**	1.018***
	(0.518)	(0.386)
Method	OLS	OLS
Individual FE	No	No
Labor data source	Phone data	ENOE survey
Congestion measured in:	Workplace	Residence
Observations	6,333	7,219
R-squared	0.087	0.014

Notes: Standard errors clustered at week-district level in parentheses. Statistical significance at the one, five, and ten percent levels is indicated by ***, ** and *, respectively. All regressions use the baseline model.

1.5.3 Rush hour and bottleneck model

In this section, I explore whether the main results are consistent with the bottleneck model. In this model, the departure time decision is endogenous. Individuals may choose when to start their commute in response to congestion, e.g., they may choose to leave earlier from home or delay their departure time from work to avoid the morning and afternoon rush hours, respectively. To conduct this exploration, I estimate the baseline model breaking traffic congestion into congestion in the morning and afternoon rush hours. I define the morning rush hour as from 6 am to 10 am, and the afternoon rush hour as from 5 pm to 8 pm.

Table 1.6 shows that the positive effect of traffic congestion on hours worked is driven by traffic congestion during the afternoon rush hour. In column (1), the outcome variable is daily number of hours spent at work (constructed using the smartphone data). There is a positive and statistically significant effect of traffic congestion during the afternoon rush hour on hours worked, but a negative and statistically insignificant coefficient for traffic congestion during the morning rush hour.

Table 1.6: The effect of rush hour traffic congestion on hours worked, arrival time to work, and departure time from work

	Dependent variable:		
	Hours worked	Arrival time	Departure time
	(1)	(2)	(3)
Traffic congestion (log)			
AM rush hour	-0.274	-0.189	-0.462
(6-10am)	(0.572)	(0.297)	(0.380)
PM rush hour	1.564***	-0.407	1.157***
(5-8pm)	(0.536)	(0.254)	(0.378)
Observations	6,243	6,243	6,243
R-squared	0.089	0.101	0.041

Notes: Standard errors clustered at week-district level in parentheses. Statistical significance at the one, five, and ten percent levels is indicated by ***, **, and *, respectively. All regressions use the baseline model and smartphone data.

Columns (2) and (3) in table 1.6 show that individuals are spending more time at work due to delayed departure times. Results suggest that traffic congestion has a negative effect on the time workers arrive to work. However, the estimated coefficients are not statistically significant. Instead, traffic congestion, particularly during the afternoon rush hour, delays the time that individuals depart from their jobs. The estimated coefficient is statistically significant at the 1 percent level.

These results are consistent with the bottleneck model. I find evidence that workers are spending more time at work because they are delaying departure to avoid the afternoon rush hour. These results may also suggest the presence of labor market frictions that prevent workers from arriving late to work or departing earlier. Hence, the mitigation strategies available to them result in longer hours.

1.5.4 Mitigation and adaptation

Table 1.7 shows evidence that workers do not mitigate the effect of traffic congestion via intertemporal labor substitution. One way to investigate this is to regress weekly hours worked (using the smartphone data) on weekly traffic congestion. Column (1) shows the estimated OLS coefficient exploiting cross-sectional variation across weeks and districts. Evidence of intertemporal labor substitution would be supported by an estimated coefficient close to zero, reflecting the idea that workers compensate for extra time at work one day by working less on another day. However, I do not find evidence supporting such compensatory behavior. I find a positive and statistically significant effect of weekly traffic congestion on weekly hours worked. The size of the coefficients suggests a cumulative effect of the single-day effect across business days. This result also goes in line with the presence of labor rigidities. In a context where there are frictions preventing workers from leaving earlier, it is unlikely that we can observe intertemporal substitution of time allocated to work.

Table 1.7: Mitigation and adaptation to traffic congestion

		Hours worked (daily)	
	Hours worked (weekly)	High-congestion districts	Low-congestion districts
	(1)	(2)	(3)
Traffic congestion (weekly, log)	6.446***		
	(1.778)		
Traffic congestion (daily, log)		-0.525	2.804***
		(1.382)	(0.650)
Observations	2,671	3,508	2,825
R-squared	0.074	0.058	0.150

Notes: Standard errors clustered at week-district level in parentheses. Statistical significance at the one, five, and ten percent levels is indicated by ***, **, and *, respectively. All regressions use the baseline model and smartphone data. High-congested districts have traffic congestion above the median for all of Mexico City.

Table 1.7 also shows that workers adapt to traffic congestion. Columns (2) and (3) present estimates for the baseline model where I separate districts by the intensity of traffic congestion. In column (2), I consider only the sub-sample of individuals working in high-congestion districts. In column (3), I include only individuals working in low-congestion districts. I define a district as highly congested if its congestion is above the median traffic congestion for the entirety of Mexico City. Results indicate that traffic congestion has no effect on hours worked for individuals working in high-congestion areas. On the other hand, traffic congestion has a positive and statistically significant effect on hours worked for individuals working in low-congestion areas. I interpret these results as evidence of adaptation. Individuals working in high-congestion areas are less sensitive to shocks in traffic congestion. Instead, individuals working in low-congestion areas are more affected by shocks in traffic congestion in terms of increased work time.

1.5.5 Labor income

If individuals are staying longer hours at work, are they getting paid more as well? Table A.4 reports the effect of traffic congestion on labor income. The outcome variable, monthly self-reported labor income, is measured in logs and it comes from ENOE household surveys. Traffic congestion is aggregated at the monthly level and is also measured in logs. Estimated coefficients are therefore elasticities. The number of observations decreases due to the monthly aggregation, as well as because a large proportion of respondents do not provide income information in the survey. Column (1) shows OLS estimates pooling all individuals and including month fixed effects. Column (2) exploits the fact that some individuals were interviewed in multiple months by adding individual fixed effects. Column (3) presents results using 2SLS instrumenting for traffic congestion with monthly road accidents.

I do not find evidence of increased income as a result of increased hours at work. The results from table A.4 suggest that traffic congestion is not increasing labor income. Hence, individuals are staying longer at work without earning more income. Note that one potential explanation is that workers are rewarded in the future. However, this cannot be explored in this study due to data limitations.

1.6 Discussion

Road accidents Figure A.7 in the appendix suggests that workers react to shifters of congestion, namely road accidents, that happen before the afternoon rush hour. It displays the coefficient estimates and 90% confidence intervals of reduced form regressions. Estimated coefficients are reported in Table A.3. Panel (a) in Figure A.7 shows the results of regressing the number of accidents on the number of hours spent at work. I split the number of accidents into three categories: (i) before the morning rush hour, (ii) during the morning rush hour and before the afternoon rush hour, and (iii) during and after the afternoon rush hour. Panel (b) shows similar results using departure time as the outcome variable instead. We can observe that the estimated coefficient associated with the number of accidents between 6 am and 5 pm is not statistically significant. This suggests that workers react to accidents that happen before the afternoon rush hour. However, the magnitude of all coefficients is similar. Nonetheless, another piece of evidence suggests that workers can react to accidents during the day. In 2019, the navigation company Waze was already operating in the city and had two million active users monthly.

Labor supply To what extent the time spent at work measured with smartphone data is capturing labor supply? Time at work does not always equal work time. For example, individuals may stay one hour longer at work, but they may be partially working that hour or not working at all. Moreover, individuals may leave work and continue working at home. Hence, in principle, it seems that "hours spent at the workplace" is far from capturing labor supply. Moreover, we observe only an equilibrium outcome of the supply and demand in the labor market.

Table 1.5 suggests the "hours spent at the workplace" are not far from capturing hours worked. Columns (1) and (2) present the main results using smartphone data and ENOE household surveys. We can observe that both estimated coefficients are similar in magnitude, and the coefficient using ENOE is slightly lower. We can take this as evidence that hours stay at work slightly overestimates hours worked. Workers may be staying extra time at work, and part of that time is allocated to work. Figure A.6 may also indicate that hours spent at work are related to hours worked. If output per labor hour is increasing with congestion, and hours worked are increasing with congestion, this means that output should also be increasing in congestion. If people were only shirking at work, then we would not observe an increase in output. If the extra hours at work were artificial, then we should not see an increase in productivity.

In our short-run setting, it is unlikely that labor demand factors play a role. Therefore, we could use changes in hours worked to approximate shifts in labor supply. However, differences between the number of hours spent at work and hours worked reported above suggest being cautious and interpreting the results as changes in work-time allocation instead of labor supply.

Labor productivity If individuals are staying longer hours at work, are they producing more output per labor hour? Figure A.6 presents suggestive evidence that this may be the case. It displays the correlation between monthly traffic congestion and monthly labor productivity. Labor productivity is calculated as total output value divided by total hours worked using information from manufacturing firm surveys (EMIM) in 2019. However, this evidence should be taken with caution given that we cannot find a causal relationship between traffic congestion and labor productivity from Figure A.6. Moreover, this relationship may seem counterintuitive as one may expect that if traffic is inducing individuals to work more hours, workers are not as productive during these extra hours as they were during earlier hours. Hence, the positive correlation may be explained by other factors such as seasonality. Labor productivity and traffic congestion may be higher in particular months of the year (*e.g.* December). EMIM collects more granular firm-level data. Unfortunately, only aggregate numbers used in Figure A.6 are publicly available. This exercise can be replicated in the future when access to the fully EMIM microdata is provided.

Welfare and Inequality I find evidence that workers are staying more hours at work due to congestion. However, suggestive evidence implies that workers are not earning more for this extra hour. Regardless of whether individuals are conducting actual work or shirking during that extra hour, there is evidence that this time is not being remunerated. This extra hour represents 10 percent of the average shift reported in Table 1.2. Hence, traffic congestion may be reducing the welfare that individuals obtain from participating in the labor market by 10 percent. However, we first need to test whether this extra hour at work is not rewarded in the medium or long term. On the other hand, traffic congestion may be a shifter of inequality. Workers are staying more time at work without being compensated for it, but firms are enjoying more labor productivity.

1.7 Conclusion

Traffic congestion is a major and yet unsolved concern in most cities in the world. This paper studies the effect of traffic congestion on work-time allocation, a previously unquantified externality. I exploit highly granular smartphone data to measure daily work-time allocation, including arrival and departure times from work. I combine these data with daily exposure to traffic congestion measured using GPS sensors installed in circulating vehicles in Mexico City in 2019. My identification strategy exploits within-district daily variation in traffic congestion. I complement this approach using road accidents as an exogenous shifter of traffic congestion.

The results suggest that traffic congestion increases time allocated to work. Facing twice as much traffic congestion leads to an additional hour spent at work. This finding is robust to using self-reported hours worked from household surveys, as well as to using road accidents as an instrumental variable for congestion. I find that workers stay longer largely because they delay their departure time from work to avoid traffic congestion during the afternoon rush hour. Moreover, workers seem to respond to congestion shifters (i.e, accidents) that occur before the afternoon rush hour. I do not find evidence that workers mitigate the effect of traffic congestion through intertemporal labor substitution; for example, a worker who stays longer today does not compensate by leaving work early tomorrow. I do find evidence of adaptation in the sense that individuals working in high-congestion areas are less affected than individuals working in low-congestion areas. I also find suggestive evidence that workers are not earning more even though they are staying longer hours at work, but labor productivity is increasing. However, this study has some limitations. It is focused only on the short-run effects of traffic congestion on work-time allocation. The findings are also silent on modes of transportation, which may be another channel that individuals use to avoid traffic congestion.

Staying longer hours at work has detrimental effects on well-being, with wide implications for human health, productivity, and the quality of leisure time (e.g., time spent on hobbies and with those we love). By prompting people to stay longer at work, traffic congestion may be mitigating the substantial benefits that cities offer to workers. In addition, rescheduling the timing of activities has important welfare effects (Small, 1982). Time is the ultimate finite resource, which puts time allocation at the heart of the human experience. In our setting, doubled traffic congestion may reduce the welfare that individuals obtain from participating in the labor market by 10 percent.

Hence, the externality of traffic congestion on work-time allocation likely has major impacts on broader well-being.

This study is an example of the use of smartphone data to study human behavior. Smartphone data have huge potential. More research is needed to investigate the representativeness of these data to the whole population, their statistical reliability, and possible synergies with household- and firm-level surveys to learn more and better our behavioral patterns and new developments in the labor markets, such as the great resignation. Further research is also needed to understand long-term effects and to explore the role of modes of transportation. Likewise, future research should address the effect of traffic congestion on productivity across all sectors of the economy.

Chapter 2

Nodody's Watching: COVID-19 impacts on the Amazon Rainforest

2.1 Introduction

Reducing deforestation has the potential to mitigate around one-third of global human-caused carbon emissions (Shukla et al., 2019). Deforestation is most prevalent in developing countries' tropical forests (Burgess et al., 2012, Jayachandran, 2013), however, its effects will be felt globally through climate change and biodiversity loss. Therefore, understanding the factors facilitating deforestation is crucial to curb its effects.

This paper examines the effect of the COVID-19 pandemic on Amazon deforestation, using data from Peru. It provides insights into how a developing country's capacity for environmental monitoring and enforcement was constrained by the pandemic, and how these constraints damaged environmental outcomes such as forest conservation. We provide new evidence of the impact of COVID-19 on deforestation and shed light on the underlying mechanisms driving the estimated causal impact. Peru is an ideal context for such study because is one of the countries hardest hit by the pandemic and places fourth worldwide among countries with the largest extension of tropical forests.

In our empirical setting, we build a rich district-level panel dataset with information about annual deforestation covering the period 2015-2020. Deforestation is derived from high-resolution Landsat satellite imagery. We then exploit two sources of variation. We first exploit the time variation in deforestation before and after the pandemic in an event study design as our baseline specification. Next, we complement this approach with a difference-in-difference design that exploits the inter-district variation in COVID-19 cases and deaths.

The empirical analysis yields several significant findings. First, we observe a substantial increase in deforestation during the COVID-19 pandemic. Compared to pre-pandemic levels, deforestation in Peru rose by approximately 35%, resulting in a national forest loss of 54 thousand hectares. This finding holds true across various identification strategies, providing robust evidence of the impact.

Second, this surge in deforestation carries significant costs. In 2020 alone, COVID-19-induced deforestation led to emissions exceeding 17 million $tCO_2 - eq$ at the national level. This amounts to an additional social cost of US$131.38 million, three times the budget allocated for forest management in Peru in 2019.

Third, we document a potential mechanism that explains the pandemic's impact on deforestation, namely a decrease in forest monitoring efforts coupled with an increase in illegal deforestation activities. Investments in forest monitoring declined in 2020 at both national and regional levels. Moreover, we observe a spike in illegal activities related to coca leaf production and mining during the same period. Our analysis of heterogeneous effects reveals that districts engaged in coca production or characterized by informal or illegal mining experienced exacerbated levels of deforestation. Moreover, the variation in our deforestation outcome seems to be driven by illegal deforestation, since legal logging activities actually decreased in 2020.

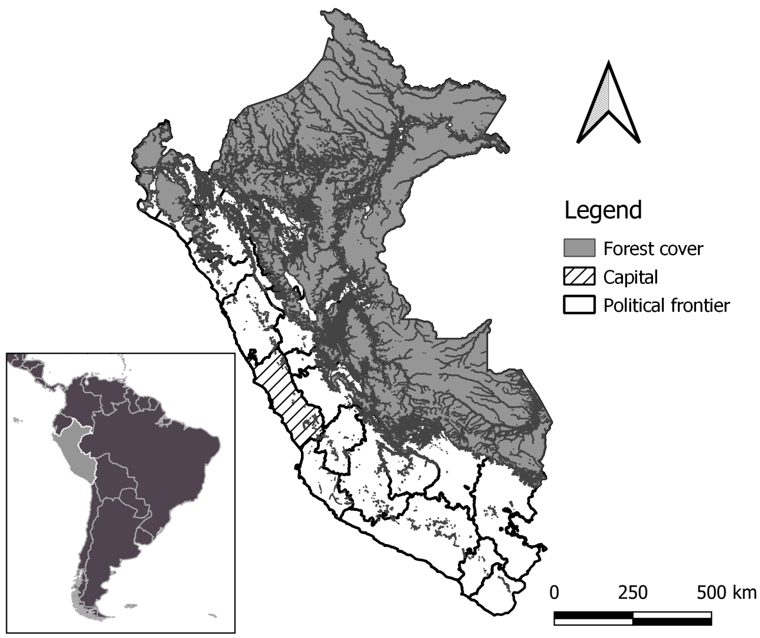
This study contributes to the literature on COVID-19 and environmental outcomes. Previous studies have analyzed the pandemic's impacts on air quality (Blackman et al., 2023, Brodeur et al., 2021, Dang and Trinh, 2021), wildlife (Madhok and Gulati, 2022), and environmental regulation (Vale et al., 2021). Regarding the impacts of COVID-19 on deforestation, most studies trace potential impacts based on theoretical models (Wunder et al., 2021) or descriptive analysis (Brancalion et al., 2020, Lopez-Feldman et al., 2020). The closest in spirit to our study is Saavedra (2020). It uses a difference-in-difference approach to study the effect of national-level lockdowns on deforestation using 70 countries, and it finds no statistically significant effects overall. However, the outcome variable in this study ("vegetation cover change alerts" instead of deforestation) is prone to measurement error that may be attenuating the statistical significance. Moreover, the outcome was measured between January 1, 2019 and July 12, 2020. This disregards a great part of the dry season in the Amazon region (usually between June and November) when slash-and-burn practices are intensified, given the higher prevalence of environmental conditions favoring the flammability of fallen forests (Aragao et al., 2008).

The rest of the paper is organized as follows. Section 2.2 describes the background. Section 2.3 describes the data and the empirical approach. Section 2.4 presents the main results, robustness checks, and heterogeneous effects. Section 2.5 presents a discussion about the mechanisms and the social cost of deforestation. Section 2.6 concludes.

2.2 Background

We use the context of Peru, one of the countries hardest hit by COVID-19 (Higa et al., 2022) and one with the largest extension of tropical forests (Keenan et al., 2015). Peru holds one of the highest COVID-19 mortality rates worldwide, even above countries such as Brazil and India. As of July 2021, there were 2 million COVID-19 cases and more than 600 COVID-19-related deaths per 100,000 inhabitants. Peru implemented stay-at-home orders and social distancing at the start of the pandemic in March 2020. Similar to other countries, the scope of the social restrictions has fluctuated in response to demands to open the economy and based on the number of COVID-19 cases. On the other hand, more than half of Peru's territory (53%) is covered by rainforests. However, on average, more than 128 thousand hectares were annually deforested nationwide between 2001 and 2019, which is equivalent to losing more than 20 soccer fields every hour. Forests in Peru are threatened by activities related to commercial agriculture, gold mining, coca production, and cattle ranching, among others (Finer and Novoa, 2017, Piotrowski, 2019). Figure 2.1 displays the location of rainforests in the territory.

Figure 2.1: Forest cover in Peru, 2020



Source: MINAM, 2022

In Peru, the Law of Forestry and Wildlife (FWL) constitutes the main policy oriented to guarantee the sustainable provision of the benefits generated by the forests. In the last two decades, two FWLs have been introduced in the country. The first one was enacted in 2000 and started to be enforced in 2001. Nevertheless, its numerous reforms were found to be insufficient to effectively halt deforestation across the country (Sears and Pinedo-Vasquez, 2011). As a consequence, a second FWL was introduced in 2011 and enforced since 2015, after a long process of consultation with several groups, including indigenous communities and other stakeholders involved in the forestry sector (e.g., mestizo farmers and small- and medium-sized companies). One of the main reforms introduced by the new law was the creation of SERFOR (Forestry and Wildlife National Service) as the national ruling agency of the forestry sector, which operates jointly with regional forestry authorities. This has boosted the country's capacity to regulate forestry activities across the territory and track forest-related faults. Given the institutional landmark that the SERFOR creation represents in terms of the country's capacity for monitoring deforestation, we focus our analysis on the period after the new FWL, that is from 2015 onwards.

2.3 Data and Methods

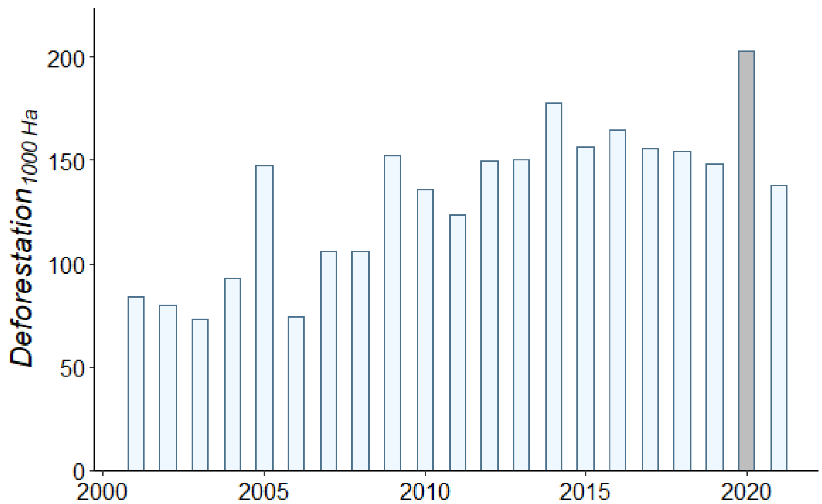
This paper explores the effects of COVID-19 on deforestation. We rely on the official annual forest cover loss data collected by the Ministry of Environment in Peru, which has been available since 2001 (MINAM, 2022). This data is derived from interpreting Landsat satellite imagery with high spatial resolution of up to 30 meters. This fine resolution allows the detection of small-scale deforestation and the distinction between clearing and forest degradation due to natural events (Potapov et al., 2014, Vargas et al., 2021). This data mainly accounts for forest loss due to anthropogenic factors, given the efforts of the Ministry of Environment to exclude natural forest degradation. Likewise, potential forest-cover confusions (e.g., plantations) and false positives as a consequence of prediction errors are taken into account in this data. The Ministry of Environment conducts several validation exercises to reduce systematic measurement error in the forest cover loss measurement. These validation exercises include the use of RapidEye imagery with 5 meters resolution. The information is available only for the 400 districts with tropical forest coverage, about 20% of the 1896 districts in the country. Figure 2.2, Panel (a), displays the number of deforested hectares by year and highlights that 2020 was the year with the largest deforestation on record. On the other hand, we use COVID-19 data collected by the Ministry of Health in Peru (MINSA, 2022). This data contains information about the number of COVID-19 cases and deaths caused by COVID-19 in each district in the country. A description of the main variables used in the analysis and summary statistics are provided in Tables 1.1 and 1.2, respectively, in the Appendix.

Baseline approach Figure 2.2, Panel (b), shows a positive correlation between the number of COVID-19 cases and the deforestation in each district, indicating a possible effect of the pandemic on deforestation. To study this potential effect, we use an event study and estimate the following model:

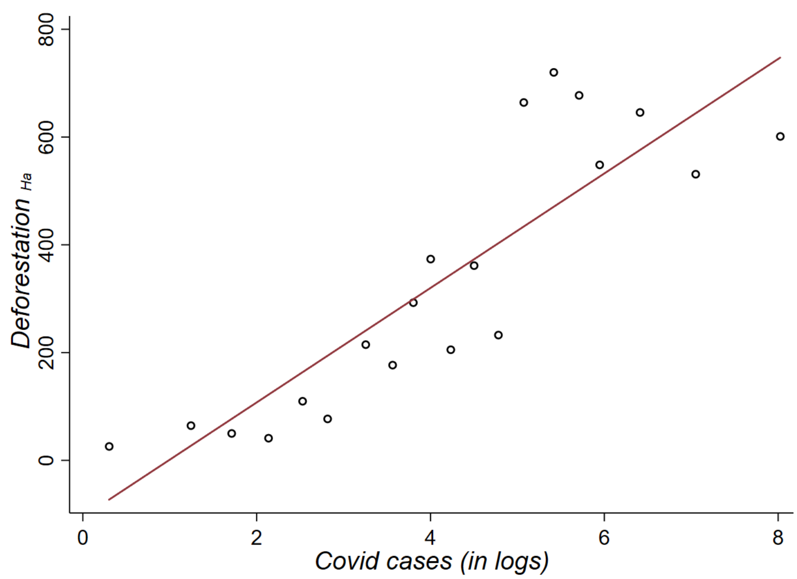
$$y_{it} = \sum_{Q=0}^{1} \beta_Q D_Q + \gamma_i + \epsilon_{it} \tag{2.1}$$

where the unit of observation is district i in year t. y_{it} represents the deforestation outcome variable. D_Q is an indicator variable that equals one when the year is 2020 and zero otherwise. The omitted category is the pre-pandemic year 2019. γ_i includes district fixed effects. Standard errors are clustered at the district level. Our identification strategy exploits time variation across years: we compare changes in deforestation before (2019) and after (2020) the pandemic. We also conduct robustness checks by extending the pre-pandemic period to 2015-2019.

Figure 2.2: Deforestation and COVID-19



(a) Annual country-level deforestation, 2001-2021



(b) Correlation between COVID-19 and deforestation

Notes: Panel (a) depicts the annual deforestation ('000 ha) from 2001-2021. Panel (b) depicts a binscatter with deforestation (ha) on the vertical axis and COVID-19 infections (in logs) on the horizontal axis. Information comes from MINAM (2022), MINSA (2022).

Difference-in-Difference design We complement our baseline regression using a difference-in-difference approach. Specifically, we estimate the following model using our panel of districts:

$$y_{it} = after_t + \text{COVID-19}_i + \beta(\text{COVID-19}_i * after_t) + \epsilon_{it} \tag{2.2}$$

where the unit of observation is district i in year t. y_{it} is the deforestation outcome variable. $after_t$ is a dummy variable that equals one if the period corresponds to 2020 or zero if the period encompasses the years 2015-2019. COVID-19$_i$ is our treatment indicator variable that equals one if the district had COVID-19 cases above the national median in 2020 (i.e., 28 cases). Hence, we exploit the district-level variation in exposure to COVID-19 to identify the effects of the pandemic on deforestation.

2.4 Results

Table 2.1 shows our main results. Column 1 presents estimates from our event study in Equation 2.1. Columns 2 and 3 present difference-in-difference (DiD) estimates from Equation 2.2.

Our main results suggest that deforestation in Peru increased significantly due to the pandemic. The average deforestation per district increased by approximately 35% in 2020 compared to pre-pandemic levels. This corresponds to an additional reduction in forest cover of 54 thousand hectares at the national level, which is equivalent to the surface of more than 77 thousand soccer fields. We can identify the combined effect of shocks associated with COVID-19 on deforestation, but we cannot single out a particular policy. Reverse causality is less worrisome in our context given the evidence that forest loss does not affect the incidence of respiratory diseases (Berazneva and Byker, 2017).

Identification concerns Results in Column (1) may be biased under the presence of unobserved time-varying confounders. For example, we cannot disentangle annual trends or other non-COVID-19 related shocks in Equation 2.1. To attenuate this concern, in Column (3), we complement our baseline approach with a difference-in-difference design that controls for district-specific time trends. The stability of our outcome variable before the pandemic also helps mitigate the concern (see Panel (b) in Fig 2.3).

There may be some identification concerns regarding our DiD complementary approach. First, it relies on the assumption that the treatment and control groups have a common trend over time in deforestation. Figure 2.3 presents evidence that both groups may have been experiencing similar trends in the outcome variable prior to treatment. Panel (a) depicts the estimates from an event study that assesses whether there are differences in deforestation between treated and control districts every year. A district is treated if the number of COVID-19 cases (in 2020) in their jurisdiction was above the national median. Circles represent point estimates from regressing

Table 2.1: Main results

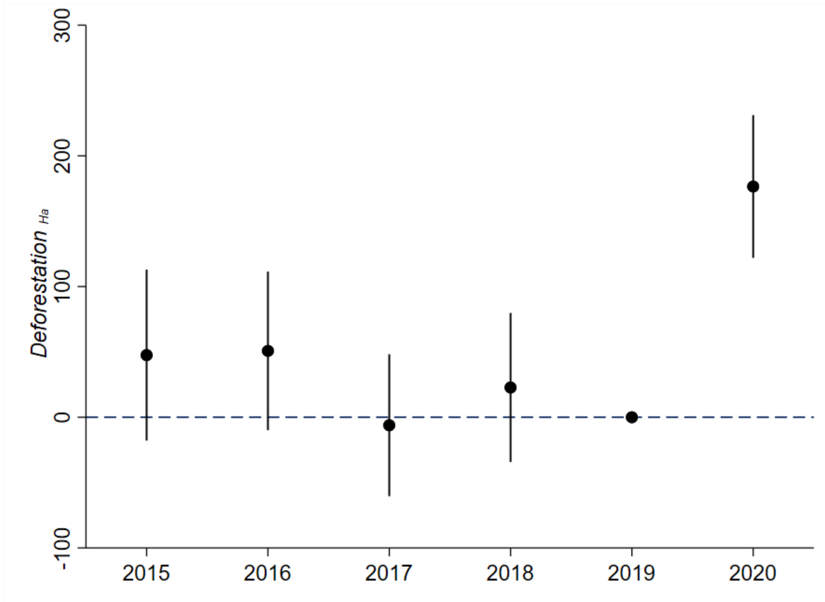
	Dependent variable: Deforestation (ha)		
	(1)	(2)	(3)
Year 2020	137.12***		
	(17.589)		
DiD		153.54***	140.11***
		(35.68)	(31.62)
Design	Event study	DiD	DiD
District-specific time trends	No	No	Yes
Pre-pandemic period	2019	2015-2019	2015-2019
Mean outcome (pre-pandemic)	371.1	390.1	390.1
N	800	2,394	2,394
R-squared	0.966	0.04	0.04

Notes: Estimated standard errors, reported in parentheses, are clustered at the district level. Significance at the one, five, and ten percent levels is indicated by ***, **, and *, respectively. See Table B.3 in the Appendix for similar results using our event study design but expanding the pre-pandemic period to 2015-2019, and using our difference-in-difference approach but restricting the pre-pandemic period to 2019.

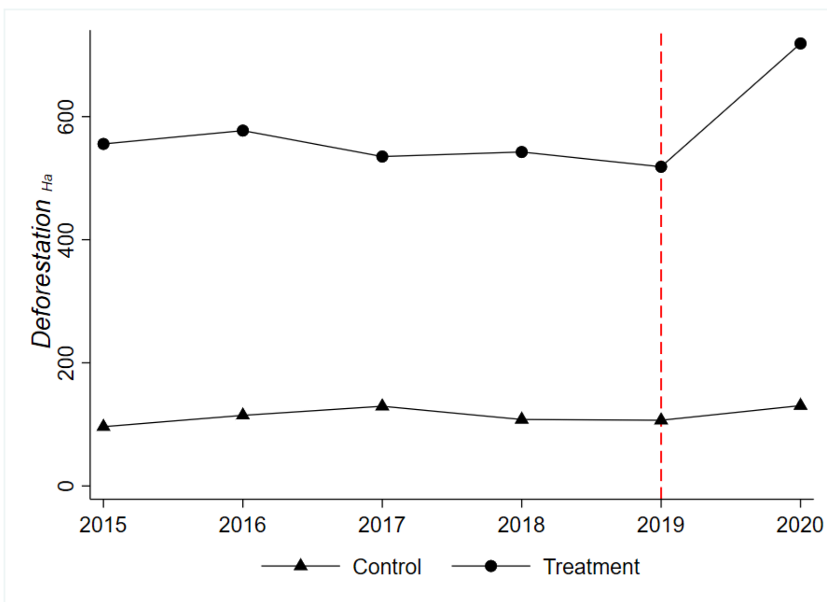
deforestation on dummy variables corresponding to the interaction between the year and treatment dummies, controlling for year and district-fixed effects. The omitted category is the year 2019. Vertical lines show 95 percent confidence intervals, calculated using standard errors clustered at the district level. Except for 2020, the differences in deforestation between the treatment and control groups are not statistically significant at a five percent level of significance. Likewise, Panel (b) provides more graphical evidence of parallel trends over time. We also use recent developments to test violations of parallel trends, following Rambachan and Roth (2023). We find that our result is robust to allowing for violations of post-treatment parallel trends up to 1.5x as big as the maximum violation in the pre-treatment period (see Figure B.3 in the Appendix).

Second, our treatment may be correlated with unobserved events that differently affected the treatment and control groups. To address this concern, we include district-specific time trends to allow the treatment and control districts to follow different trends (see Column (3) in Table 2.1). Finally, the composition of the treatment and control groups may have changed between the pre-treatment and post-treatment periods. Given that our unit of observation is a district, it is unlikely that its composition and characteristics changed dramatically before vs. after the pandemic. However, in order to mitigate this potential concern, we combine our difference-in-difference design with propensity score matching (PSM).

Figure 2.3: Evidence of parallel trends



(a) Event study



(b) Average deforestation 2015-2020

Notes: Panel (a) depicts the estimates from an event study that assesses whether there are differences in deforestation between treated and control districts every year. Panel (b) depicts the average deforestation for treated and control districts during 2015-2020. A district is treated if the number of COVID-19 cases (in 2020) in their jurisdiction was above the median in the country.

Additional robustness checks Table 2.2 shows that our main results are robust across different specifications. Column (1) displays estimated coefficients from Equation 2.2 after restricting the sample to districts with forests that covered 20% or more of their territories in 2015. We use this threshold to avoid potential biases by including districts with substantially different environmental features in the sample. Results suggest that the higher the coverage percentage in the district, the higher the magnitude of the effect. To address concerns regarding measurement error in our treatment variable, in Column (2), we use COVID-19 deaths instead of COVID-19 infection cases as an alternative measurement of our treatment. Here, a district is treated if it registered deaths due to COVID-19. Columns (4) and (5) display the results from applying the difference-in-difference approach to samples matched by their corresponding propensity score. Columns (4) and (5) use COVID-19 infection cases and deaths as the measures of treatment, respectively. We construct control groups based on the conditional probability of districts to be assigned to the treated group, given proxies for biophysical, geographical, and socioeconomic drivers that could exert some influences on deforestation (Busch and Ferretti-Gallon, 2017). We find larger coefficients than those in Table 1 as potential confounders are controlled. We use information from 2019 or earlier on district-level total surface area (IGN, 2022); river area, road number, and distance to the capital city (MTC, 2022); altitude (ECLAC, 2022); slopes (Farr et al., 2007); forest cover (MINAM, 2022); population density (ECLAC, 2022); and human development index scores (ECLAC, 2022). Post-matching covariate balance shows that the procedure achieves important bias reductions in the resulting samples (see Figure B.1 in the appendix).

Table 2.2: Robustness Checks

	Dependent variable: Deforestation (ha)			
	(1)	(2)	(3)	(4)
DiD	193.93***	149.50***		
	(46.64)	(27.77)		
DiD-PSM			171.44*	220.43***
			(91.60)	(44.84)
Measure of COVID-19	Cases	Death	Cases	Death
N	1,728	2,394	1,710	1,705
R-squared	0.05	0.04	0.04	0.06

Notes: All regressions control for district fixed-effects and consider the years 2015-2019 as the pre-pandemic period. Estimated standard errors, reported in parentheses, are clustered at the district level. Significance at the one, five, and ten percent levels is indicated by ***, **, and *, respectively.

Heterogeneous effects We also explore the heterogeneous effects of COVID-19 on deforestation based on the incidence of coca production, illegal or informal mining, and protected areas in the districts. We focus our analysis on coca production and mining due to their notorious potential to trigger short-term land use changes across the Peruvian Amazon (Swenson et al., 2011, Young, 1996). We also analyze the role of protected areas because this is the most frequent policy applied to deter deforestation processes in the country. Instead of being a single command-and-control instrument, it encompasses a wide range of governance regimes with different levels of national agencies participation. One-quarter of the Peruvian Amazon region is under some protected area regime (SERNANP, 2022). Aside from these two drivers, commercial agriculture and cattle ranching have also been identified as relevant deforestation drivers across the Peruvian Amazon. However, both these activities follow complex dynamics that delay by several years the transition from forests to temporary land uses (Armas et al., 2009). Therefore, they are less likely to play a role in the impact of the pandemic on deforestation.

Table 2.3: Heterogeneous effects

	Dependent variable: Annual rate of forest change, 2019-2020	
	(1)	(2)
Year 2020	2.19***	2.17***
	(0.21)	(0.21)
Mining	0.17*	0.18*
	(0.09)	(0.09)
Coca	0.21**	0.19*
	(0.10)	(0.10)
Protected areas	-0.34***	-0.44***
	(0.08)	(0.08)
N	394	394
R-squared	0.53	0.54

Notes: Estimated standard errors, reported in parentheses, are clustered at the district level. Significance at the one, five, and ten percent levels is indicated by ***, **, and *, respectively. Regressions include other district characteristics such as population density, river area, altitude, number of roads, distance to the capital city, and slope (see Table B.4 in the Appendix).

Table 2.3 displays our results. We find that deforestation is exacerbated in districts with coca production, or with informal or illegal mining. On the other hand, the presence of protected areas mitigated the effects of deforestation in the district, with increasing efficacy as the areas under protection get larger. This result is in line with previous findings that highlight the role of decentralized management models of protected areas in successfully mitigating deforestation (Schleicher et al., 2017). Our outcome variable is the annual rate of forest change between 2020 and 2019. Given that deforestation was higher in 2020 compared to 2019, the annual rate of forest change is a negative number. To ease the interpretation of results, we multiply the estimated rate by -1. We then regress our modified outcome variable on dummy variables that capture whether the year is 2020, whether coca production was recorded in the district in 2017 (using data from UNODC (2017)), and on whether there was illegal or informal mining activity in the district (using data provided by MINAM (2016)). All the regressions control for the average district slope, the total district area, and the extension of rivers and national roads. Columns (1) and (2) present the estimates from the regression. The only difference between the columns is the treatment of the variable protected areas. It equals one if there are protected areas in the district in Column (1), while in Column (2) it equals one if protected areas represent 10% or more of the territory in the district. We made this differentiation to assess the extensive margin and the intensity of the protected areas policy in mitigating the COVID-19 deforestation effects.

In summary, we identify the weakened institutional capacity of the country to conduct monitoring and enforcement activities as a mechanism through which COVID-19 emergence enabled an increase in deforestation. This institutional weakening led to an exacerbation of illegal and informal activities driving forest loss.

2.5 Discussion

2.5.1 CO2 emissions and social cost

We estimate the impact of the deforestation caused by the pandemic in terms of carbon (CO) emissions and the corresponding social costs. We use the following equation to calculate the released tonnes of equivalent CO:

$$tCO_2\text{-}eq = (Def_{2020} \times n \times E) * 3.67 \tag{2.3}$$

Where tonnes of $tCO_2\text{-}eq$ are estimated using the increase in deforestation observed in 2020 (Def_{2020}) and captured by our estimates from Equation 2.1 (Column 1 of Table 2.1). We multiplied this by the number of districts in our data ($n = 400$), and by a parameter representing a fixed amount of tonnes of CO released per deforested hectare ($E = 84.54\ tCO/ha$) that was previously estimated by the Ministry of the Environment in Peru, considering the different types of forest in the country (Malaga et al., 2014). This is then transformed to $tCO_2\text{-}eq$ by multiplying by 3.67 (i.e., the factor to transform carbon to carbon dioxide).

Our result shows that COVID-19 could have contributed to the release of 12.7 to 21.3 million tCO_2-eq in 2020. Using the social cost of carbon of 7.72 USD/tCO_2-eq estimated for Peru (MEF, 2021), we calculate the associated economic losses to be around USD 98.2 million and 164.5 million (see Table B.5 in the Appendix). This cost represents almost three times the national annual budget allocated to forest protection in the country.

2.5.2 Mechanisms

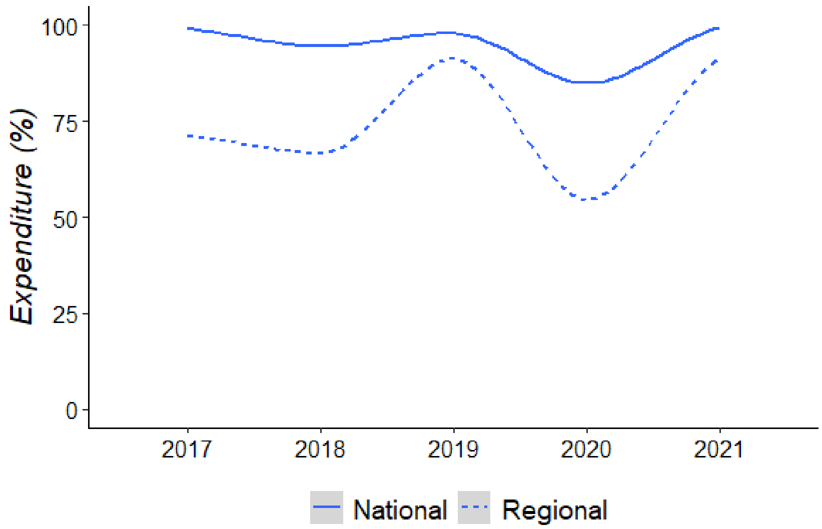
A potential mechanism explaining the impact of the pandemic on deforestation is a decrease in forest monitoring efforts, paired with an increase in illegal deforestation activities. Figure 2.4 shows that the investment in activities related to forest monitoring between 2019 and 2020, at the national and regional levels, experienced a reduction of 13 and 37 percentage points, respectively. Likewise, we have anecdotal evidence from journalistic reports that monetary resources allocated to enforcement were redirected to provide economic support to the population. In addition, personnel in charge of enforcement were reassigned to enforce social restrictions and lockdowns. This reallocation of resources and personnel also varied with the intensity of the pandemic since during the second half of 2020, three regions (Huanuco, San Martin, and Madre de Dios) out of the five that explain 80% of the forest loss in the country (Potapov et al., 2014) remain under quarantine while others had the restrictions lifted. In addition, data from OSINFOR (Monitoring Agency of Forest Resources and Wildlife), reports that the number of supervision conducted by them decreased by 57% in 2020 compared to pre-pandemic levels. We interpret all this information as evidence that monitoring activities may have decreased, making it more difficult to detect illegal activities in forested areas in the Amazon.

Illegal coca production and illegal mining are also related to higher deforestation during the pandemic. Results from the heterogeneity section show that deforestation was exacerbated in districts with coca production or with informal or illegal mining. In addition, Panel (a) in Figure 2.5 shows that illegal coca production reached a peak in 2020, possibly caused by the reduction of eradication efforts during the pandemic, as shown in Panel (b). Likewise, journalistic reports (e.g., Vera (2020)) attest to the intensification of artisanal mining in highly forested areas in 2020.

The variation in our deforestation outcome seems to be driven by illegal deforestation. Our outcome accounts for both legal (concession logging, agriculture, etc.) and illegal deforestation (mining, coca). However, legal logging activities decreased in 2020, as shown in Figure 2.6, which displays information regarding round wood production in the last decade. This suggests that our findings are mostly linked to deforestation related to the illegal activities described earlier.

The economic crisis generated by the pandemic offers two additional mechanisms: (i) the trade-off between livelihoods and the forests, and (ii) migration. The economic crisis and the lack of employment could have led to people clearing more forests. Individuals participating in forest-

Figure 2.4: Budget expenditures (%) in forest monitoring, 2017-2021

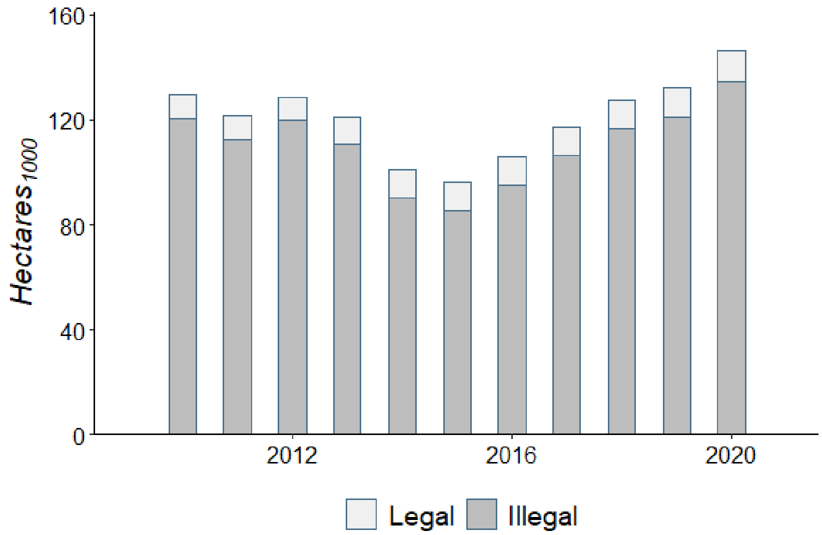


Notes: Figure depicts the budget expenditures (%) at the national and regional level for activities related to forest monitoring by the end of each year, from 2017 to 2021.
Source: MEF (2022)

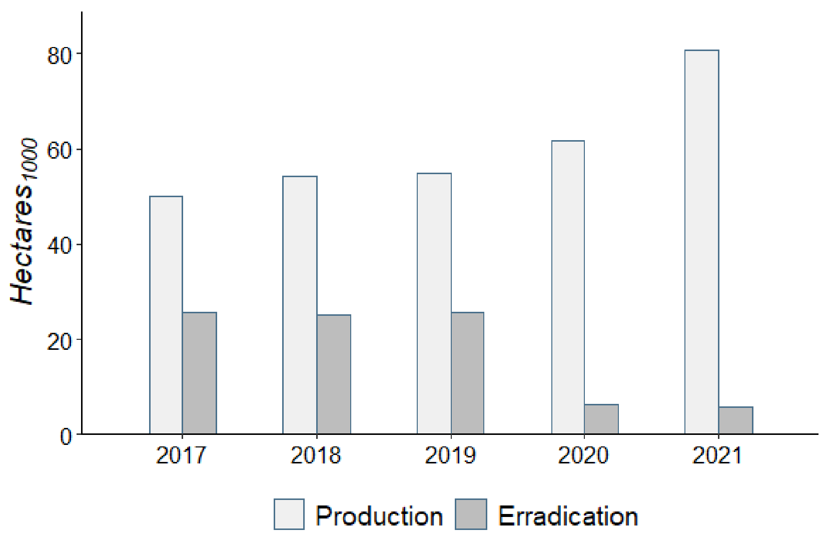
clearing activities could have intensified their efforts (intensive margin) or individuals could have switched from other activities to clearing forests (extensive margin). However, there is evidence that tropical forests have a key role in rural contexts in providing wild food (fish, bushmeat, fruits, etc.) to the community (Van Vliet et al., 2017). Moreover, we observe that deforestation and a food vulnerability index are negatively correlated (see Figure B.2 in the Appendix). Overall, although this could have explained some of the deforestation, it would be unlikely to explain the huge increase we observe. In terms of migration, the economic downturn pushed migrants located in cities to move back to their home rural areas. This increase in population could have put pressure on forests through residents' participation in economic activities driving deforestation. Fort et al. (2021) explore this hypothesis and find a weak correlation (0.086) between deforestation in 2019-2020 and the number of people returning to their cities of origin due to the pandemic. While this figure considers only returning migrants and not all migrants, due to a lack of data we cannot explore further this channel.

Other mechanisms that may explain the increase in deforestation are: (i) the possibility that individuals may have been incentivized to do selective logging due to increasing prices of roundwood. Second, the economic crisis generated by the pandemic which was more pronounced in the service sector (where in-person interactions are more intrinsic), might have generated a reallocation of the labor force from the service sector towards agriculture pressing the demand for land, and therefore inducing deforestation.

Figure 2.5: Coca production and eradication



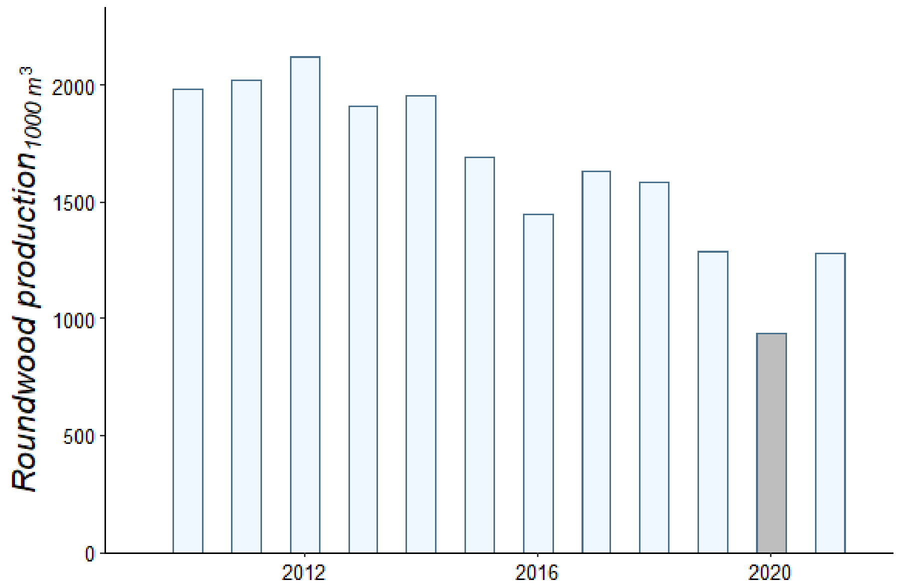
(a) Legal and illegal coca production, 2010-2020



(b) Hectares of Coca produced and eradicated, 2017-2021

Notes: Panel (a) depicts the production of coca leaves in metric tonnes for legal and illegal markets. Panel (b) depicts the hectares of coca leaves produced and eradicated. Data is from DEVIDA (2022).

Figure 2.6: Roundwood production, 2010-2020



Notes: This figure depicts the roundwood production for 2010-2021 recorded by the national government.
Source: SINIA, 2022

2.6 Concluding remarks

This paper provides evidence that COVID-19 increased deforestation in the Amazon. This surge in deforestation might have had a considerable negative net impact on Peru's climate commitments. Our findings also unveil the role that illegal and informal activities, such as coca leaf production and mining, have had on tropical forest loss during the COVID-19 pandemic in the country, and highlight the importance of protected areas in significantly mitigating the deforestation triggered by the pandemic across the Peruvian Amazon.

Further research should be done to better understand the role of governance regimes in preventing forest loss during the COVID-19 pandemic, and to evaluate the consequences of the pandemic on biodiversity loss due to the increase in deforestation.

Chapter 3

Is it Too Hot to Work? Evidence from Peru

3.1 Introduction

Time is central to humans' well-being as well as the consequences of climate change. Evidence from the US suggests that rising temperatures may reduce the time allocated to work in industries highly exposed to weather. However, the damages at a global scale might be significant. For instance, the International Labour Organization (ILO) projects that due to global warming, the world will lose 2.2% of total working hours (equivalent to 80 million full-time jobs) by 2030, leading to a reduction of the world's GDP in US$2,400 billion.

However, the evidence of the effects of climate change on time use is still sparse, particularly in developing countries. Understanding this labor-temperature relationship is particularly relevant for developing countries since they concentrate 80% of the world's labor force (Behrman, 1999), they are located in tropical areas where changes in climate will occur faster and with more intensity (Aragon et al., 2020), they are expected to face higher associated costs of climate change (Dell et al., 2014, Jessoe et al., 2016), and they have a high incidence of asset-poor households with lack of access to adaptation strategies or avoidance behavior (Jessoe et al., 2016).

This paper seeks to answer the question: what is the effect of temperature on working hours? The main contribution of the paper is to highlight the relevance of segmented labor markets when assessing the impacts of climate change on labor market outcomes. To answer this question, I combine worker longitudinal microdata from household surveys with meteorological reanalysis data for Peru covering the period 2007-2015. I exploit presumably random year-to-year variation in temperature within residential localities to estimate whether weekly working hours for a particular individual are higher or lower in years that are warmer. Although this setting minimizes the omitted variable bias problem (Deschenes and Greenstone, 2007), I control for rainfall, humidity, and daylight hours which are correlated with temperature and could potentially affect the number of working hours.

This paper documents three main findings. First, I find a negative effect of high temperatures on overall working hours. In general, individuals reduce weekly work time when experiencing an additional day with a temperature above 27°C compared to a day with 'comfort zone' temperatures for human beings (i.e. between 18-21°C). This finding is consistent with ILO (2019) that states that temperatures above 26°C are associated with negative impacts on an individual's work capacity.

I also find that the negative effect of high temperatures on work time is driven by informal jobs instead of jobs in industries highly exposed to weather. Previous studies for developed countries have found that high temperatures may reduce hours worked in industries where jobs are primarily performed outdoors. However, the type of industry loses relevance once we consider a labor market that is highly segmented between formal and informal jobs as observed in developing countries. In a highly segmented labor market, informal workers are negatively affected by high temperatures regardless of whether they have outdoor or indoor jobs.

The last main result is that labor market segmentation can lead to misleading conclusions regarding intertemporal labor supply. Workers seem to substitute work time across weeks due to high temperatures in the aggregated labor market. However, the aggregate intertemporal labor substitution is masking two opposite effects from the segmented labor market. It combines the positive effect of high temperatures on formal workers with the negative effect on informal workers.

This paper contributes to a growing literature on the impact of temperature on labor market outcomes such as work time (Connolly, 2008, Garg et al., 2020, Kruger and Neugart, 2018, Schwarz, 2018, Zivin and Neidell, 2014); wages (Schwarz, 2018); productivity (Dell et al., 2014, LoPalo, 2020, **?**); and labor reallocation (Colmer, 2020, Jessoe et al., 2016). However, none of them have explored the role of highly segmented labor and the implications for understanding the future impact of climate change on the labor market. This study also contributes to the literature on weather and intertemporal labor supply, most of which is focused on developed countries such as the US and Germany (Connolly, 2008, Kruger and Neugart, 2018, Zivin and Neidell, 2014) with one exception for China (Garg et al., 2020) that did not exploit the high labor market segmentation feature observed in developing countries.

The remainder of the paper is organized as follows. Section 3.2 describes the data used to measure weather and working hours. Section 3.3 discusses the empirical approach. Section 3.4 presents the results. Section 3.5 presents the main robustness checks. Finally, Section 3.6 presents the limitations and conclusions.

3.2 Data

To study the temperature-labor supply relationship, I require variables capturing information at the individual-, household-, and location-level. I combine worker-level data from household surveys with meteorological data for Peru.[1] The Peruvian case is relevant given that the country satisfies seven out of the nine requirements defined by the United Nations to consider a nation particularly vulnerable to climate change.[2] In addition, Peru also has a large variation in weather since it has more than 70% of all the types of climates in the world.[3]

3.2.1 Labor Data

I use two sets of panel data from the Peruvian Living Standard Survey (ENAHO, Spanish acronym) covering the period 2007-2015. This survey is nationally representative and collects information along the year at the household- and individual levels. It also keeps a record of the date of the interview and provides the residential geographic coordinates (i.e. longitude and latitude) for all participating households.[4]

In particular, I use the module about employment to calculate the outcome of interest: working hours. In this module, working-age individuals report their time allocated to work for each of the days during the *reference week* (i.e. the last week previous to the interview date), their economic sector (e.g. agricultural or non-agricultural), their occupations, and whether they are wage workers or self-employees. Importantly, every worker on the panel is interviewed during the same month across different years. The sample consists of 117,430 person-year-observations.

3.2.2 Weather Data

Temperature and Humidity. I use data from ERA5 which is the latest reanalysis data produced by the European Center for Medium-Range Weather Forecasting (Munoz Sabater, 2019). It has a much higher spatial and temporal resolution and supersedes the ERA-Interim archive which is one of the most commonly used reanalysis products (Auffhammer et al., 2013).[5] ERA5 provides hourly

[1]Table C.1 in the appendix shows summary statistics for the main variables. Temperature is measured in degree Celsius and precipitation in mm.

[2]According to MINAM (2015), the seven requirements are: low-lying coastal area; arid and semi-arid lands; areas liable to flood, drought, and desertification; fragile mountain ecosystems; disaster-prone areas; areas with high urban atmospheric pollution; and economies highly dependent on income generated from the production and use of fossil fuels.

[3]See Figure C.1 in the appendix for the temperature distribution in Peru.

[4]We exclude from our sample a few households because they did not have information about geographic coordinates. The number of households is 6, 5, 3, and 4 households for 2007, 2008, 2011, and 2013, respectively.

[5]More strengths of ERA5 compared to Era-Interim can be found here: https://confluence.ecmwf.int//pages/viewpage.action?pageId=74764925

data for surface air temperature, humidity,[6] and precipitation on a 0.25×0.25 degree latitude-longitude grid.

Precipitation. I use data from the Peruvian Interpolated data of SENAMHI's Climatological and Hydrological Observations (SENAMHI PISCOp) and from Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS). SENAMHI PISCOp developed by Aybar et al. (2019) provides monthly precipitation on a 0.1×0.1 degree latitude-longitude grid, but their estimation combines information from monitoring stations and CHIRPS. Note that these precipitation estimates achieve their highest performance for the Pacific coast and the western flank of the Andes in Peru. On the other hand, I use monthly precipitation on a 0.05×0.05 degree latitude-longitude grid from CHIRPS for robustness check.

Daylight. I calculate daylight hours for each day of the working week by taking the difference between the sunset and sunshine time at each household location. Sunset and sunshine times were calculated using astronomical algorithms taking as input the date each individual worked and the residential geographic coordinates reported in the ENAHO.[7]

For each household location, I calculate the daily average, maximum, and minimum temperature, daily humidity, and daily precipitation.[8] Note that temperature is available either at the hourly or daily level. Hence, I aggregate population-weighted temperatures at the weekly level to match the periodicity of our labor outcome as in Carleton et al. (2020). Likewise, monthly precipitation is divided by the number of days in the corresponding month to calculate daily rainfall.[9] Importantly, as reported by Aragon et al. (2020), satellite data do not perform very well for the jungle in Peru. Therefore, I will exclude this region from all the analysis.[10]

3.3 Empirical Approach

3.3.1 Residual Variation

To identify the effect of climate change on working hours I rely on weather variation. However, my empirical approach considers some form of fixed effects that may absorb substantial variation

[6] We use the dewpoint temperature to proxy relative humidity. Note that the formula to calculate relative humidity take only as inputs temperature and the dewpoint temperature.

[7] I use the package suncalc in R to calculate daylight hours.

[8] In case temperature is missing for a particular day and residential location, I use the temperature data for the same day from other households located in the same district to impute the average temperature across all these households to the residential location with missing temperature.

[9] There are very few cases where the reference week does not correspond to the year of analysis. For instance, if we are analyzing 2007, the reference week in January 2007 may correspond to December 2006. In those cases, we assign the rain information from January 2007.

[10] Note that the jungle represents approximately only 13% of the labor market in Peru.

in weather. How much variation in weather is left after controlling for fixed effects? This section provides the answer to this question.

How can I measure the remaining variation in weather? I follow the literature on this (Fisher et al., 2012, Guiteras, 2009, Jessoe et al., 2016, Schwarz, 2018) and regress temperature on different fixed effects specifications and time trends. The residual from these regressions measures the remaining variation in weather. Ideally, to identify the effect of climate change on working hours, I would like the remaining variation in weather to be as large as the predicted changes in weather by climate change models. Table C.2 reports the R^2 of the regression of average and maximum temperature on fixed effects, the standard deviation of the residuals, and the fraction of the observations that have a residual with an absolute value larger than the predicted change in weather (in our case 1°C).

The remaining variation in weather is larger when using maximum temperature and models without individual fixed effects. For instance, I can observe rows 7, 21, 23, and 24 that use fixed effects as in Zivin and Neidell (2014), Schwarz (2018), and Garg et al. (2020). There, the fraction of residuals larger than 1°C is between 5 to 9 percentage points higher when using maximum temperature instead of average temperature. On the other hand, this fraction is around 40% for models in rows 7 and 21, while it is around 25% for models in rows 23 and 24. As mentioned above, the larger this fraction the better for identification. Hence, the data for Peru seems more suitable for models using fixed effects as rows 7 and 21.[11] In fact, our preferred specification described in the next section will use a model as in row 7 similar to Zivin and Neidell (2014). Nonetheless, I will use a model with individual fixed effects as a robustness check.

Finally, our empirical strategy relies on variation in working hours within individuals. To show how much variation there is in our main outcome variable, I follow Kruger and Neugart (2018) and report in Table C.4 a between and within decomposition of the standard deviation of working hours. I can observe that between variation in working hours is higher than the within variation. Thus, Table C.4 shows that there is variation in the observed labor data.

3.3.2 Empirical Model

My main specification is described in Equation 3.1. It uses a panel of individuals with district and time fixed effects for reasons described in the previous section. I exploit presumably random year-to-year variation in temperature within localities to estimate whether working hours for a particular individual living in a given locality are higher or lower in years that are warmer.

[11] Table C.3 in the appendix shows that I obtain similar conclusions if I regress each temperature bin on the different fixed effects models and time trends.

$$\text{labor}_{idt} = f(\beta, w_{dt}) + \alpha_d + \lambda_t + \delta Z_{dt} + \theta X_{idt} + \varepsilon_{idt} \tag{3.1}$$

In Equation 3.1 the outcome variable is total working time. It is constructed as the summation of all hours worked during all days of the *reference week* in all jobs for a given year.[12] Note that individuals, localities, and years are indexed by i, d, and t, respectively. Location and time fixed effects are represented by α_d and λ_t, respectively. Time fixed effects include year-month and weekly dummies to account for seasonality. The error term ε_{idt} is clustered at the region-month level to address temporal and spatial correlation in temperature as in Zivin and Neidell (2014).[13]

Our parameters of interest are the β in the nonlinear function of daily maximum temperature $f(\beta, w_{dt})$. Temperature is divided into 7 bins of 3°C increments.[14] The bin 18-21°C is taken as the reference group for interpretation because it overlaps with the "comfort zone" temperature band for human beings (i.e. between 18-22°C) according to Heal and Park (2016).[15] Hence, β can be interpreted as the effect on the hour worked of shifting a day from the reference bin to bin j during the working week.

Although the setting in equation 3.1 minimizes the omitted variable bias problem (Deschenes and Greenstone, 2007), I include rainfall, humidity, and daylight hours in Z_{dt} to control for other locality-level weather variables that are correlated with temperature and could potentially affect the number of working hours. Finally, I include individual-level demographics in X_{idt} such as age, gender, education (high school dropout, high school graduate, some college), employment status (on vacation, full-time employee), and other labor information (whether the job is informal, whether the contract is permanent, size of the company). I also include dummies for economic activity, household income, whether they live with a partner, information about dependents (number of children and percentage of family members over 65), and whether they live in rural or urban areas.

3.4 Results

I find evidence of a negative effect of high temperatures on hours worked for the overall labor market. Figure 3.1 shows the estimates for β in equation 3.1 and their confidence interval at 95%

[12] In our sample 26% have a secondary job and 1% reported zero hours of work.

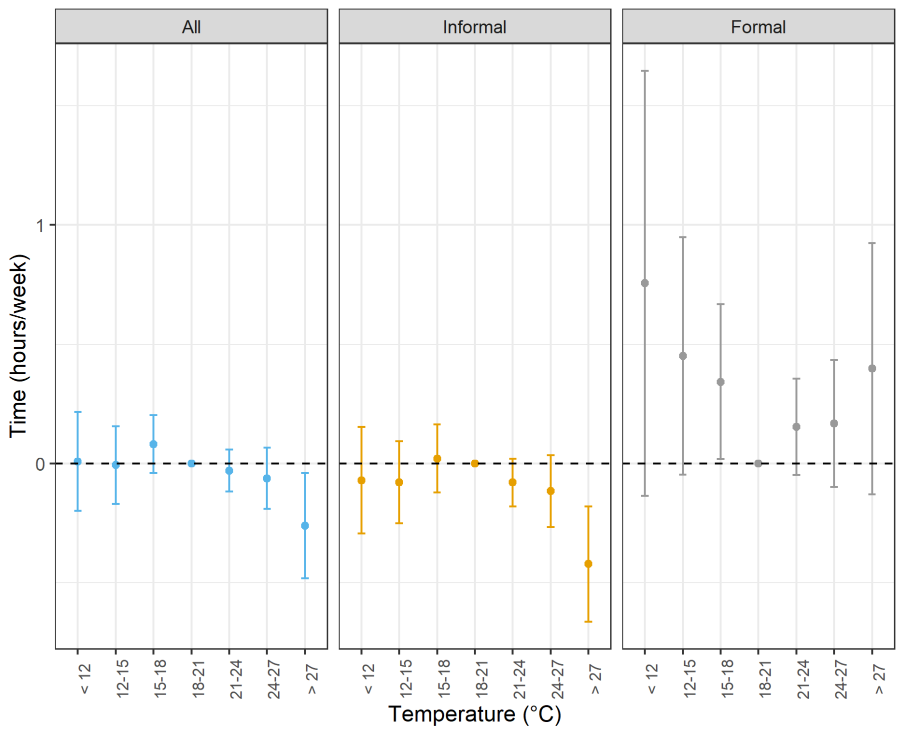
[13] In section 5, I implement Conley (1999, 2010) standard errors using the algorithm developed by Colella et al. (2019).

[14] The number of bins was selected so that each bin contains information for at least 15% of the sample. As a result, on average, each bin contains at least 3/4 of a day out of the seven days of the reference week. Note that in section 5 we present results with 9 bins of 3°C increments and also with bins of 2°C increments, and results remain.

[15] Note that the average maximum temperature in our sample lies also in the 18-21°C bin. In addition, ILO (2019) states that temperatures above 24°C are not comfortable for workers.

significant level.[16] These results are in line with findings from other developing countries such as China. For instance, Garg et al. (2020) find that extreme low and high temperatures reduce working hours. However, this is not compatible with findings in the US (Zivin and Neidell, 2014) and in Mexico (Schwarz, 2018). For instance, in the US there seems to be no effect of temperatures on working hours in the entire labor market while in Mexico only low temperatures reduce hours worked.

Figure 3.1: Effect of Temperature on Work Time.



Notes: Figure depicts the estimates of the effect on hours worked of shifting a day from the reference bin (18-21°C) to a given bin during the working week for all workers. Circles represent point estimates from regressing total working hours on temperature bins, controlling for precipitation, humidity, daylight hours, sociodemographics, and location and time fixed-effects (see Table C.5 for the regression output). Vertical lines show 95 percent confidence intervals calculated using standard errors clustered at the region-month level. The figure uses maximum temperature, ERA5 data, and excludes the jungle.

The magnitude of the effects of temperature on hours worked is not large in general. Figure 3.1 shows that shifting a day with 'comfort zone' temperatures (i.e. 18-21°C) to a day above 27°C reduces 18 minutes of work time during a week. This is below the weekly reduction of 1.2 hours reported for China for the same temperature range in Garg et al. (2020). This may be explained

[16] See Table C.5 in the appendix for the regression output.

by the differences in the rate of self-employment which is higher in Peru, and therefore the labor market has more flexibility to adapt to changes in temperature, leading to lower estimates.

The negative effect of temperature on hours worked seems to be driven by workers with informal jobs.[17] Figure 3.1 shows separate regressions for each group of workers (i.e. informal vs. formal) instead of adding interaction terms between each temperature bin and an indicator variable for labor informality.[18] There we can see that all the negative effect of temperature is mainly driven by workers with informal jobs. However, 67.5% of informal workers in the sample have outdoor jobs.[19] To disentangle whether the effect is driven by outdoor or informal jobs, I explore within outdoor and indoor jobs whether the effect is driven by informal workers. I run separate regressions for each group of workers (e.g. outdoor-informal or outdoor-formal workers), and I find that all the effect is driven by informal jobs.[20] Hence, informal workers are negatively affected by high temperatures regardless of whether they have outdoor or indoor jobs. This might not be surprising since more than 75% of outdoor and indoor jobs are informal. This reflects the high labor market segmentation between informal and formal jobs observed in developing countries that are not present in the US where Zivin and Neidell (2014) found that only outdoor workers are responsive to high temperatures. Note that in this country setting, 87% of government agencies do not use AC, and most firms in the private sector do not use AC except firms in the mining sector where only 44% of them use AC (MINEM, 2013). Therefore, this information suggests that both informal and formal jobs lack access to AC. On the other hand, a large share of informal jobs are associated with self-employees who may have more flexibility to allocate hours to work. Thus, informality may be confounded with the degree of flexibility in the job to change hour worked. Likewise, informal jobs may be more exposed to demand shocks (e.g., on hot days people may spend less time outdoors, and therefore informal workers do not have customers, and in response to this they change their allocation of hours to work) that affect hours worked in combination with the supply shock associated to high temperatures.

[17]The ENAHO survey provides a variable to identify workers with formal or informal jobs. A job is informal if the production unit is not registered for tax purposes or the worker is not covered by social security.

[18]I performed a test to evaluate whether the effects of the other control variables are the same across groups. I found that those effects are statistically different between subgroups. Hence assuming that those effects are equal between subgroups, as we do when we use interactions, does not seem reasonable in this context.

[19]Individuals with outdoor jobs work in high-exposure industries to changes in weather such as agriculture, fishing, mining, manufacturing, transportation, and utilities. This corresponds to the “high-risk industries" categorization in Zivin and Neidell (2014) that is based on definitions from the National Institute for Occupational Safety and Health (NIOSH). Note that high-risk industries are those where the work is primarily performed outdoors.

[20]See results reported in Figure C.1. Table C.6 reports results from a model with interactions and I reach the same conclusion.

3.4.1 Intertemporal Substitution

In this section, I explore the possibility that workers may substitute working hours across adjacent weeks due to unpleasant temperatures. Zivin and Neidell (2014) and Kruger and Neugart (2018) suggest that time allocation may depend more on the weekly weather instead of the daily weather, making the analysis across weeks more relevant in this study.

I find evidence of intertemporal substitution of work time due to changes in temperature for the overall labor market. Figure 3.2, panel (a), shows that temperatures above 27°C decrease working hours for the same week of exposure (solid line). In contrast, hot temperatures from the previous week increase working hours in the current week (dashed line), albeit the coefficient is not statistically significant.[21] In addition, the effect of hot temperatures on work time across the two adjacent weeks is not statistically significant as shown in Panel (b) at the 95% level.[22] This may suggest that workers are substituting work hours across weeks so that their work time is not affected by high temperatures. Moreover, this also may suggest that intertemporal labor substitution mutes the effects reported in figure 3.1 which then might be only capturing temporary or non-persistent effects.

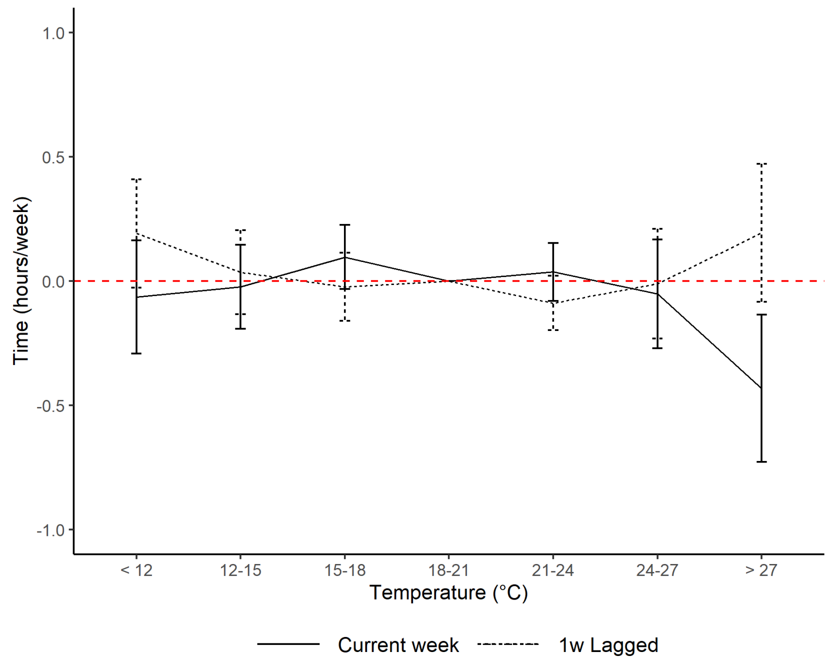
However, the intertemporal labor substitution for the overall labor market is masking two opposite patterns from the high segmentation between informal and formal jobs. Figure 3.3, panel (a), shows point estimates of the effect across the two adjacent weeks for formal workers. The estimates for temperatures below 12°C and above 27°C are both positive and statistically significant at the 95% and 90% levels, respectively. This results from having the current week's temperature not affecting the current week's work time, and from having the previous week's low and high temperatures increasing the current week's work time. For instance, the previous week's temperature below 12°C makes them work one extra hour during the current week. Likewise, high temperatures above 27°C make them work approximately 40 minutes more during the current week, albeit the estimate is statistically significant at the 90% level.[23] Thus, there is no evidence of intertemporal labor substitution for formal workers. Instead, there is evidence of a cumulative positive effect of extreme temperatures on hours worked. Regarding informal workers, the pattern described above is almost reversed. Figure 3.3, panel (b), shows that the cumulative effect across adjacent weeks is negative for high temperatures suggesting as well the absence of intertemporal labor substitution. In fact, there is no statistically significant effect of the previous week's extreme temperature on hours worked during the current week. However, the current week's temperatures above 27°C make individuals work 30 minutes less during the same week.

[21]Note that the point estimate for the bin <12°C is positive as well and significant at the 90% level.

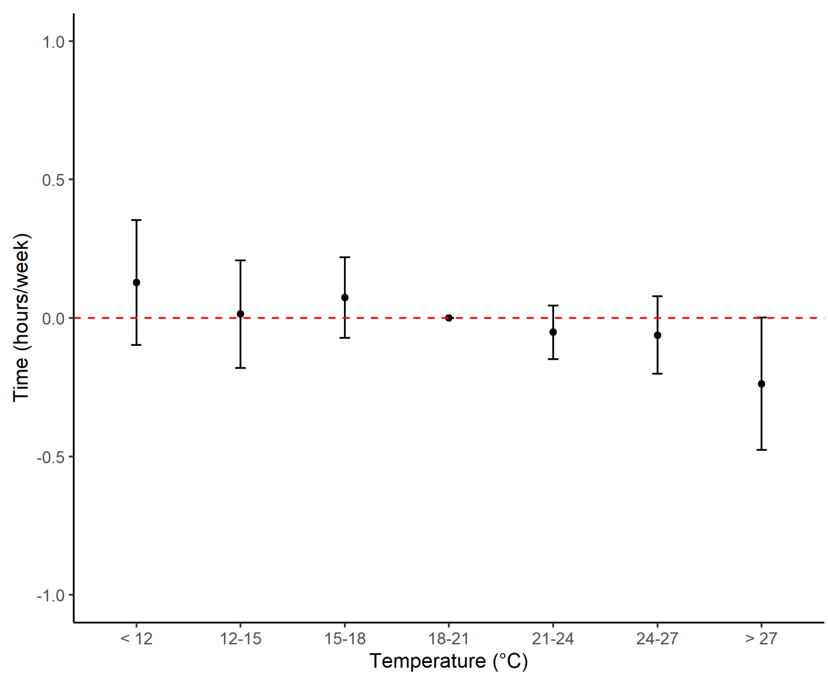
[22]Table C.7 in the appendix reports the regression output for figure 3.2. The effect for bin >27°C is negative and statistically significant at the 90% level.

[23]See the estimates in Figure C.4 in the appendix.

Figure 3.2: Intertemporal Labor Substitution.



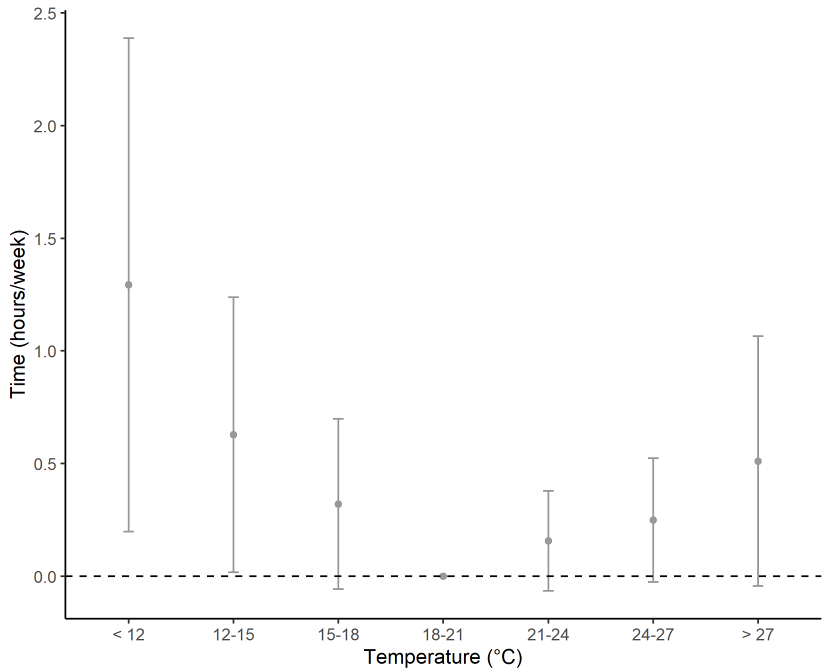
(a) Current and lagged effects of temperature



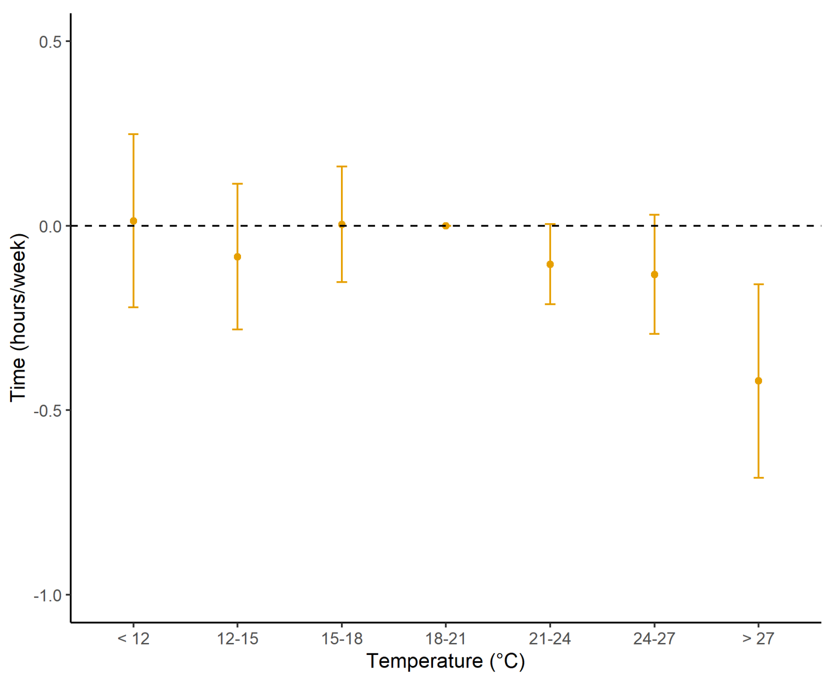
(b) Summation of current and lagged effects of temperature

Notes: Figure depicts the estimated coefficients from regressing working hours on temperature bins for the current and previous week for all workers. Panel (a) plots the effect of contemporaneous and one-week-lagged temperature bins on working hours. Panel (b) plots the summation of the effect across the contemporaneous and lagged weeks. All regressions control for precipitation, humidity, daylight hours, sociodemographics, and location and time fixed-effects. Vertical lines show 95 percent confidence intervals calculated using standard errors clustered at the region-month level. The figure uses daily maximum temperature, ERA5 data, and excludes the jungle.

Figure 3.3: Intertemporal Labor Substitution: Informal vs. Formal



(a) Formal workers



(b) Informal workers

Notes: Figure depicts the summation of the effect of temperature on working hours across the contemporaneous and (one-week) lagged weeks for formal and informal workers. All regressions control for precipitation, humidity, daylight hours, sociodemographics, and location and time fixed effects. Vertical lines show 95 percent confidence intervals calculated using standard errors clustered at the region-month level (see Figure C.4 in the appendix for a similar figure showing 90 percent confidence intervals). The figure uses daily maximum temperature, ERA5 data, and excludes the jungle.

In summary, the aggregated intertemporal labor substitution found in Figure 3.2 combines the positive effect of high temperatures on formal workers and the negative effect on informal workers. Hence, the aggregate intertemporal labor substitution is masking two opposite effects from a segmented labor market.

3.5 Robustness

Table 3.1 presents the robustness checks to our baseline specification. Column (1) shows our baseline model from equation 3.1. In column (2), I try to address the concern that our results may be driven by changes in labor demand instead of labor supply given that weather potentially affects both labor demand and supply. Following Kruger and Neugart (2018), I exclude from our sample workers in areas where the demand for labor is more affected by changes in temperature such as agriculture, forestry, fishing, and construction. Our main results are almost identical in magnitude and sign; however, they are no longer statistically significant. More conceptually, I could claim that I am focusing on short-run year-to-year variations, where it is less likely to see changes in wages, changes in employer, or changes in the contract as in Connolly (2008).

In column (3), I implement Conley (1999, 2010) standard errors to correct for spatial and temporal correlation using the algorithm developed by Colella et al. (2019). The results are identical to the ones in column (1). Alternatively, in column (4), I clustered the standard error at the district level since this is the cross-sectional level of exogenous variation in temperature as in Schwarz (2018) and Garg et al. (2020). Results are also very similar to the baseline model.

In column (5), I deal with residential sorting. I exclude individuals who are not currently living in the same district they were born. This assumes that individuals who were born and currently live in the same district never moved out during our period of analysis. Results from the main model still hold.

In columns (6), I use alternative sources for precipitation. The main findings are invariant to instead using rainfall data from CHIRPS. I also replicate as close as possible other specifications used in the literature of temperature and working hours. In column (7), I estimate a model with individual fixed effects instead of district fixed effects as in Garg et al. (2020) to minimize potential biases due to omitted variables. Even though we lose statistical significance, the coefficient for temperatures above 27°C has the same sign and it is similar in magnitude to the baseline model in column (1). I reach a similar conclusion in column (8) where I estimate a model with individual effects but without controlling for key potential omitted variables such as precipitation, humidity, and daylight hours following Schwarz (2018). In column (9), I also follow Schwarz (2018) and estimate a model that only controls for time fixed effects and demographics. Results are still similar to the baseline model.

Finally, I explore alternative measures of temperature. In Table C.8, I present results using 9 bins of 3°C increments and I get comparable results. In Table C.9, I present results using narrow bins of temperature that increase at 2°C. Even though we lose statistical significance, results are still comparable to those from the baseline model in sign and magnitude. In Table C.10, only one parametric measure of temperature seems to capture the non-parametric relationship estimated in figures 3.1.

Table 3.1: Robustness Checks.

	Baseline	**Demand**	**Conley S.E.**	**District S.E.**	**Sorting**	**Chirps**	**Model I**	**Model II**	**Model III**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
Temperature (Â°C)									
< 12	0.02	-0.20	0.02	0.02	0.13	0.03	-0.11	-0.08	0.01
	(0.10)	(0.18)	(0.12)	(0.13)	(0.13)	(0.10)	(0.27)	(0.27)	(0.20)
12-15	0.00	0.19	0.00	0.00	0.07	0.01	-0.05	-0.01	0.02
	(0.08)	(0.15)	(0.10)	(0.10)	(0.11)	(0.08)	(0.25)	(0.24)	(0.19)
15-18	0.08	0.16*	0.08	0.08	0.06	0.09	0.02	0.04	-0.04
	(0.06)	(0.09)	(0.08)	(0.08)	(0.10)	(0.06)	(0.21)	(0.22)	(0.14)
21-24	-0.03	-0.03	-0.03	-0.03	-0.11	-0.04	-0.02	-0.02	-0.08
	(0.04)	(0.06)	(0.04)	(0.05)	(0.08)	(0.04)	(0.12)	(0.12)	(0.07)
24-27	-0.06	-0.07	-0.06	-0.06	-0.08	-0.06	-0.01	0.04	-0.06
	(0.07)	(0.08)	(0.05)	(0.06)	(0.10)	(0.07)	(0.21)	(0.19)	(0.08)
> 27	-0.26**	-0.24	-0.26**	-0.26**	-0.33**	-0.25**	-0.25	-0.15	-0.40***
	(0.11)	(0.15)	(0.10)	(0.11)	(0.16)	(0.11)	(0.27)	(0.25)	(0.12)
Precipitation	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Humidity	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Daylight	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
Individual FE	No	No	No	No	No	No	Yes	Yes	No
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	117,428	54,828	117,428	117,428	67,110	117,428	104,997	104,997	117,428

Notes: This table presents estimated coefficients and standard errors from regressing total working hours for all workers on temperature bins. Column (1) considers our baseline model. Column (2) excludes agriculture, forestry, fishing, and construction from the sample. Column (3) considers a model with Conley standard errors correcting for spatial and temporal correlation. Column (4) considers a model with standard errors clustered at the district level. Column (5) considers only individuals who are currently living in the same district they were born. Column (6) considers our baseline model using CHIRPS precipitation data. Column (7) replicates the empirical model in Garg et al. (2020). Columns (8) and (9) replicate the empirical model in Schwarz (2018). Estimated standard errors, reported in parentheses, are clustered at the region-month level, except for Columns (3), (4), (7), (8) and (9) where errors are clustered at the district level. Significance at the one, five, and ten percent levels is indicated by ***, **, and *, respectively.

3.6 Conclusion

In this study, I investigate the effect of temperature on working hours in Peru for the period 2007-2015. Using worker and meteorological data for Peru, I find evidence that in general working hours are negatively affected by hot temperatures. I also document that the negative effect of high temperatures on work time is driven by informal jobs instead of jobs in industries highly exposed to the weather. Lastly, labor market segmentation can lead to misleading conclusions regarding intertemporal labor supply. Aggregate intertemporal labor substitution is masking the positive effect of high temperatures on formal workers with the negative effect on informal workers. Given the differential impact of temperature on working time between informal and formal jobs there is room for public policies to compensate the former for the hours of work lost due to heat. In addition, policies should promote mechanisms that encourage work time substitution across weeks, in particular for those with formal jobs.

There are some caveats in this study that would need to be addressed in further research. For instance, behavioral explanations should be taken into account (e.g. individuals not changing their working hours due to changes in temperature because they are following deadlines or are reluctant to look like shirking in the workplace). Second, a key assumption in the study is that individuals work and live in the same district. Hence, a measure of the geographic location of individuals' workplaces would improve the study. Third, given the nature of the data, I also cannot rule out an intra-day substitution of working time. Fourth, (out/in) migration could change the composition of the labor market over time. Fifth, given the correlation between informal jobs and jobs in "outdoor" industries, there is a need for new methods to better identify who is working outdoors. It is essential to thoroughly pin down this group to understand the impacts of temperature on labor supply. Finally, this is a short-run analysis that does not fully capture the gradual long-term changes in temperature due to climate change and behavioral responses. Therefore, understanding long-term dynamics in the labor market due to weather shocks is necessary to identify better the effects of climate change.

Bibliography

Akbar, P., V. Couture, G. Duranton, and A. Storeygard (2020). Mobility and Congestion in World Cities: Evidence from Google Maps. *work in progress.*

Aragao, L. E. O., Y. Malhi, N. Barbier, A. Lima, Y. Shimabukuro, L. Anderson, and S. Saatchi (2008). Interactions between rainfall, deforestation and fires during recent years in the brazilian amazonia. *Philosophical Transactions of the Royal Society B: Biological Sciences 363*(1498), 1779–1785.

Aragon, F., F. Oteiza, and J. P. Rud (2020). Climate change and agriculture: Farmer responses to extreme heat. *American Economic Journal: Economic Policy*, Forthcoming.

Armas, A., J. Borner, M. Rugnitz Tito, L. Diaz Cubas, S. Tapia Coral, S. Wunder, L. Reymond, and N. Nascimento (2009). Inventario nacional de gases de efecto invernadero del año 2016. Technical report, SERNANP.

Arnott, R., A. De Palma, and R. Lindsey (1990). Economics of a bottleneck. *Journal of urban economics 27*(1), 111–130.

Arnott, R., A. De Palma, and R. Lindsey (1993). A structural model of peak-period congestion: A traffic bottleneck with elastic demand. *The American Economic Review*, 161–179.

Arnott, R. and K. Small (1994). The economics of traffic congestion. *American scientist 82*(5), 446–455.

Auffhammer, M., S. Hsiang, W. Schlenker, and A. Sobel (2013). Using weather data and climate model output in economic analyses of climate change. *Review of Environmental Economics and Policy 7*(2), 181–198.

Aybar, C., C. FernÃ¡ndez, A. Huerta, W. Lavado, F. Vega, and O. Felipe-Obando (2019). Construction of a high-resolution gridded rainfall dataset for peru from 1981 to the present day. *Hydrological Sciences Journal.*

Becker, G. S. (1965). A theory of the allocation of time. *The economic journal 75*(299), 493–517.

Behrman, J. R. (1999). Labor markets in developing countries. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3, Chapter 43.

Beland, L.-P. and D. A. Brent (2018). Traffic and crime. *Journal of Public Economics 160*, 96–116.

Bento, A., K. Roth, and A. R. Waxman (2020). Avoiding traffic congestion externalities? the value of urgency. Technical report, National Bureau of Economic Research.

Berazneva, J. and T. S. Byker (2017). Does forest loss increase human disease? evidence from nigeria. *American Economic Review 107*(5), 516–21.

Black, D. A., N. Kolesnikova, and L. J. Taylor (2014). Why do so few women work in new york (and so many in minneapolis)? labor supply of married women across us cities. *Journal of Urban Economics 79*, 59–71.

Blackman, A., J. A. Bonilla, and L. Villalobos (2023). Quantifying covid-19âs silver lining: Avoided deaths from air quality improvements in bogotá. *Journal of Environmental Economics and Management 117*, 102749.

Brancalion, P. H., E. N. Broadbent, S. De-Miguel, A. Cardil, M. R. Rosa, C. T. Almeida, D. R. Almeida, S. Chakravarty, M. Zhou, J. G. Gamarra, et al. (2020). Emerging threats linking tropical deforestation and the covid-19 pandemic. *Perspectives in ecology and conservation 18*(4), 243–246.

Brodeur, A., N. Cook, and T. Wright (2021). On the effects of covid-19 safer-at-home policies on social distancing, car crashes and pollution. *Journal of environmental economics and management 106*, 102427.

Burgess, R., M. Hansen, B. A. Olken, P. Potapov, and S. Sieber (2012). The political economy of deforestation in the tropics. *The Quarterly journal of economics 127*(4), 1707–1754.

Busch, J. and K. Ferretti-Gallon (2017). What drives deforestation and what stops it? a meta-analysis. *Review of Environmental Economics and Policy 11*(1), 3–23.

Calatayud, A., S. Sanchez Gonzalez, F. Bedoya Maya, F. Giraldez Zuniga, and J. M. Marquez (2021). *Congestion urbana en America Latina y el Caribe: Caracteristicas, costos y mitigacion.* IDB.

Carleton, T. A., A. Jina, M. T. Delgado, M. Greenstone, T. Houser, S. M. Hsiang, A. Hultgren, R. E. Kopp, K. E. McCusker, I. B. Nath, J. Rising, A. Rode, H. K. Seo, A. Viaene, J. Yuan, and A. T. Zhang (2020, July). Valuing the global mortality consequences of climate change accounting for adaptation costs and benefits. Working Paper 27599, National Bureau of Economic Research.

Colella, F., R. Lalive, S. O. Sakalli, and M. Thoenig (2019). Inference with arbitrary clustering.

Colmer, J. (2020). Temperature, labor reallocation, and industrial production: Evidence from india. *American Economic Journal: Applied Economics*, forthcoming.

Conley, T. G. (1999). Gmm estimation with cross sectional dependence. *Journal of econometrics 92*(1), 1–45.

Conley, T. G. (2010). Spatial econometrics. In *Microeconometrics*, pp. 303–313. Springer.

Connolly, M. (2008). Here comes the rain again: Weather and intertemporal substitution of leisure. *Journal of Labor Economics 26*(1), 73–100.

Couture, V., J. Dingel, A. Green, J. Handbury, and K. Williams (2022). Measuring movement and social contact with smartphone data: a real-time application to covid-19. *Journal of Urban Economics 127*.

Dang, H.-A. H. and T.-A. Trinh (2021). Does the covid-19 lockdown improve global air quality? new cross-national evidence on its unintended consequences. *Journal of Environmental Economics and Management 105*, 102401.

Dell, M., B. F. Jones, and B. A. Olken (2014). What to we learn from the weather? the new climate-economy literature. *Journal of Economic Literature 52*(3), 740–798.

Deschenes, O. and M. Greenstone (2007). The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather. *American Economic Review 97*(1), 354–385.

ECLAC (2022). Informacion para el planeamiento a nivel departamental, provincial y distrital. Website: https://www.ceplan.gob.pe/informacion-sobre-zonas-y-departamentos-del-peru/, accessed on September 22nd, 2022.

Farr, T. G., P. A. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, M. Paller, E. Rodriguez, L. Roth, et al. (2007). The shuttle radar topography mission. *Reviews of geophysics 45*(2).

Finer, M. and S. Novoa (2017). Patrones y drivers de deforestacion en la amazonia peruana. MAAP: Sintesis 2. Accessed July 15, 2022. https://www.maaproject.org/2017/maap-sintesis2/.

Fisher, A. C., W. M. Hanemann, M. J. Roberts, and W. Schlenker (2012, December). The economic impacts of climate change: Evidence from agricultural output and random fluctuations in weather: Comment. *American Economic Review 102*(7), 3749–60.

Fort, R., M. Espinoza, and A. Espinoza (2021). Covid-19 y las migraciones de la ciudad al campo en el peru: Identificacion de amenazas y oportunidades para el uso sostenible del capital natural (resumen).

Fu, S. and B. Viard (2019). Commute costs and labor supply: evidence from a satellite campus. *Journal of Economic Geography 19*(3), 723–752.

Garg, T., M. Gibson, and F. Sun (2020). Extreme temperatures and time use in china. *Journal of Economic Behavior and Organization 180*, 309–324.

Guiteras, R. (2009). The impact of climate change on indian agriculture. Working paper.

Gutiérrez-i-Puigarnau, E. and J. N. van Ommeren (2010). Labour supply and commuting. *Journal of Urban Economics 68*(1), 82–89.

Gutiérrez-i-Puigarnau, E. and J. N. van Ommeren (2015). Commuting and labour supply revisited. *Urban Studies 52*(14), 2551–2563.

Hanna, R., G. Kreindler, and B. A. Olken (2017, July). Citywide effects of high-occupancy vehicle restrictions: Evidence from "three-in-one" in Jakarta. *Science 357*(6346), 89–93.

Heal, G. and J. Park (2016). Temperature stress and the direct impact of climate change: A review of an emerging literature. *Review of Environmental Economics and Policy 10*(2), 347–362.

Higa, M., C. Ospino, and F. Aragon (2022). The persistent effects of covid-19 on labour outcomes: evidence from peru. *Applied Economics Letters 0*(0), 1–12.

IGN (2022). Geoportal - ign peru. Website: https://www.idep.gob.pe/geovisor/descarga/visor.html, accessed September 22nd, 2022.

ILO (2019). Working on a warmer planet: The impact of heat stress on labour productivity and decent work. Report, International Labour Organization.

INRIX (2019). Scorecard, global traffic.

Jayachandran, S. (2013). Liquidity constraints and deforestation: The limitations of payments for ecosystem services. *American Economic Review 103*(3), 309–313.

Jessoe, K., D. T. Manning, and J. E. Taylor (2016). Climate change and labour allocation in rural mexico: Evidence from annual fluctuations in weather. *The Economic Journal 128*(February), 230â–261.

Keenan, R. J., G. A. Reams, F. Achard, J. V. de Freitas, A. Grainger, and E. Lindquist (2015). Dynamics of global forest area: Results from the fao global forest resources assessment 2015. *Forest Ecology and Management 352*, 9–20. Changes in Global Forest Resources from 1990 to 2015.

Kim, J. (2019). Estimating the social cost of congestion using the bottleneck model. *Economics of Transportation 19*, 100119.

Kreindler, G. (2022). Peak-hour road congestion pricing: Experimental evidence and equilibrium implications. *Working paper*.

Kruger, J. J. and M. Neugart (2018). Weather and intertemporal labor supply: Results from german time-use data. *Labour 32*(1), 112–140.

LoPalo, M. (2020). Temperature, worker productivity, and adaptation: Evidence from survey data production. Working paper.

Lopez-Feldman, A., C. Chavez, M. A. Velez, H. Bejarano, A. B. Chimeli, J. Feres, J. Robalino, R. Salcedo, and C. Viteri (2020). Environmental impacts and policy responses to covid-19: a view from latin america. *Environmental & resource economics*, 1.

Madhok, R. and S. Gulati (2022). Ruling the roost: Avian species reclaim urban habitat during india's covid-19 lockdown. *Biological Conservation*, 109597.

Malaga, N., R. G. Granados, C. V. Gonzales, and E. R. Bae (2014). Estimacion de los contenidos de carbono de la biomasa aerea en los bosques de peru. Technical report, MINAM.

MEF (2021). Nota tecnica para el uso del precio social del carbono en la evaluacion social de proyectos de inversion. Technical report.

MEF (2022). Transparencia economica peru - consulta de ejecucion del gasto mensual. Website: https://apps5.mineco.gob.pe/transparencia/mensual/, accessed September 22nd, 2022.

MINAM (2015). Intended nationally determined contribution (indc) from the republic of peru. Technical report.

MINAM (2016). Zonas con mineria ilegal e informal. Technical report.

MINAM (2022). Geobosques. plataforma de monitoreo de cambios sobre la cobertura de bosques. Website: http://geobosques.minam.gob.pe/, accessed September 22nd, 2022.

MINEM (2013). Habitos de consumo de energia en el sector publico, productivo y de servicios y oferta de equipos elÃ©ctricos eficientes. Technical report.

MINSA (2022). Datos abiertos minsa y gestion del conocimiento en covid-19. Website: https://www.minsa.gob.pe/datosabiertos/, accessed September 22nd, 2022.

MTC (2022). Descarga de datos espaciales. Website: https://portal.mtc.gob.pe/estadisticas/descarga.html, accessed September 22nd, 2022.

Munoz Sabater, J. (2019). Era5-land hourly data from 1981 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), https://cds.climate.copernicus.eu/cdsapp#!/home.

Noland, R. and K. A. Small (1995). Travel-time uncertainty, departure time choice, and the cost of morning commutes. *Transportation research record 1493*, 150–158.

Piotrowski, M. (2019). Nearing the tipping point. drivers of deforestation in the amazon region. Technical report, Andes Amazon Fund.

Potapov, P., J. Dempewolf, Y. Talero, M. Hansen, S. Stehman, C. Vargas, E. Rojas, D. Castillo, E. Mendoza, A. Calderon, et al. (2014). National satellite-based humid tropical forest change assessment in peru in support of redd+ implementation. *Environmental Research Letters 9*(12), 124012.

Rambachan, A. and J. Roth (2023). A more credible approach to parallel trends. *Review of Economic Studies*, rdad018.

Saavedra, S. (2020). Do covid-19 lockdowns affect deforestation? *Working paper*.

Schleicher, J., C. A. Peres, T. Amano, W. Llactayo, and N. Leader-Williams (2017). Conservation performance of different conservation governance regimes in the peruvian amazon. *Scientific reports 7*(1), 1–10.

Schwarz, A. I. (2018). Too hot to handle? weather as a driving force of labour market fluctuations in mexico. Working paper.

Sears, R. R. and M. Pinedo-Vasquez (2011). Forest policy reform and the organization of logging in peruvian amazonia. *Development and Change 42*(2), 609–631.

SERNANP (2022). Geo anp - visor de las areas naturales protegidas. Website: https://geo.sernanp.gob.pe/visorsernanp/,accessed September 22nd, 2022.

Shukla, P. R., J. Skea, E. Calvo Buendia, V. Masson-Delmotte, H. O. Pörtner, D. Roberts, P. Zhai, R. Slade, S. Connors, R. Van Diemen, et al. (2019). Ipcc, 2019: Climate change and land: an ipcc special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems.

Small, K. A. (1982). The scheduling of consumer activities: work trips. *The American Economic Review 72*(3), 467–479.

Swenson, J. J., C. E. Carter, J.-C. Domec, and C. I. Delgado (2011, 04). Gold mining in the peruvian amazon: Global prices, deforestation, and mercury imports. *PLOS ONE 6*(4), 1–7.

UNODC (2017). Peru - monitoreo de cultivos de coca 2017. Technical report.

Vale, M. M., E. Berenguer, M. A. de Menezes, E. B. V. de Castro, L. P. de Siqueira, and Q. P. Rita de Cássia (2021). The covid-19 pandemic as an opportunity to weaken environmental protection in brazil. *Biological Conservation 255*, 108994.

Van Vliet, N., J. L. Moreno Calderón, J. Gomez, W. Zhou, J. E. Fa, C. Golden, R. R. Nobrega Alves, and R. Nasi (2017). Bushmeat and human health: assessing the evidence in tropical and sub-tropical forests.

Vargas, R., C. Fonseca, G. Hareau, M. Ordinola, W. Pradel, V. Robiglio, and V. Suarez (2021). Health crisis and quarantine measures in peru: Effects on livelihoods of coffee and potato farmers. *Agricultural Systems 187*, 103033.

Vera, E. (2020). Madre de dios: nuevo foco de mineria ilegal amenaza a indigenas del pariamanu.

Viard, B. and S. Fu (2015). The effect of beijing's driving restrictions on pollution and economic activity. *Journal of Public Economics 125*, 98–115.

Vickrey, W. (1969). Congestion theory and transport investment. *The American Economic Review 59*(2), 251–260.

Wolff, H. (2014). Value of time: Speeding behavior and gasoline prices. *Journal of Environmental Economics and Management 67*(1), 71–88.

Wunder, S., D. Kaimowitz, S. Jensen, and S. Feder (2021). Coronavirus, macroeconomy, and forests: What likely impacts? *Forest Policy and Economics 131*, 102536.

Young, K. R. (1996). Threats to biological diversity caused by coca/cocaine deforestation in peru. *Environmental Conservation 23*(1), 7â15.

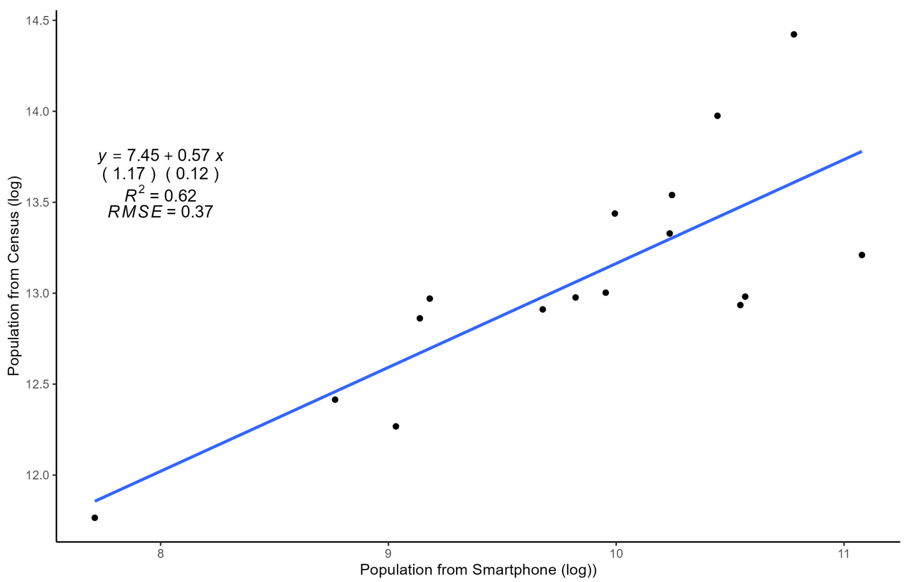
Zivin, J. G. and M. Neidell (2014). Temperature and the allocation of time: Implications for climate change. *Journal of Labor Economics 32*(1), 1–26.

Appendix A

Chapter 1

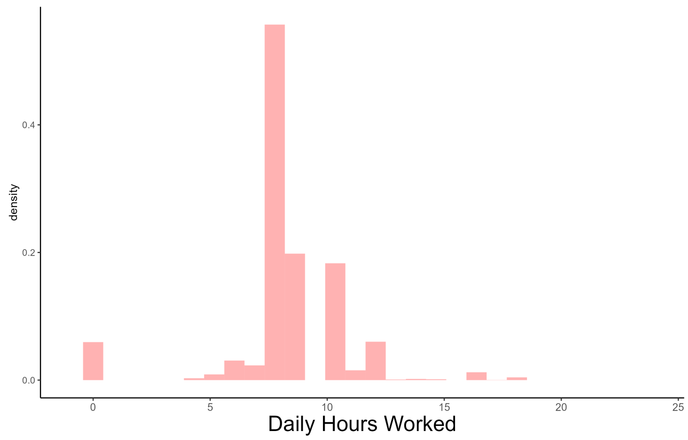
A.1 Additional figures

Figure A.1: Correlation of population size from smartphone and census data

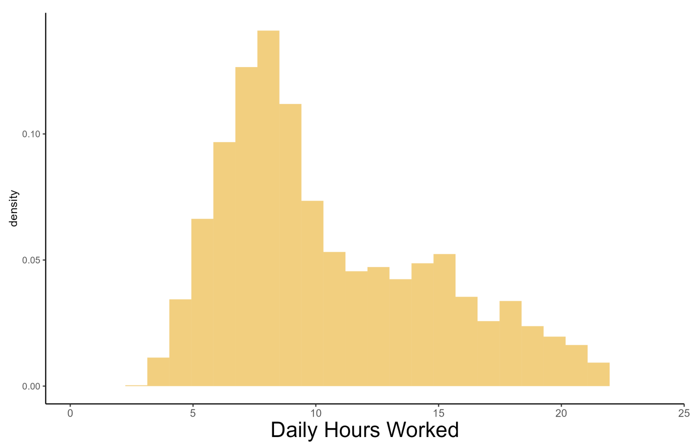


Notes: The figure depicts a comparison of the total population in Mexico City according to the Census 2020 (vertical axis) with the total number of smartphone devices in 2019. Each dot is one of the 16 districts. Linear regression line in blue.

Figure A.2: Distribution of daily hours worked: ENOE vs. Smartphone



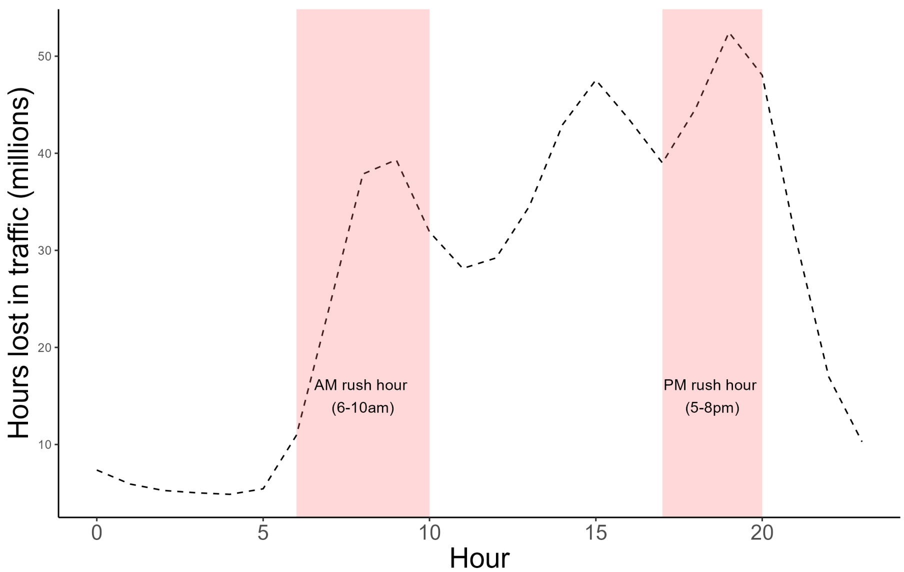
(a) Labor survey (ENOE)



(b) Smartphone

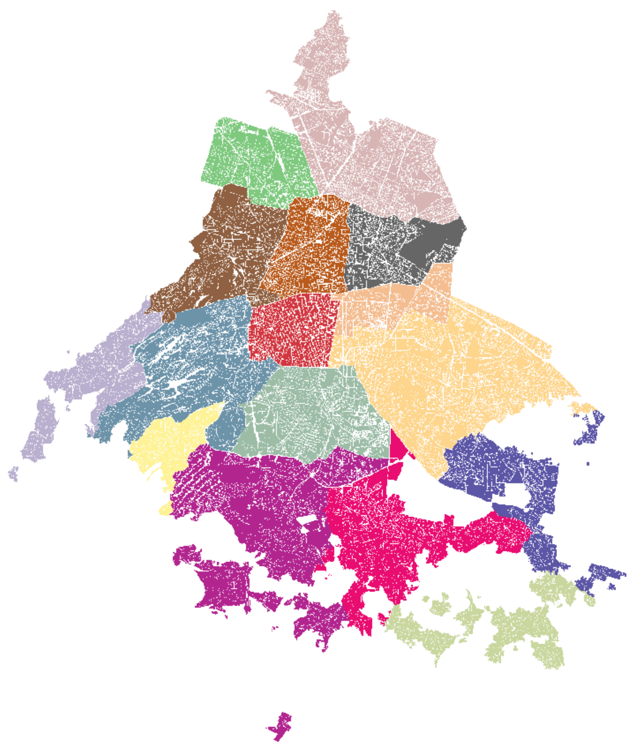
Notes: The figure depicts self-reported daily hours worked from the labor household survey ENOE (panel a) and daily hours spent at the workplace from the smartphone data (panel b) both conditional on being manufacturing or office workers in firms with 50 or more employees. The first week of January and the last week of December are excluded, as well as Saturdays and Sundays.

Figure A.3: Distribution of traffic congestion using data from Waze



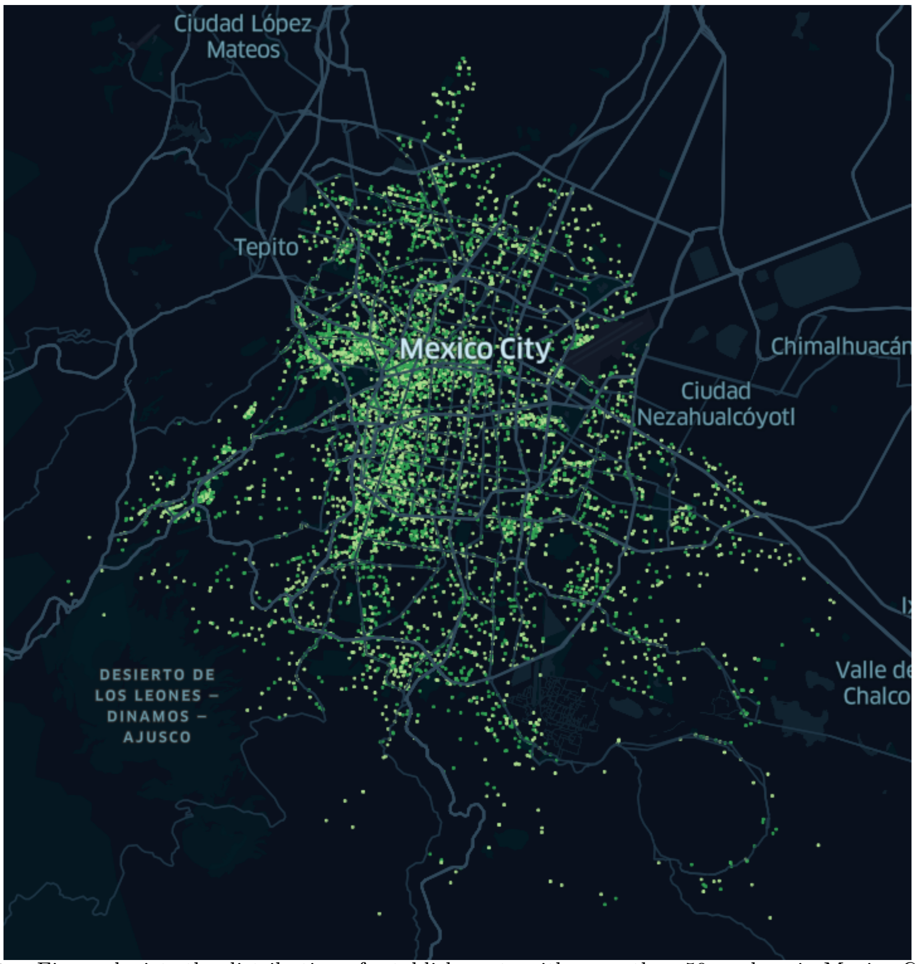
Notes: The figure depicts the distribution of traffic congestion per hour using data from Waze for Mexico City in 2019. Data provided by Calatayud et al. (2021). Morning (6 am-10 am) and afternoon (5 pm-8 pm) rush hours are highlighted in yellow.

Figure A.4: Map of Mexico City



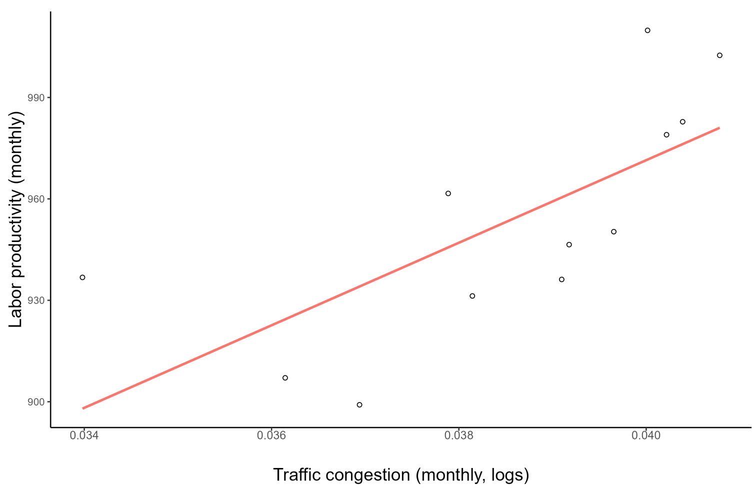
Notes: Figure depicts the map of Mexico City. Each color represents one of the 16 districts.

Figure A.5: Distribution of establishments



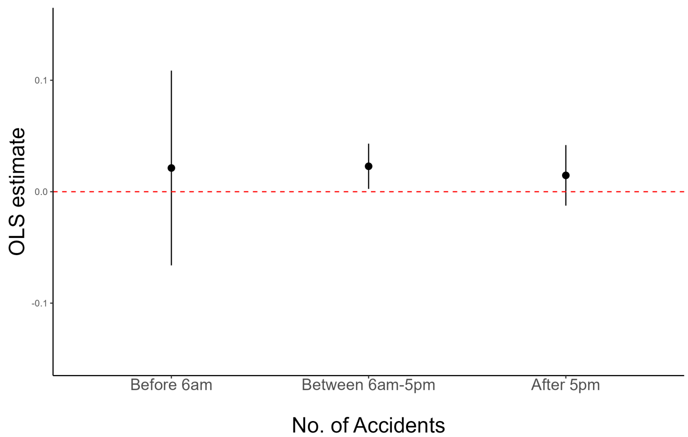
Notes: Figure depicts the distribution of establishments with more than 50 workers in Mexico City. The location of establishments is provided by the National Statistical Directory of Economic Units (DENUE) 2019.

Figure A.6: Correlation between traffic congestion and labor productivity

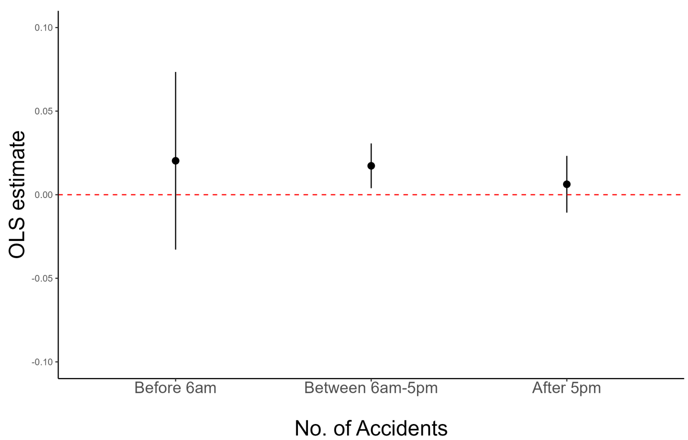


Notes: Figure depicts the correlation between monthly traffic congestion and monthly labor productivity (total output value divided by total hours worked). Each dot represents information for a month in 2019. Labor productivity comes from monthly manufacturing firm surveys (EMIM). Linear regression line in red.

Figure A.7: The effect of road accidents on hours worked and departure time from work



(a) Hours worked



(b) Departure time

Notes: The figure depicts see OLS estimates and 90% confidence intervals. See regressions output in Table A.3.

A.2 Additional tables

Table A.1: Additional robustness analysis for main results

	Dependent variable: Hours worked	
	(1)	(2)
Traffic Congestion (log)	0.756*	3.483
	(0.435)	(6.579)
Change in specification	Distance >4km	2SLS + individual FE
Observations	1,159	6,242
R-squared	0.035	0.022

Notes: Standard errors clustered at week-district level in parentheses. Statistical significance at the one, five, and ten percent levels is indicated by ***, **, and *, respectively. All regressions use the baseline model and smartphone data.

Table A.2: Additional robustness analysis for mitigation and adaptation

	Hours worked (weekly)		Arrival time		Departure time	
	High	Low	High	Low	High	Low
	(1)	(2)	(3)	(4)	(5)	(6)
Congestion (weekly, log)	-3.694	10.296***				
	(5.685)	(2.258)				
Congestion (log)			0.977	-1.740***	0.452	1.065***
			(0.676)	(0.344)	(0.899)	(0.406)
Observations	1,474	1,197	3,508	2,825	3,508	2,825
R-squared	0.059	0.108	0.070	0.167	0.026	0.077

Notes: Standard errors clustered at week-district level in parentheses. Statistical significance at the one, five, and ten percent levels is indicated by ***, **, and *, respectively. All regressions use the baseline model and smartphone data.

Table A.3: The effect of road accidents on hours worked and departure time from work

	Hours worked	Arrival Time	Departure time	Congestion (log)
	(1)	(2)	(3)	(4)
No. of accidents				
before AM rush hour	0.021	-0.001	0.020	0.010***
(before 6am)	(0.053)	(0.029)	(0.032)	(0.003)
during and after AM rush hour	0.023*	-0.006	0.017**	0.010***
(6am-5pm)	(0.012)	(0.006)	(0.008)	(0.001)
during and after PM rush hour	0.015	-0.008	0.006	0.008***
(5pm-midnight)	(0.016)	(0.009)	(0.010)	(0.001)
Observations	6,709	6,709	6,709	6,333
R-squared	0.088	0.104	0.038	0.433

Notes: Standard errors clustered at week-district level in parentheses. Statistical significance at the one, five, and ten percent levels is indicated by ***, **, and *, respectively. All regressions use the baseline model and smartphone data.

Table A.4: The effect of traffic congestion on labor income (ENOE)

	Labor income (monthly, log)		
	(1)	(2)	(3)
Traffic congestion (monthly, log)	-0.506**	-1.689	-1.389**
	-0.217	-1.85	-0.688
Method	OLS	OLS	2SLS
Fixed effects	Month	Month, individual	Month
Observations	684	684	684
R-squared	0.035	0.982	-0.015

Notes: Standard errors clustered at month-district level in parentheses. Statistical significance at the one, five, and ten percent levels is indicated by ***, **, and *, respectively. Traffic congestion is measured as the inverse of the daily average speed at the district level. All regressions include monthly weather (temperature, precipitation, humidity), monthly daylight hours, and month fixed-effects, and exclude income from the top 1% of earners. In column (3), the Kleibergen-Paap rk Wald F-statistic is 10.77, and the first-stage coefficient is 0.0001 and statistically significant at the one percent level.

A.3 Data appendix

A.3.1 Supplemental Data

Weather data I use hourly data about temperature, precipitation, and relative humidity recorded by ground stations in Mexico City. The data was provided by the national meteorological agency (CONAGUA).

Daylight hours I calculate daylight hours for each day of the working week by taking the difference between the sunset and sunshine time at each district location. Sunset and sunshine times were calculated using astronomical algorithms taking as input the date each individual worked and the geographic coordinates of the districts where they work. I use the package suncalc in R to perform these calculations.

ENOE It is the main labor market household survey in Mexico providing monthly and quarterly information. It is conducted by the National Statistics Office (INEGI) and it collects information from individuals aged 15 years or more continuously every week from Monday to Sunday throughout the year. It has a rotating panel design where every five quarters 20% of the sample is replaced. The quarterly sample size is around 126,000 housing units. It is representative at the national level, and also at the level of cities such as Mexico City.

Accidents I use administrative records about road accidents collected by the centralized emergency center in Mexico City under the supervision of the local government in Mexico City. The administrative records contain information about the location (latitude and longitude coordinates), the date, and the time of the road accident. It also provides information about the type of accidents and whether the accident involved victims among other details.

EMIM It is a monthly establishment-level survey representative of the manufacturing sector in Mexico. All establishments report information about the number of employees, earnings, output value, and sales among other economic characteristics. They report this information every month of the year. The sample size for 2019 is 10,447.

A.3.2 Smartphone and Traffic Congestion Data Providers

Quadrant It is a global leader in mobile location data, POI data, and corresponding compliance services. Quadrant provides anonymized location data solutions that are fit for purpose, authentic, easy to use, and simple to organize. They offer data for almost all countries in the world, with hundreds of millions of unique devices and tens of billions of events per month, allowing our clients to perform location analyses, derive location-based intelligence, and make well-informed business decisions. Their data is gathered directly from first-party opt-in mobile devices through a server-to-server integration with trusted publisher partners, delivering genuine and reliable raw GPS data, unlike other location data sources. Their consent management platform, QCMP, ensures that their data is compliant with applicable consent and opt-out provisions of data privacy laws governing the collection and use of location data. More information about the company can be found here: https://www.quadrant.io/

Dat's Why It is a leading mobility intelligence platform with +70M smartphones, vehicles, and sensors collecting in real-time +40B data points annually in Latin America. Using its real-time Big Data network of Geobehavior, the largest in Mexico, monitors various traffic parameters and creates smart mobility solutions and analytics. More information about the company can be found here: https://datswhy.com/

A.3.3 Identify Work Location

- Step 1: Initial sample selection
  - Using SQL in Amazon AWS, select those devices with error in location accuracy < 50m, which is approximately half of a street block.
  - Select those devices observed more than seven days in a month to avoid tourists or sporadic users.
  - Select those devices observed at least twice a day to potentially know the arrival and departure time from a location (e.g. home or work).
- Step 2: Location of establishments
  - Use geocoded establishment-level data from DENUE that provides the latitude and longitude coordinates for each establishment. Use the location of the establishment as a point of interest (POI).
  - Using Python, draw a circular geofence of radius 50m around the POI.
  - Using Python, convert the POI with circular geofence into geohash grids (precision 8, +- 20m).
- Step 3: Combine the smartphone data with the establishment-level data using the geohash grids.
- Step 4: Use algorithm inspired in Couture et al. (2022)
  - Use pings observed from Monday to Friday and between 9 am and 5 pm. These are the days and times for regular daytime work. Note that the sample here is restricted only for the purposes of finding the work location. For the statistical analysis, we use all pings observed during the entire day.
  - Calculate how much time each device spends at workplaces (i.e circular geofences around the POI). As a result, one device may have more than one candidate as a potential work location given that individuals move around.
  - Assign the device to the workplace venue with the longest duration.
  - If the duration is 0, then the workplace is the work location with the most daytime visits.
  - Finally, the device needs to visit this workplace venue at least 3 times a week.

A.3.4 Identify Home Location

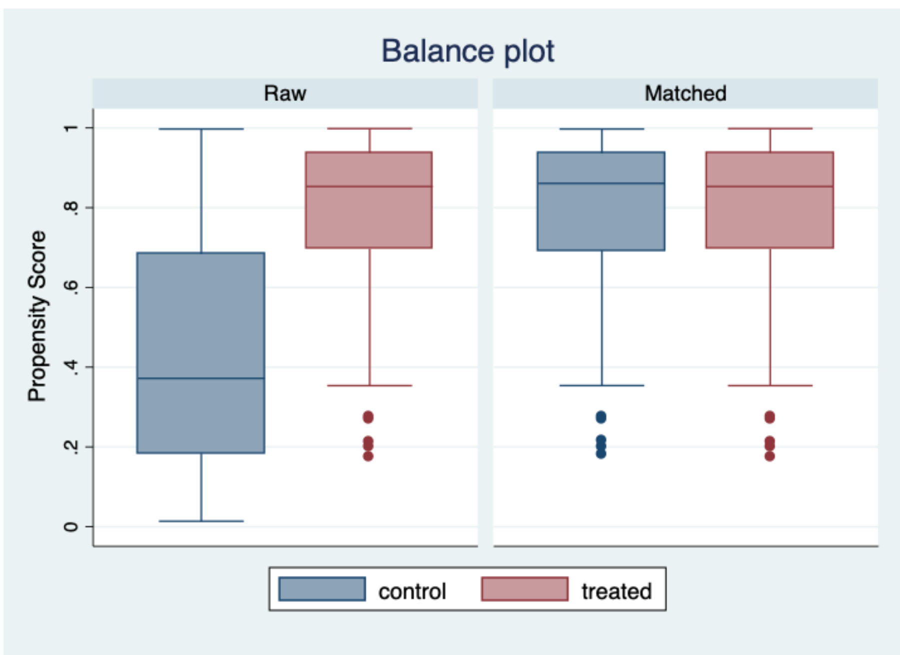
- Step 1: Initial sample selection
  - Using SQL in Amazon AWS, select those devices with error in location accuracy < 50m, which is approximately half of a street block.
  - Select those devices observed more than seven days in a month to avoid tourists or sporadic users.
  - Select those devices observed at least twice a day to potentially know the arrival and departure time from a location (e.g. home or work).
- Step 2: Location of residential areas
  - Using Python, convert the block from the National Geostatistical Framework (MGN) into polygon geofences.
  - Use Census to identify which blocks contain residential places.
  - Using Python, convert the polygon geofence into geohash grids (precision 8, +- 20m).
- Step 3: Combine the smartphone data with the residential data using the geohash grids.
- Step 4: Use algorithm inspired in Couture et al. (2022)
  - Select blocks with inhabited private homes regardless of whether there are establishments as well in the block.
  - Use pings observed between 9 pm and 5 am. Presumably, these are the times when daytime workers are at home. Note that the sample here is restricted only for the purposes of finding the work location. For the statistical analysis, we use all pings observed during the entire day.
  - Calculate how much time each device spends at potential home locations. As a result, one device may have more than one candidate as a potential home location given that individuals move around.
  - Assign the device to the residential venue with the longest duration.
  - If the duration is 0, then the residential place is the home location with the most daytime visits.
  - Finally, the device needs to visit this residential venue at least 3 times a week.

Appendix B

Chapter 2

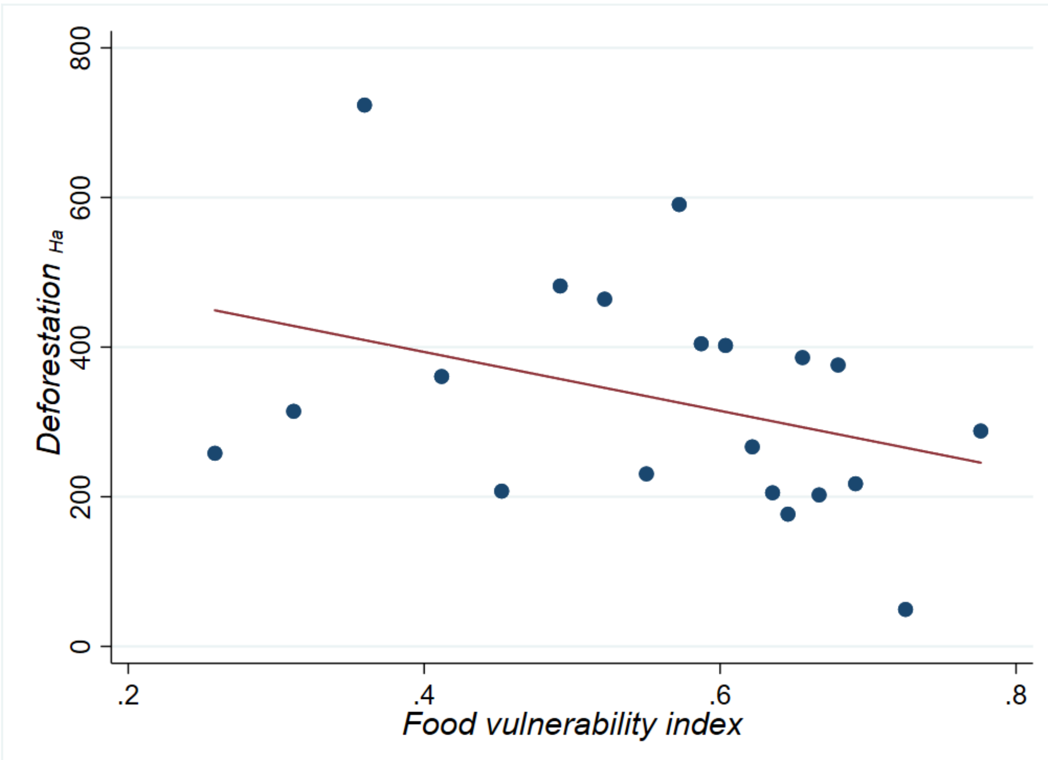
B.1 Additional figures

Figure B.1: Balance plot - COVID-19 Cases



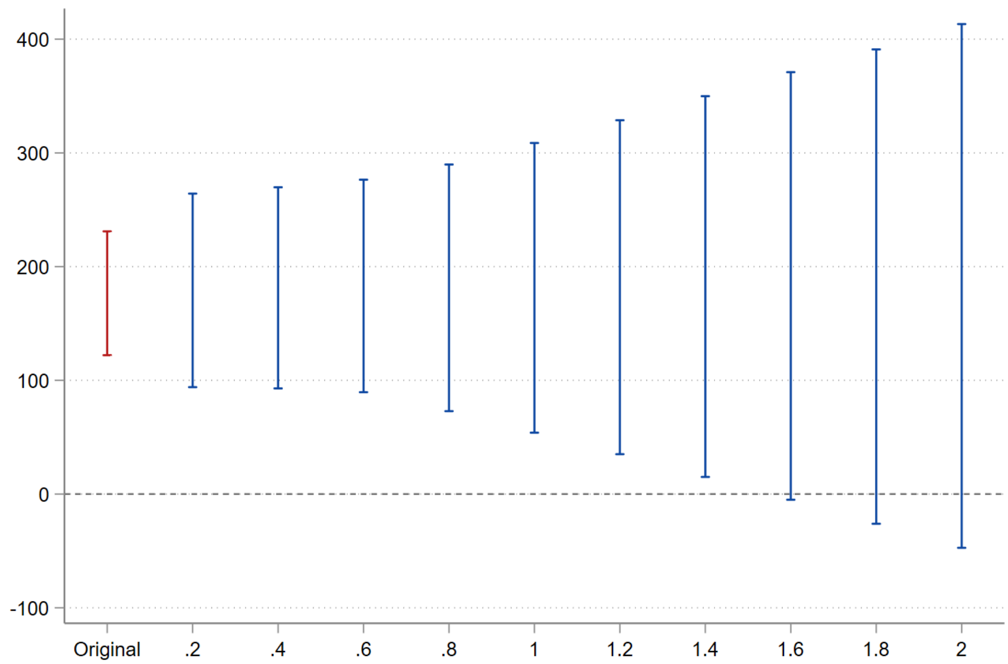
Notes: Sample balance before (left panel) and after (right panel) applying the propensity score matching procedure, using COVID-19 cases number per district as treatment variable (1 = above the country median).

Figure B.2: Deforestation and Food Vulnerability Index



Notes: Figure depicts a binscatter with deforestation (ha) in the vertical axis and food vulnerability index in the horizontal axis. Both are for the year 2018. *Source:* CEPLAN, 2022, MINAM, 2022

Figure B.3: Sensitivity analysis of violations of parallel trends



Notes: Figure depicts a sensitivity analysis of post-treatment violations of parallel trends. It displays in red the confidence interval for the estimated coefficient associated with the year 2020 in the original event study in Figure 3.1. It displays in blue robust confidence intervals that allow for post-treatment violation of parallel trends to be no more than some constant (e.g. 0.2, 0.4, etc.) larger than the maximum violation of parallel trends in the pre-treatment period.

B.2 Additional tables

Table B.1: Variables description and data sources

Variable	Description	Availability	Data Source
Panel A. Forest			
Deforestation	Forest cover reduction (hectares) in districts	2015-2020	MINAM, 2022
Panel B. Covid			
Cases	Number of people that tested positive to COVID-19 per district	2020	MINSA, 2022
Deaths	Number of people that died due to COVID-19 per district	2020	MINSA, 2022
Panel C. District characteristics			
Coca	Hectares of coca leaf production per district	2017	COVIDA, 2017
Mining	Districts with presence of illegal or informal mining	2016	MINAM, 2016
ANP	Districts with presence of a Natural Protected Area	2017	SERNAP, 2017
Roads	Number of national roads in the district	2017	MTC, 2022
Rivers	River's area in the district (km2)	2017	MTC, 2022
Population	Population in the district	2017 & 2020	CEPLAN, 2020
Area	District's area (km2)	2019	CEPLAN, 2020
HDI	Human development index	2015 & 2019	CEPLAN, 2020
Slope	Average slope of district	2007	Farr et al, 2007
Altitud	Average altitude of district (masl)	2020	CEPLAN, 2022

Notes: Table displays the description, availability, and source for the main variables. All variables are at the annual level.

Table B.2: Summary statistics by treated and control districts

	Treated districts	**Control districts**
	(1)	(2)
Panel A. Forest		
Deforestation	555.61	96.26
	(875.58)	(333.92)
Panel B. Covid		
Cases	503.26	8.88
	(1045.34)	(7.38)
Deaths	32.99	2.6
	(99.03)	(14.62)
Panel C. District characteristics		
Coca	540.58	323.64
	(842.53)	(647.61)
Mining	0.15	0.04
	(0.36)	(0.20)
ANP	0.17	0.06
	(4.69)	(0.24)
Roads	1.38	1.06
	(0.38)	(2.44)
Rivers	32.21	5.6
	(63.54)	(17.82)
Population (2020)	18796.32	3970.17
	(25626.16)	(6892.91)
Area	2732.18	956.52
	(4259.16)	(2330.02)
HDI (2019)	0.39	0.34
	(0.10)	(0.08)
Indigenous populations	2370.08	939.14
	(3923.48)	(2033.48)
Observations	257	142

Notes: Table displays summary statistics of the main variables. Column (1) shows the mean values for districts with COVID-19 cases above the median. Column (2) shows the mean values for districts with COVID-19 cases below the median. In Panel A, deforestation corresponds to 2015.

Table B.3: Main results with different periods

	Dependent variable: Deforestation (ha)	
	(1)	(2)
Year 2020	118.1***	
	(22.284)	
DiD		176.54***
		(27.76)
Design	Event study	Difference-in-Difference
Pre-pandemic period	2015-2019	2019
N	800	798
R-squared	0.940	0.18

Notes: Estimated standard errors, reported in parentheses, are clustered at the district level. Significance at the one, five and ten percent levels is indicated by ***, ** and *, respectively. Column (1) and Column(2) present the results of estimating Equation 2.1 and Equation 2.2, respectively.

Table B.4: Heterogeneous effects with all regressors

	Dependent variable: Annual rate of forest change, 2019-2020	
	(1)	(2)
Year 2020	2.19***	2.17***
	(0.21)	(0.21)
Mining	0.17*	0.18*
	(0.09)	(0.09)
Coca	0.21**	0.19*
	(0.10)	(0.10)
Protected areas	-0.34***	-0.44***
	(0.08)	(0.08)
Population density	-0.00	-0.00
	(0.00)	(0.00)
River area (m2)	-0.00***	-0.00***
	(0.00)	(0.00)
Altitud (m)	-0.00***	-0.00***
	(0.00)	(0.00)
Number of vias	-0.01	-0.01
	(0.01)	(0.01)
Distance to Lima	-0.00***	-0.00***
	(0.00)	(0.00)
Slope	-0.02***	-0.03***
	(0.01)	(0.01)
N	394	394
R-squared	0.53	0.54

Notes: Estimated standard errors, reported in parentheses, are clustered at the district level. Significance at the one, five, and ten percent levels is indicated by ***, **, and *, respectively.

Table B.5: CO2 emission and economic loss

	tCO_2(*millions*)	Economic loss (millions USD)
Lower bound	12.7	98.25
Average	17	131.38
Upper bound	21.3	164.53

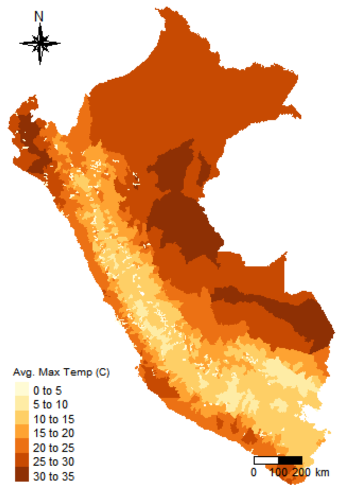
Notes: Table displays the estimates of $tCO_2\text{-}eq$ and economic losses (million USD) caused by the deforestation originated by COVID-19. We use the estimate from column 1 of Table **??**, as well as the confidence interval to obtain the lower and upper bound.

Appendix C

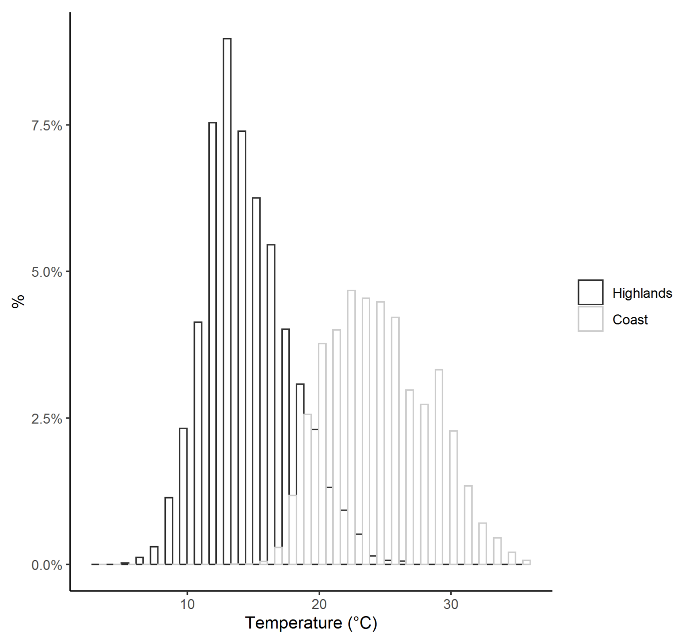
Chapter 3

C.1 Additional figures

Figure C.1: Temperature Distribution in Peru.



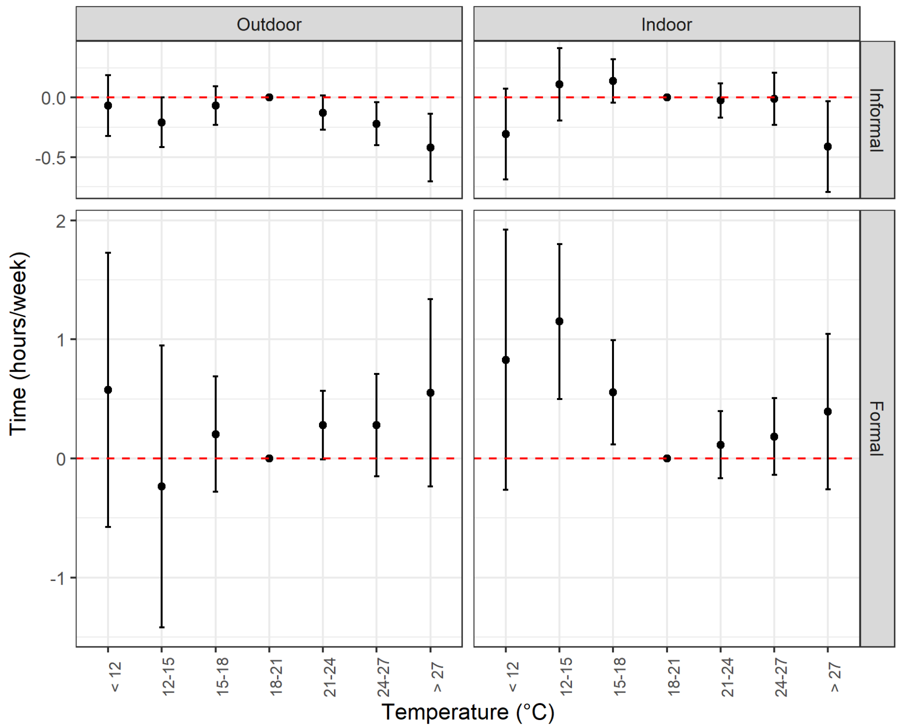
(a) Avg. Max Temperature (°C) in 2015 (district level)



(b) Coast vs. Highlands

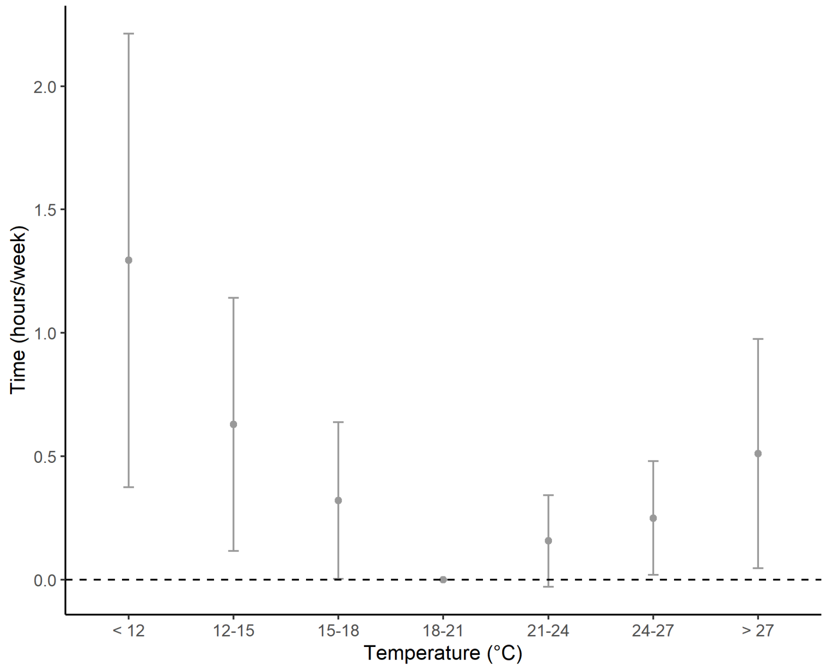
Notes: Figure depicts the temperature distribution for Peru. Panel (a) shows the average temperature for each district for the year 2015. Panel (b) shows the distribution of temperature on the coast and the highlands for the period 2007-2015. The figure uses maximum temperature and ERA5 data.

Figure C.2: Effect of Temperature on Work Time: Outdoor vs. Informal.

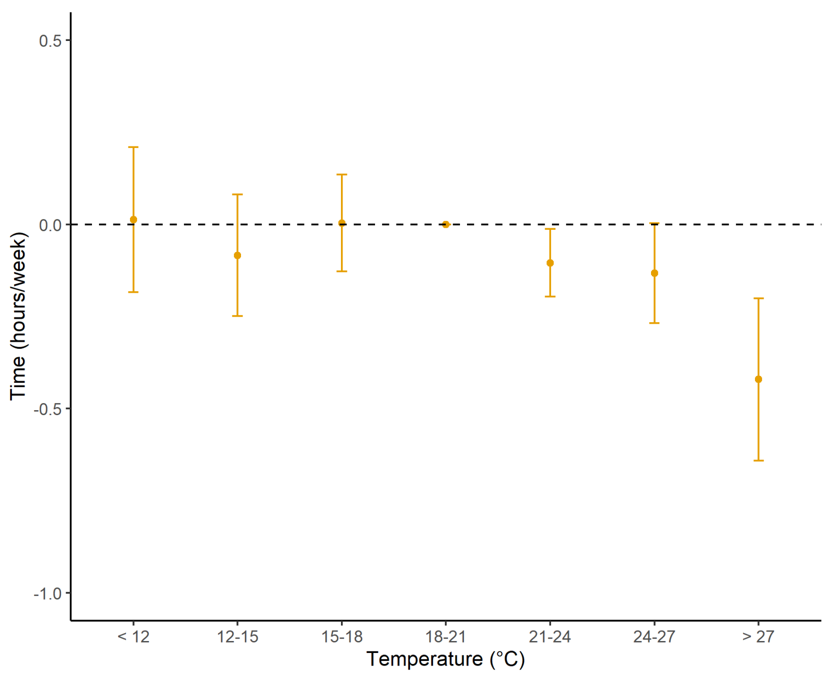


Notes: Figure depicts the estimates of the effect on hours worked of shifting a day from the reference bin (18-21°C) to a given bin during the working week. Outdoor jobs consider only workers in high-risk industries such as agriculture, fishing, mining, manufacturing, transportation, and utilities. Jobs are informal if the production unit is not registered for tax purposes or the worker is not covered by social security. Circles represent point estimates from regressing total working hours on temperature bins, controlling for precipitation, humidity, daylight hours, sociodemographics, and location and time fixed-effects. Vertical lines show 95 percent confidence intervals calculated using standard errors clustered at the region-month level. The figure uses maximum temperature, ERA5 data, and excludes the jungle.

Figure C.3: Intertemporal Labor Substitution: Informal vs. Formal



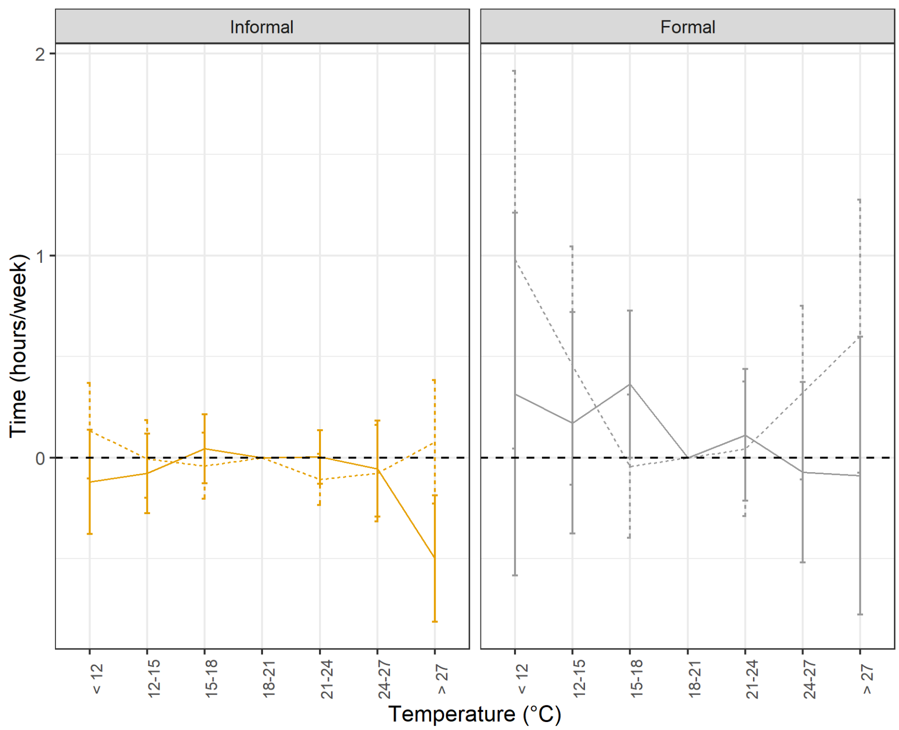
(a) Formal workers



(b) Informal workers

Notes: Figure depicts the cumulative effect of temperature on working hours across the contemporaneous and (one-week) lagged weeks for formal and informal workers. All regressions control for precipitation, humidity, daylight hours, sociodemographics, and location and time fixed-effects. Vertical lines show 90 percent confidence intervals calculated using standard errors clustered at the region-month level. The figure uses daily maximum temperature, ERA5 data, and excludes the jungle.

Figure C.4: Intertemporal Substitution of Working Hours: Informal vs. Formal.



Notes: Figure depicts the effects of contemporaneous (solid line) and one-week lagged (dashed line) temperature bins on working hours for workers with informal and formal jobs. The regression includes precipitation, humidity, daylight hours, sociodemographics, and location and time fixed-effects. Vertical lines show 95 percent confidence intervals calculated using standard errors clustered at the region-month level. The figure uses daily maximum temperature, ERA5 data, and excludes the jungle.

C.2 Additional tables

Table C.1: Summary Statistics.

	All (1)	**Informal** (2)	**Formal** (3)	**Outdoor** (4)	**Indoor** (5)
Hours (weekly)					
Main job	39.71	37.60	49.05	38.48	41.32
Secondary job	15.82	16.17	13.32	16.63	14.67
All jobs	43.46	41.73	51.13	42.56	44.64
Weather (average)					
Max Temperature (°C)	19.79	19.47	21.24	18.90	20.97
Min Temperature (°C)	12.21	11.65	14.70	10.85	13.99
Precipitation (mm)	30.94	35.25	11.85	39.99	19.10
Humidity (°C)	10.97	10.46	13.25	9.70	12.63
Daylight (hours)	12.10	12.10	12.08	12.10	12.10
Sociodemographics					
Income (logs)	9.84	9.71	10.40	9.61	10.13
Education (years)	8.94	8.23	12.10	7.90	10.30
Age (years)	38.87	38.87	38.89	39.62	37.90
Married	0.32	0.31	0.37	0.35	0.29
Female	0.45	0.47	0.35	0.32	0.62
Spouse at home	0.80	0.80	0.78	0.84	0.74
Number of children	0.43	0.47	0.29	0.47	0.39
Percent over age 65	0.08	0.08	0.06	0.09	0.07
Permanent	0.04	0.00	0.18	0.03	0.04
Full-time	0.54	0.47	0.81	0.51	0.57
On Vacation	0.01	0.01	0.03	0.02	0.01
Rural	0.29	0.35	0.04	0.46	0.08
N	117,430	101,901	15,529	75,665	41,765

Notes: This table presents the mean of the main outcomes and covariates for different samples. Column (1) considers all workers. Column (2) considers only workers with informal jobs. Jobs are informal if the production unit is not registered for tax purposes or the worker is not covered by social security. Column (3) considers only workers with formal jobs. Column (4) considers only workers in high-risk industries such as agriculture, fishing, mining, manufacturing, transportation, and utilities. Column (5) considers individuals working in all the other industries not cataloged as high risk. Hours are reported for the reference week which is the last week previous to the interview week. All weather variables are daily averages, except precipitation which is measured at the monthly level. Income is proxied by household expenditure. Education and age are reported in years. All columns use ERA5 data and exclude the jungle.

Table C.2: Temperature Variation under different Fixed Effects and Trends.

		Average Temperature (°C)			**Maximum Temperature (°C)**		
		R^2	σ_e	$\|e\| > 1$	R^2	σ_e	$\|e\| > 1$
		(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
1	Constant		6.13	93.4%	0	5.90	89.9%
2	District FE	0.94	1.49	41.1%	0.93	1.56	44.1%
3	District FE, Linear Year	0.94	1.48	41.3%	0.93	1.56	44.2%
4	District FE, Quadratic Year	0.94	1.48	41.1%	0.93	1.56	44.1%
5	District FE, Cubic Year	0.94	1.47	40.8%	0.93	1.55	44.1%
6	Distict and Year FEs	0.94	1.46	40.5%	0.93	1.54	43.6%
7	District, Year and Month FEs	0.97	1.09	33.6%	0.95	1.33	42.9%
8	District, Year and Week FEs	0.97	1.08	33.0%	0.95	1.32	42.2%
9	Distict and Province*Year FEs	0.95	1.39	36.0%	0.94	1.46	38.9%
10	Distict, Province*Year and Province*Month FEs	0.99	0.62	10.0%	0.98	0.77	17.4%
11	Distict, Province*Year and Province*Week FEs	0.99	0.50	6.0%	0.99	0.64	11.2%
12	Distict, Year*Month and Province*Month Fes	0.99	0.64	10.9%	0.98	0.80	19.0%
13	Distict and Region*Year FEs	0.94	1.44	39.3%	0.93	1.51	42.1%
14	Distict, Region*Year and Region*Month FEs	0.98	0.80	17.0%	0.97	0.98	25.9%
15	Distict, Region*Year and Region*Week FEs	0.99	0.74	13.7%	0.98	0.92	22.4%
16	Distict, Year*Month and Region*Month FEs	0.98	0.78	16.0%	0.97	0.96	24.8%
17	Distict, Year and Region*Month FEs	0.98	0.84	19.7%	0.97	1.02	28.8%
18	Distict, Region*Year and Month FEs	0.97	1.05	32.0%	0.95	1.30	41.1%
19	Distict, Region*Year and Week FEs	0.97	1.04	31.5%	0.95	1.29	41.3%
20	Distict, Year*Month and Month FEs	0.97	1.04	32.1%	0.95	1.28	40.6%
21	Distict, Year*Month and Week FEs	0.97	1.03	31.1%	0.95	1.28	40.1%
22	Individual FE	0.98	0.62	19.5%	0.97	0.75	25.4%
23	Individual and Year and Month FEs	0.99	0.58	17.8%	0.98	0.72	23.9%
24	Individual and Region*Month and Year*Month FEs	0.99	0.51	15.4%	0.98	0.64	20.5%

Notes: This table presents the temperature variation remaining after controlling for different location and time fixed effects, and time trends. Columns (1) and (4) report the R^2 for the regression of temperature on the corresponding location and time fixed effects, and time trends. Columns (2) and (5) report the standard deviation of the residuals (remaining temperature variation). Columns (3) and (6) report what fraction of the observations have a residual that is larger than 1°C. Columns (1), (2) and (3) use average temperature, while Columns (4), (5) and (6) uses maximum temperature.

Table C.3: Temperature Bins Variation under different Fixed Effects and Trends.

		Maximum Temperature Bins (°C)						
		< 12	12-15	15-18	18-21	21-24	24-27	> 27
1	Constant	1.23	1.83	1.49	1.55	1.51	1.23	1.37
2	District FE	0.36	0.55	0.62	0.83	0.80	0.74	0.30
3	District FE, Linear Year	0.38	0.55	0.63	0.86	0.82	0.75	0.31
4	District FE, Quadratic Year	0.38	0.55	0.63	0.86	0.82	0.75	0.32
5	District FE, Cubic Year	0.40	0.56	0.63	0.85	0.82	0.75	0.33
6	Distict and Year FEs	0.42	0.56	0.64	0.87	0.84	0.76	0.34
7	District, Year and Month FEs	0.43	0.58	0.66	0.94	0.89	0.83	0.42
8	District, Year and Week FEs	0.44	0.59	0.66	0.94	0.90	0.84	0.43
9	Distict and Province*Year FEs	0.32	0.49	0.57	0.79	0.76	0.71	0.29
10	Distict, Province*Year and Province*Month FEs	0.29	0.43	0.49	0.51	0.51	0.40	0.17
11	Distict, Province*Year and Province*Week FEs	0.21	0.31	0.37	0.41	0.43	0.34	0.14
12	Distict, Year*Month and Province*Month Fes	0.40	0.52	0.56	0.57	0.60	0.45	0.23
13	Distict and Region*Year FEs	0.38	0.56	0.64	0.85	0.81	0.75	0.31
14	Distict, Region*Year and Region*Month FEs	0.37	0.55	0.61	0.64	0.66	0.52	0.24
15	Distict, Region*Year and Region*Week FEs	0.36	0.53	0.59	0.62	0.64	0.51	0.23
16	Distict, Year*Month and Region*Month FEs	0.43	0.57	0.62	0.67	0.72	0.56	0.28
17	Distict, Year and Region*Month FEs	0.41	0.55	0.61	0.66	0.69	0.54	0.26
18	Distict, Region*Year and Month FEs	0.41	0.58	0.65	0.93	0.88	0.83	0.41
19	Distict, Region*Year and Week FEs	0.42	0.59	0.66	0.93	0.89	0.84	0.42
20	Distict, Year*Month and Month FEs	0.46	0.59	0.66	0.94	0.91	0.84	0.43
21	Distict, Year*Month and Week FEs	0.46	0.60	0.66	0.94	0.91	0.84	0.43
22	Individual FE	0.27	0.41	0.41	0.38	0.37	0.27	0.12
23	Individual and Year and Month FEs	0.32	0.43	0.44	0.43	0.41	0.31	0.15
24	Individual and Region*Month and Year*Month FEs	0.35	0.45	0.45	0.43	0.44	0.31	0.17

Notes: This table presents residual variation available after removing district fixed effects and other controls for each bin. The number of days in each bin is regressed on different location and time fixed effects. The absolute value of the residual is then averaged over all observations.

Table C.4: Working Hours Variation.

	All **(1)**	***Informal*** **(2)**	***Formal*** **(3)**	***Outdoor*** **(4)**	***Indoor*** **(5)**
Between	18.49	18.55	18.60	17.32	21.58
Within	11.47	11.09	9.95	10.27	11.24

Notes: This table presents the between and within decomposition of the standard deviation for the main outcome variable in our regressions (i.e. total weekly working hours).

Table C.5: Effect of Temperature on Work Time.

	All (1)	**Informal** (2)	**Formal** (3)	**Outdoor** (4)	**Indoor** (5)
Temperature (Â°C)					
< 12	0.02	-0.07	0.77*	0.00	-0.07
	(0.10)	(0.11)	(0.45)	(0.12)	(0.19)
12-15	0.00	-0.08	0.46*	-0.14	0.29*
	(0.08)	(0.09)	(0.25)	(0.10)	(0.15)
15-18	0.08	0.02	0.35**	0.01	0.26***
	(0.06)	(0.07)	(0.17)	(0.08)	(0.09)
21-24	-0.03	-0.08	0.15	-0.05	0.02
	(0.04)	(0.05)	(0.10)	(0.06)	(0.07)
24-27	-0.06	-0.12	0.16	-0.15*	0.05
	(0.07)	(0.08)	(0.14)	(0.08)	(0.09)
> 27	-0.26**	-0.42***	0.38	-0.30**	-0.20
	(0.11)	(0.12)	(0.27)	(0.13)	(0.16)
N	117,428	101,899	15,457	75,663	41,723

Notes: This table presents estimated coefficients and standard errors from regressing total working hours on temperature bins. Each column represents a separate regression. Column (1) considers all workers. Column (2) considers workers with informal jobs (i.e. jobs where the production unit is not registered for tax purposes or the worker is not covered by social security). Column (3) considers workers with formal jobs. Column (4) considers only workers in high-risk industries such as agriculture, fishing, mining, manufacturing, transportation, and utilities. Column (5) considers individuals working in all the other industries not cataloged as high risk. All regressions control for precipitation, humidity, daylight hours, sociodemographics, and location and time fixed-effects. All columns use maximum temperature, ERA5 data, and exclude the jungle. Estimated standard errors, reported in parentheses, are clustered at the region-month level. Significance at the one, five, and ten percent levels is indicated by ***, **, and *, respectively.

Table C.6: Effect of Temperature on Work Time: Informal vs. Outdoor.

	Bins (1)	**Bins*Outdoor (2)**	**Bins*Informal (5)**
Temperature (°C)			
< 12	-0.09	-0.01	0.61***
	(0.12)	(0.09)	(0.15)
12-15	0.06	-0.19***	0.30**
	(0.09)	(0.07)	(0.12)
15-18	0.16*	-0.22**	0.20
	(0.09)	(0.08)	(0.15)
21-24	-0.06	-0.03	0.15
	(0.05)	(0.07)	(0.10)
24-27	-0.02	-0.12*	0.03
	(0.07)	(0.07)	(0.09)
> 27	-0.44***	0.15*	0.45***
	(0.11)	(0.08)	(0.10)
N		117,428	

Notes: This table presents estimated coefficients and standard errors from regressing total working hours on temperature bins, temperature bins interacted with an indicator variable for outdoor jobs, and temperature bins interacted with an indicator variable for informal jobs. Column (1) shows the estimates for the temperature bins. Column (2) shows the estimates for the interaction between temperature bins and the indicator variable for outdoor jobs. Column (3) shows the estimates for the interaction between temperature bins and the indicator variable for informal jobs. A job is informal if the production unit is not registered for tax purposes or the worker is not covered by social security. Individuals with outdoor jobs work in high-risk industries such as agriculture, fishing, mining, manufacturing, transportation, and utilities. The regression controls for precipitation, humidity, daylight hours, sociodemographics, and location and time fixed-effects. All columns use maximum temperature, ERA5 data, and exclude the jungle. Estimated standard errors, reported in parentheses, are clustered at the region-month level. Significance at the one, five, and ten percent levels is indicated by ***, **, and *, respectively.

Table C.7: Effect of Current and Lagged Temperature on Working Hours.

	Weeks		
	t	**t-1**	**t + t-1**
	(1)	**(2)**	**(3)**
Temperature (°C)			
< 12	-0.06	0.19*	0.13
	(0.12)	(0.11)	(0.12)
12-15	-0.02	0.04	0.01
	(0.09)	(0.09)	(0.10)
15-18	0.10	-0.02	0.07
	(0.07)	(0.07)	(0.07)
21-24	0.04	-0.09	-0.05
	(0.06)	(0.06)	(0.05)
24-27	-0.05	-0.01	-0.06
	(0.11)	(0.11)	(0.07)
> 27	-0.43***	0.19	-0.24*
	(0.15)	(0.14)	(0.12)
N		117,428	

Notes: This table presents estimated coefficients and standard errors from regressing working hours on temperature bins. Column (1) shows estimates for the contemporaneous (t) temperature bins. Column 2 shows estimates for the one-week lagged (t-1) temperature bins. Column (3) shows the summation of the effects across weeks t and t-1. All regressions control for precipitation, humidity, daylight hours, sociodemographics, and location and time fixed-effects. All columns use daily maximum temperature, ERA5 data, and exclude the jungle. Estimated standard errors, reported in parentheses, are clustered at the region-month level. Significance at the one, five, and ten percent levels is indicated by ***, **, and *, respectively.

Table C.8: Effect of Temperature on Working Hours.

	All
	(1)
Temperature (°C)	
< 9	-0.36**
	(0.17)
9-12	0.01
	(0.11)
12-15	-0.02
	(0.08)
15-18	0.07
	(0.06)
21-24	-0.01
	(0.04)
24-27	-0.01
	(0.06)
27-30	-0.24**
	(0.11)
> 30	-0.05
	(0.14)
N	117,428

Notes: This table presents estimated coefficients and standard errors from regressing total working hours on temperature bins for all workers. Regression controls for precipitation, humidity, daylight hours, sociodemographics, and location and time fixed-effects. Estimation uses daily maximum temperature, ERA5 data, and excludes the jungle. Estimated standard errors, reported in parentheses, are clustered at the region-month level. Significance at the one, five, and ten percent levels is indicated by ***, **, and *, respectively.

Table C.9: Effect of Temperature on Working Hours - Narrow Bins.

	All
Temperature (°C)	
< 12	-0.05
	(0.11)
12-14	0.02
	(0.09)
14-16	-0.09
	(0.08)
16-18	0.11
	(0.07)
20-22	-0.00
	(0.05)
22-24	0.06
	(0.06)
24-26	0.03
	(0.07)
26-28	-0.02
	(0.11)
28-30	-0.14
	(0.12)
> 30	0.09
	(0.14)
N	117,428

Notes: This table presents estimated coefficients and standard errors from regressing total working hours for all workers on narrower temperature bins. Regression controls for precipitation, humidity, daylight hours, sociodemographics, and location and time fixed-effects. Estimation uses daily maximum temperature, ERA5 data, and excludes the jungle. Estimated standard errors, reported in parentheses, are clustered at the region-month level. Significance at the one, five, and ten percent levels is indicated by ***, **, and *, respectively.

Table C.10: Alternative Measures of Temperature.

	All
Panel A	
Sin(temperature)	-0.28***
	(0.09)
Panel B	
Temperature	0.01
	(0.06)
Panel C	
Temperature	0.03
	(0.19)
Temperature^2	-0.00
	(0.00)
Panel D	
DD (7 days)	-0.00
	(0.01)
HDD (7 days)	0.01
	(0.02)
N	117,428

Notes: This table presents estimated coefficients and standard errors from regressing working hours on different measures of temperature. All regressions control for precipitation, humidity, daylight hours, and sociodemographics. All columns use daily maximum temperature, ERA5 data, and exclude the jungle. Estimated standard errors, reported in parentheses, are clustered at the region-month level. Significance at the one, five, and ten percent levels is indicated by ***, **, and *, respectively.