Essays on Sleep, Labor, and Child Development

Jingyan (Jean) Guo

University of California, Riverside

Final Defense of Dissertation

May 29th, 2024

Chapter 1

Sleep and Fatal Vehicle Crashes: Evidence from Sunset Time in the United States

Jingyan (Jean) Guo

University of California, Riverside

Final Defense of Dissertation

May 29th, 2024

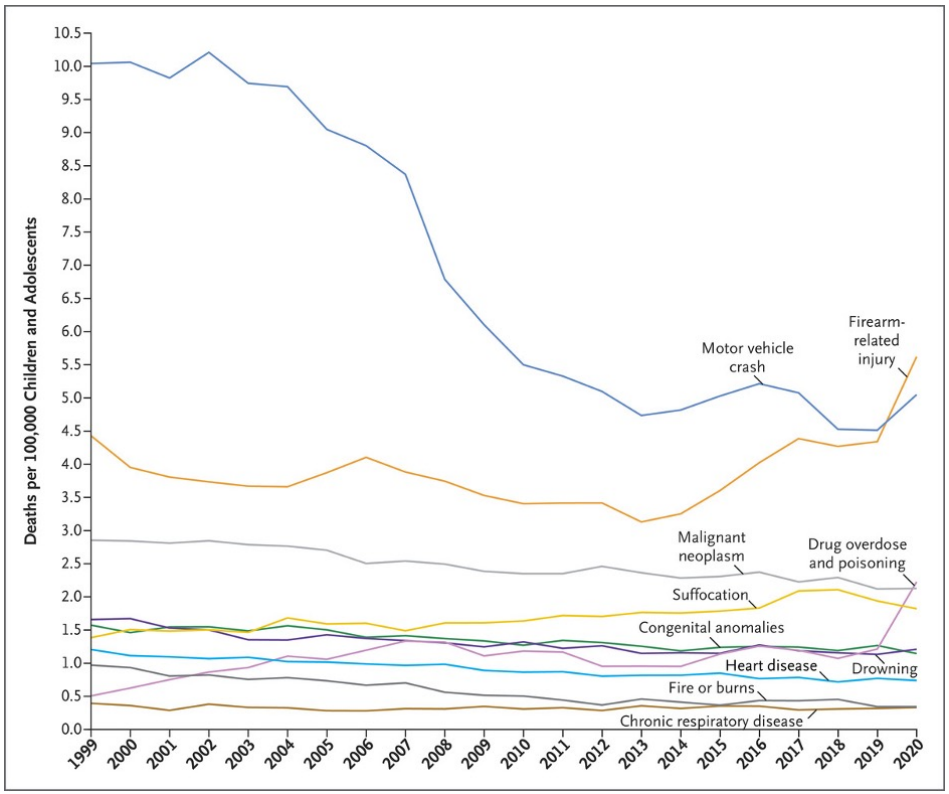
Table of Contents

Motivation

- Sleep is crucial for both human health and productivity
  - We spend one-third of our lives sleeping
  - Its importance remains largely understudied in the fields of health and labor economics
- Consequences of insufficient sleep on health and labor
  - Associates with negative health outcomes such as obesity, diabetes, cardiovascular diseases, and breast cancer (Giuntella & Mazzonna, 2019)
  - Psychological well-being (Bessone et al., 2021); lowers cognitive function (Breus, 2022); sleep and wage (Gibson and Shrader, 2018)
  - Impairs attention and leads to fatigue-related accidents and injuries (Dinges, 1995; Lockley et al., 2007; Barnes and Wagner, 2009)
  - Daylight Saving Time (DST) on fatal crashes (Smith, 2016)

Motivation

- Motor vehicle crashes was the leading cause of death among children and adolescents in the United States, from 1999 through 2019 (Goldstick, 2022)



Source: JE Goldstick et al. N Engl J Med 2022;386:1955-1956.

Research Question and Identification Strategy

- Research question
  - What are the short run (seasonal) and long run (geographical) impacts of sleep duration on fatal vehicle crashes in the United States?
- Identification strategy
  - Instrumental Variable (IV) approaches
    - ⋆ Daily sunset variation across year in one location for short-run analysis
  - Spatial Regression Discontinuity Design (RDD)
    - ⋆ Geographic variation in average sunset time across the U.S. for long-run estimates
    - ⋆ Exploits the sharp discontinuity in sunset time across time zone borders

▸ Daily Sunset Variation Graph
▸ Average Sunset Variation Graph
▸ Sunset and Distance Graph

Preview of Results

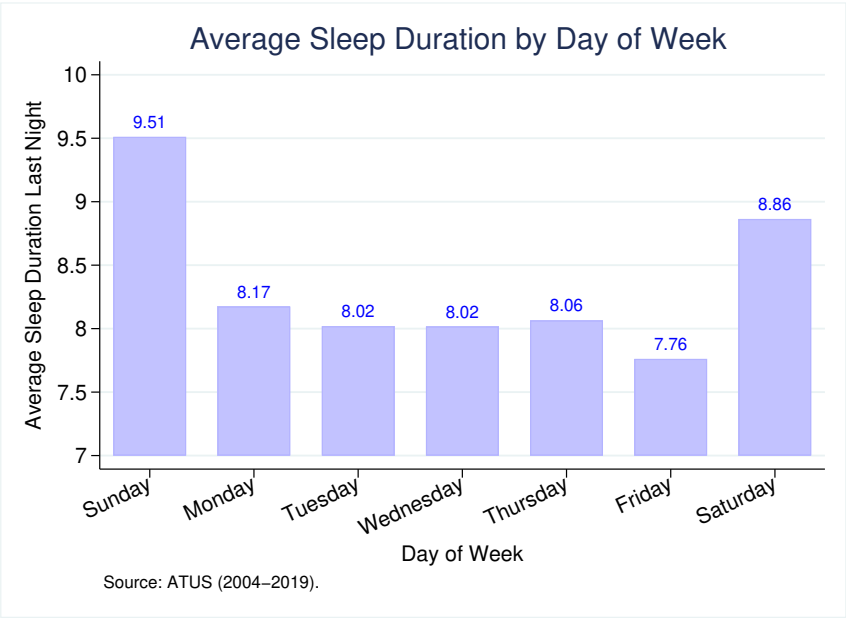
- IV short-run (seasonal) results
  - First stage: one-hour delay in sunset results in a decrease of approximately 12 minutes in weekly sleep duration
  - One-hour increase in monthly sleep leads to 2.4% reduction in fatalities
- RDD long-run (geographical) results
  - First stage: one-hour delay in sunset results in a decrease of approximately 10 minutes in average daily sleep duration
  - The results of sleep on fatalities are unstable over time

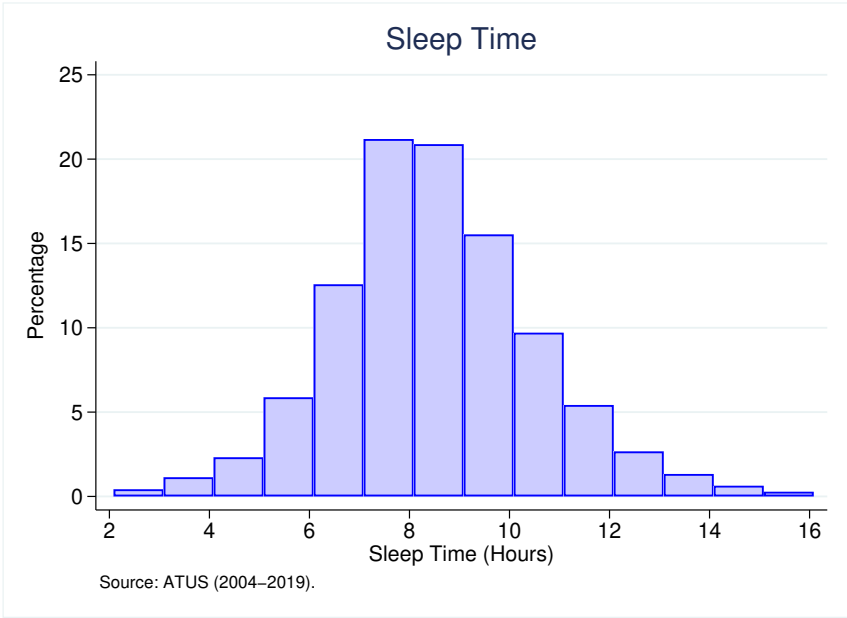
Table of Contents

Sources of Data

- **Data to measure individual sleep duration**
  - American Time Use Survey (ATUS)
- **Data to measure traffic crashes**
  - Fatality Analysis Reporting System (FARS)
- **Data to measure sunset and distance**
  - Average sunset times from package "suncalc" (R Studio)
  - County map from National Weather Service
  - Time zone borders map from U.S. Department of Transportation
  - Data are merged on the state-county levels

Sleep Duration

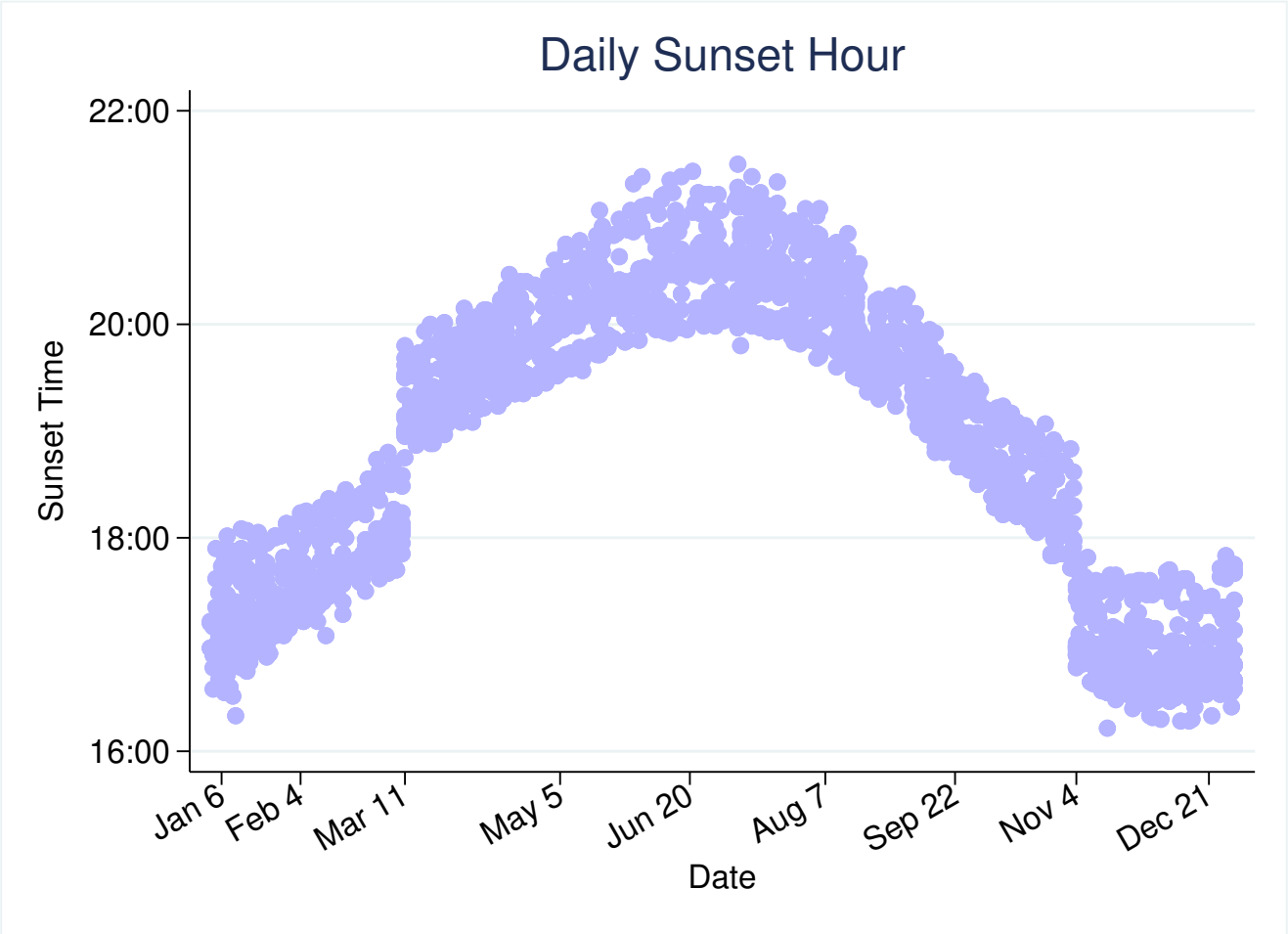




▸ Summary Statistics

Table of Contents

IV Identification Strategy - Short Run



This graph depicts the daily sunset hours for counties sampled by ATUS in the continental United States in 2012. The y-axis shows the sunset hour in 24 hour time. The setup of this graph is similar to Gibson and Shrader (2018).

◂ Identification Strategy

Estimation Equations for the IV Short Run Analysis

I estimate the following short-run first stage,

$$Sleep_{ijt} = \alpha_1 Sunset_{jt} + X'_{it}\delta_1 + \gamma_{1,j} + \eta_{1,ijt} \tag{1}$$

short-run second stage,

$$Crash_{jt} = \alpha_2 \hat{Sleep}_{ijt} + X'_{it}\delta_2 + \gamma_{2,j} + \eta_{2,ijt} \tag{2}$$

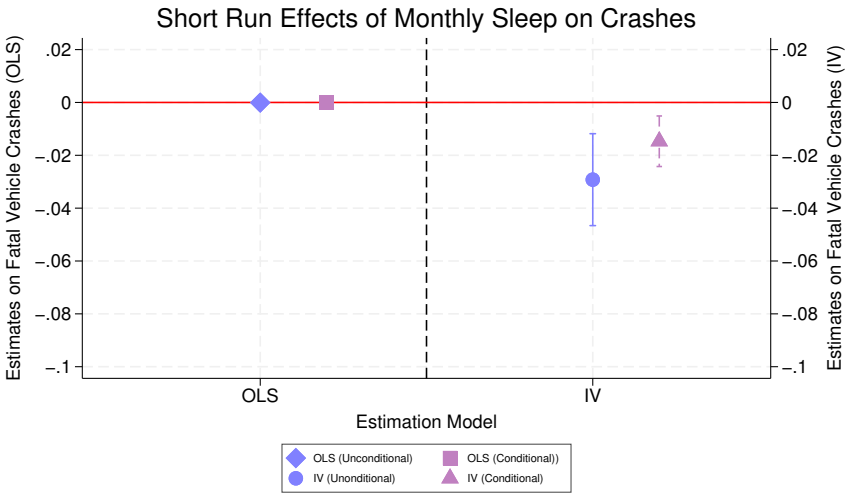
and reduced form,

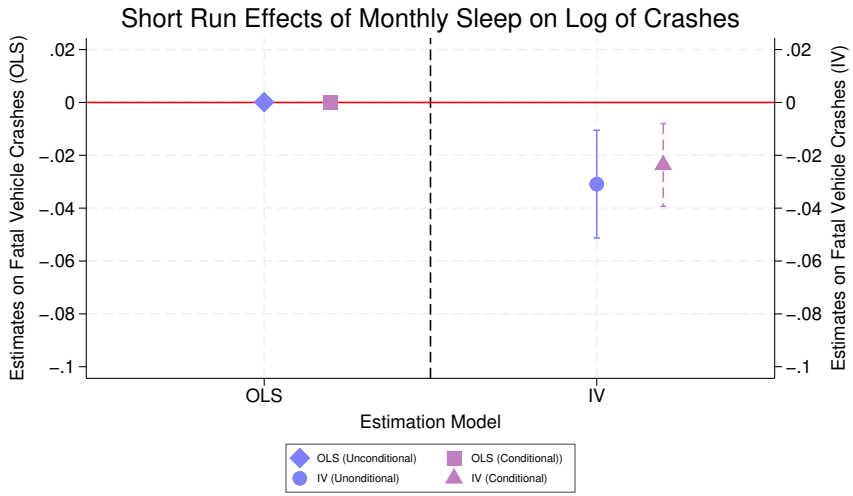
$$Crash_{jt} = \alpha_3 Sunset_{jt} + X'_{it}\delta_3 + \gamma_{3,j} + \eta_{3,ijt} \tag{3}$$

- $Sleep_{ijt}$: monthly sleep duration for individual i in county j for date t
- $Sunset_{jt}$: the sunset time on that date in that county
- $Crash_{jt}$: the fatal crashes per 100,000 population for the county j at county-year-month level
- $\gamma_{1,j}$: county fixed effects
- X_{it}: a vector of demographic, geographic, and interview controls

Table of Contents

Short Run Effects of Monthly Sleep on Crashes (IV)





◂ Graph (Version 1)

Notes: This graph shows estimates of the short run effects of monthly sleep on crashes using OLS and IV. The error bars are at 95% confidence intervals for the mean. Sleep denotes monthly average sleep hours. The dependent variable of crashes refers to fatal crashes per 100,000 population at county-year-month level. Controls include socio-demographics (age, race, sex, education, marital status, nativity status, and number of children), geographic characteristics (latitude, longitude, and indicator for large counties), and interview characteristics (indicators for holiday and weekend).

Short Run Effects of Sunset and Sleep (Log of Crashes)

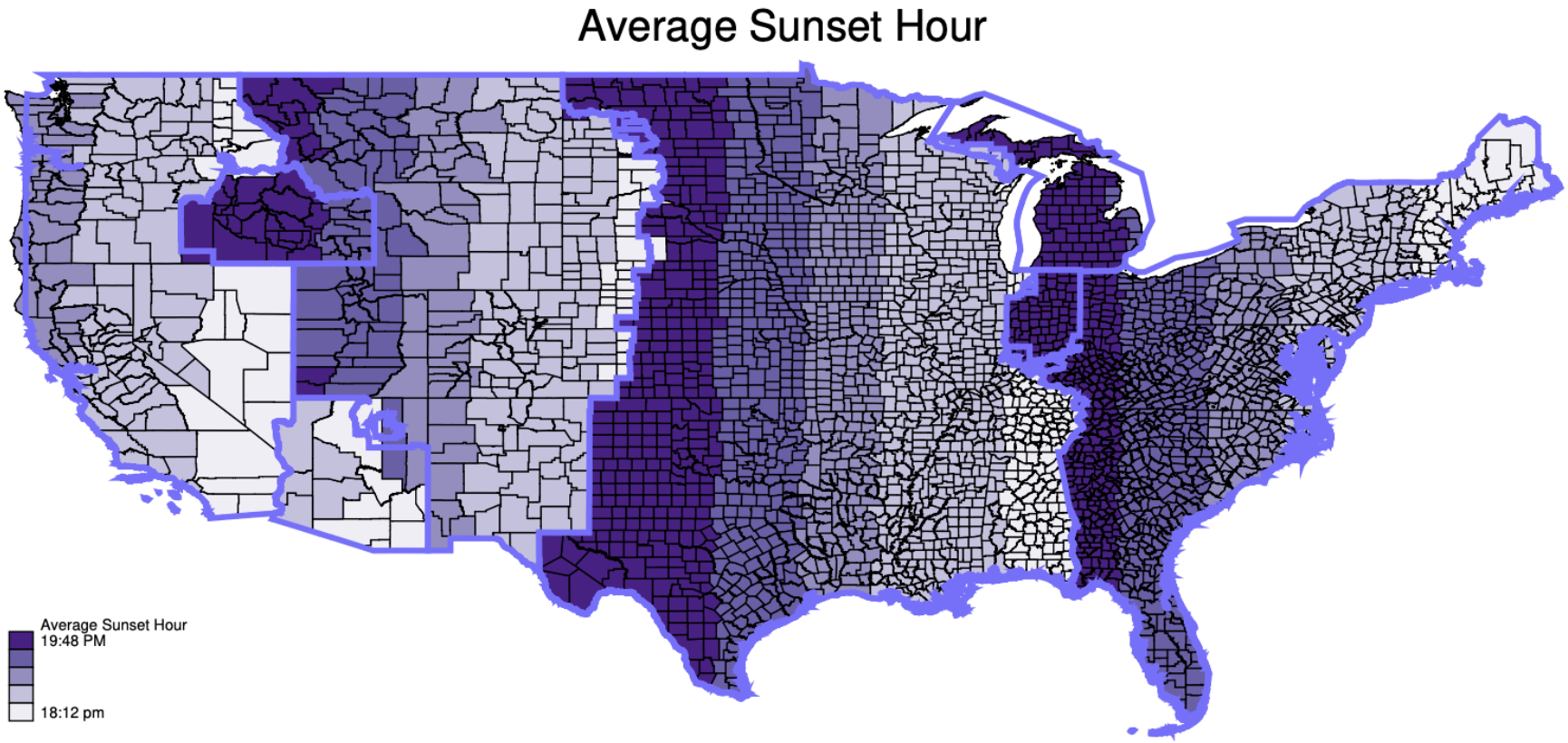
	OLS		IV(First-Stage)		IV(Second-Stage)		Reduced-Form	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Crashes	Crashes	Sleep	Sleep	Crashes	Crashes	Crashes	Crashes
	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se
Average Monthly Sleep	-0.000	-0.000			−0.044∗∗∗	−0.024∗∗∗		
	(0.00)	(0.00)			(0.01)	(0.01)		
Sunset Hour			−0.815∗∗∗	−0.778∗∗∗			0.036∗∗∗	0.019***
			(0.24)	(0.22)			(0.01)	(0.00)
Mean	-0.41	-0.41	261.28	261.28	-0.41	-0.41	-0.41	-0.41
Controls	No	Yes	No	Yes	No	Yes	No	Yes
County FEs	No	Yes	No	Yes	No	Yes	No	Yes
Observations	36296	36296	36296	36296	36296	36296	36296	36296
F test			11.99	12.55				

Notes: Sleep and sunset time are measured in hours at county level. Sleep denotes monthly average sleep hours. The dependent variable of crashes refers to fatal crashes per 100,000 population at county-year-month level. Controls include socio-demographics (age, race, sex, education, marital status, nativity status, and number of children), geographic characteristics (latitude, longitude, and indicator for large counties), and interview characteristics (indicators for holiday and weekend). The standard errors are robust to heteroscedasticity and clustered at county level (reported in parentheses). F test on the excluded instrument. Significance levels: * 0.10, ** 0.05, *** 0.01.

◂ Seasonality ◂ Road Type ◂ Light Condition

Table of Contents

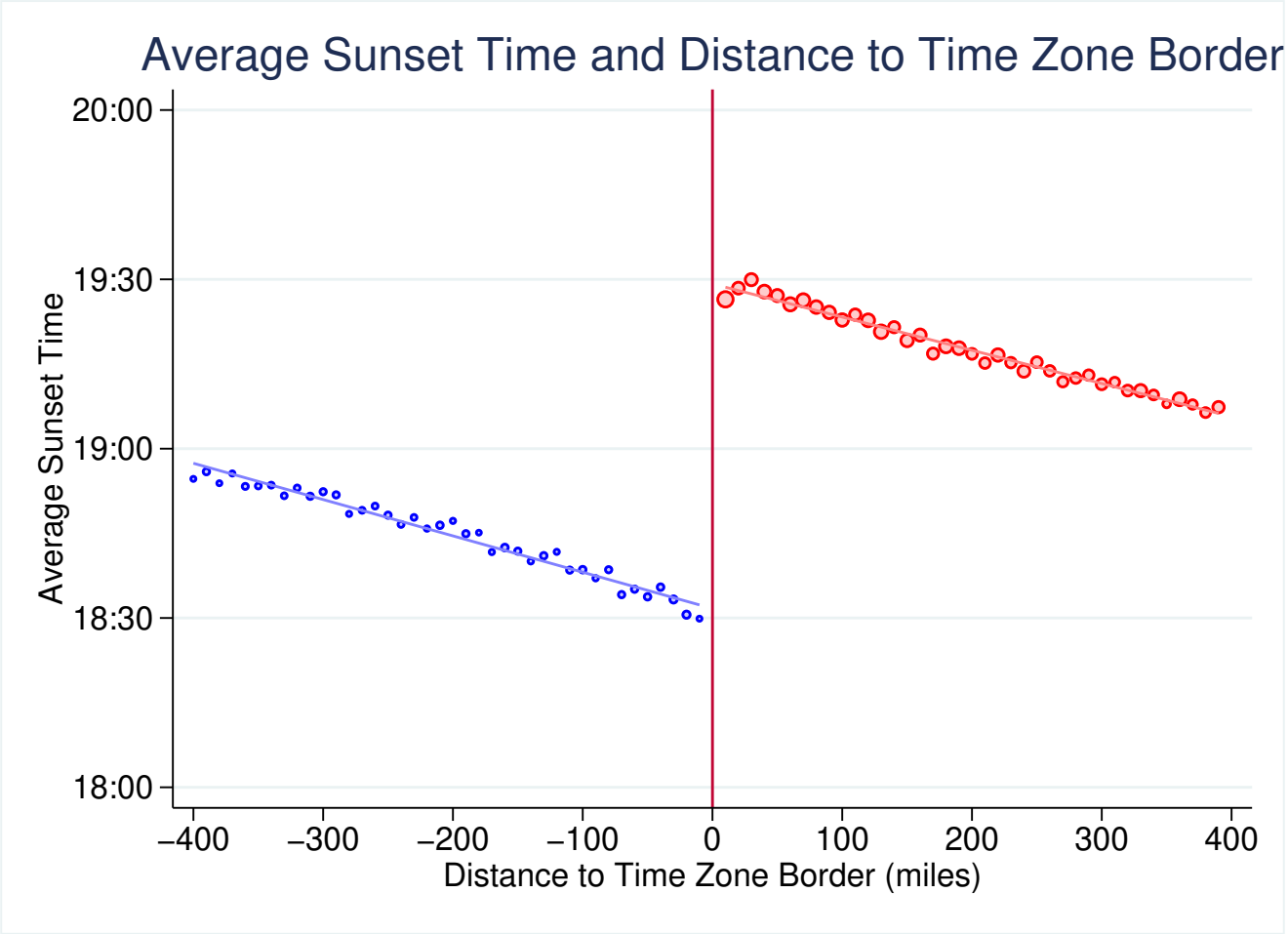
RDD Identification Strategy - Long Run



This graph shows the average sunset hours for all counties in the continental United States in 2012. Darker color implies later sunset. The time zone border lines are in blue. The setup of this graph is similar to Gibson and Shrader (2018) and Giuntella and Mazzonna (2019).

◂ Identification Strategy

RDD Identification Strategy - Long Run



This graph shows the discontinuity in sunset time over distance to time zone borders. The distance are calculated using QGIS. I used sunset time package "suncalc" from R studio to calculate the average sunset time. The scatterplot is weighted by the number of observations in distance group. The distance group is calculated using the cut command in Stata.

◂ Identification Strategy

Estimation Equations for the RDD Long Run Analysis

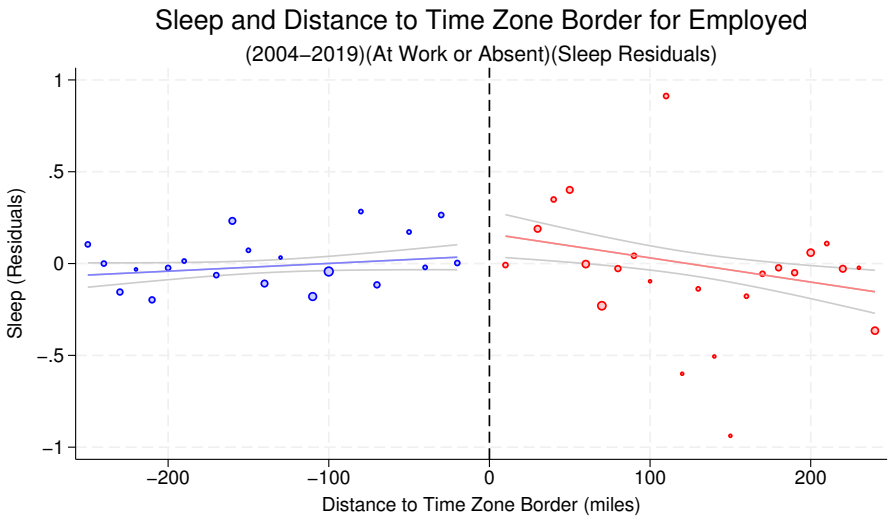
I estimate the following first stage,

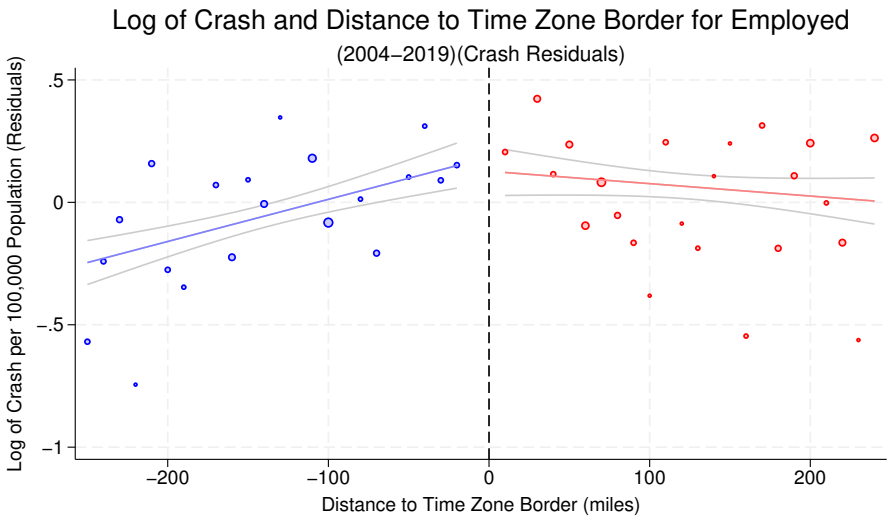
$$Sleep_{ijt}/Crash_j = \beta_0 + \beta_1 LS_j + \beta_2 f(D_j) + \beta_3 f(D_j) * LS_j + X'_{ijt}\beta_4 + u_{ijt} \quad (4)$$

- $Sleep_{ijt}$: daily sleep duration for individual i in county j for date t
- $Crash_j$: average fatal crashes per 100,000 population at county-month level
- LS_j: indicator for the county located on the late sunset side of a time zone border
- D_j: the distance to the time zone border or the "running variable"
- X_j: a vector of demographic, geographic, and interview controls

Table of Contents

Sleep and Crash (2004-2019)

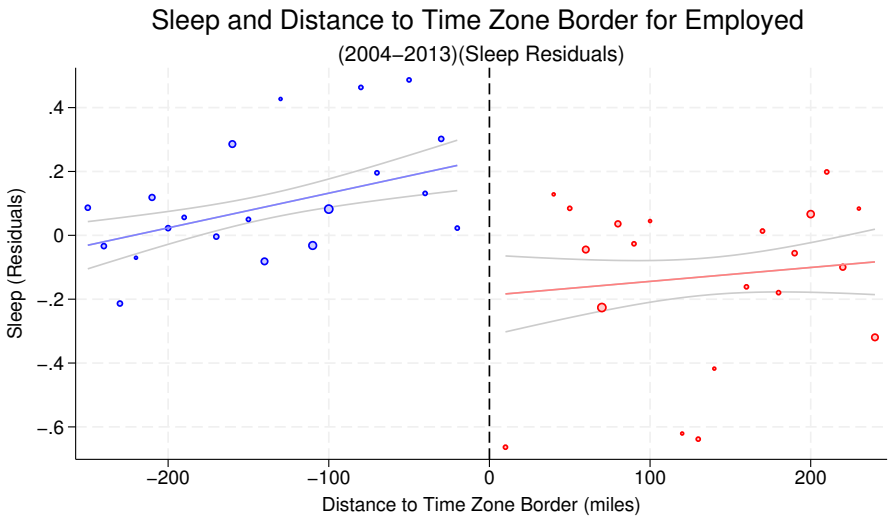


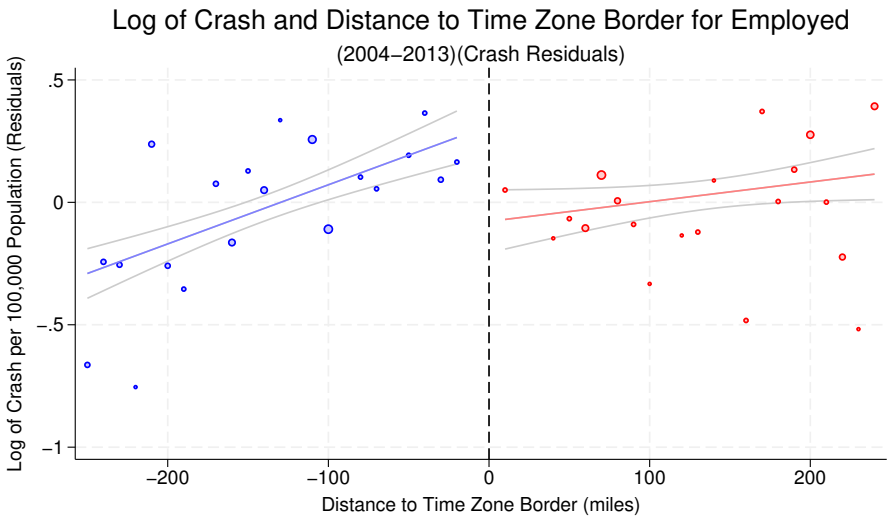


Notes: The left figure shows the discontinuity in sleep and distance to time zone border for employed and unemployed individuals. Data are from ATUS and FARS (2004-2019). Each point represents the the mean residuals (10 miles average) of sleep on a set of geographical controls (a linear control for latitude and dummy for large counties). The right figure shows the discontinuity in crash and distance to time zone border. Each point represents the mean residuals (10 miles average) of the crash per 100 millions VMT on a set of geographical controls (a linear control for latitude and dummy for large counties). The scatterplot is weighted by the number of observations in distance group. The distance group is calculated using the cut command in Stata.

▸ Table of Sleep and Distance ▸ Table of Crash and Distance

Sleep and Crash (2004-2013)

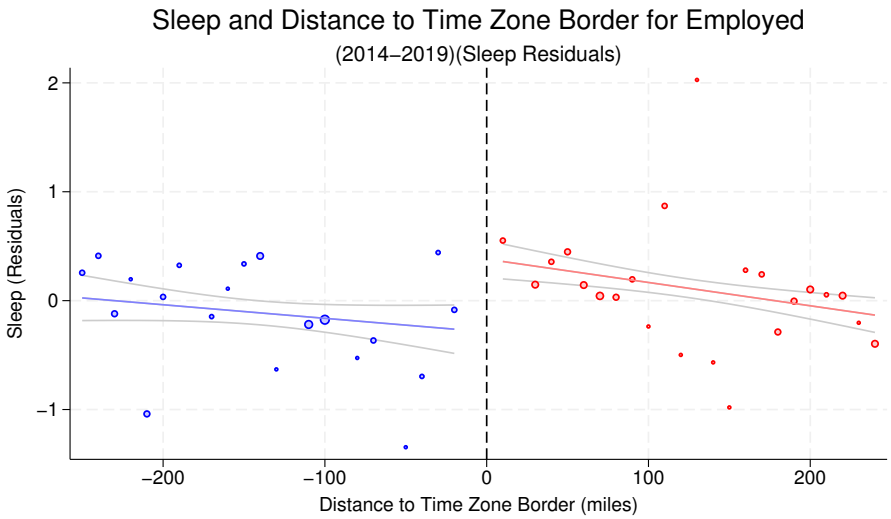


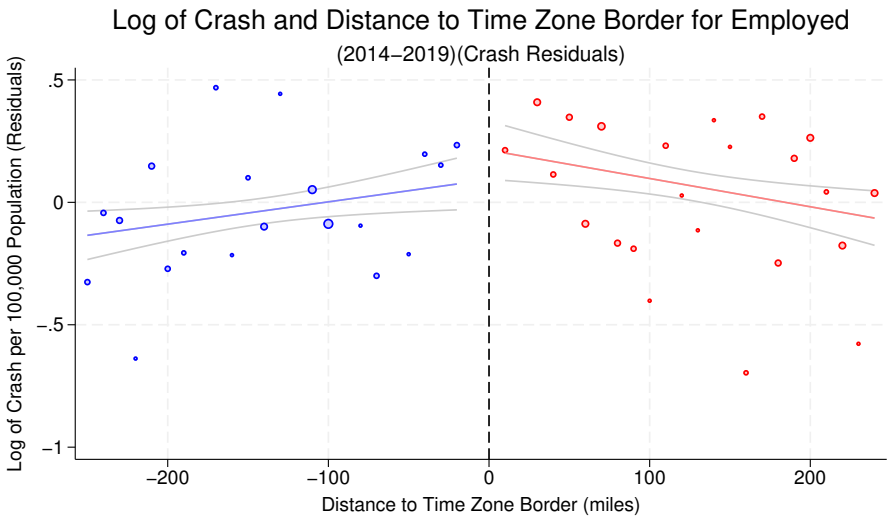


Notes: The left figure shows the discontinuity in sleep and distance to time zone border for employed and unemployed individuals. Data are from ATUS and FARS (2004-2013). Each point represents the the mean residuals (10 miles average) of sleep on a set of geographical controls (a linear control for latitude and dummy for large counties). The right figure shows the discontinuity in crash and distance to time zone border. Each point represents the mean residuals (10 miles average) of the crash per 100 millions VMT on a set of geographical controls (a linear control for latitude and dummy for large counties). The scatterplot is weighted by the number of observations in distance group. The distance group is calculated using the cut command in Stata.

▸ Table of Sleep and Distance ▸ Table of Crash and Distance

Sleep and Crash (2014-2019)





Notes: The left figure shows the discontinuity in sleep and distance to time zone border for employed and unemployed individuals. Data are from ATUS and FARS (2014-2019). Each point represents the the mean residuals (10 miles average) of sleep on a set of geographical controls (a linear control for latitude and dummy for large counties). The right figure shows the discontinuity in crash and distance to time zone border. Each point represents the mean residuals (10 miles average) of the crash per 100 millions VMT on a set of geographical controls (a linear control for latitude and dummy for large counties). The scatterplot is weighted by the number of observations in distance group. The distance group is calculated using the cut command in Stata.

▸ Table of Sleep and Distance ▸ Table of Crash and Distance

Table of Contents

Conclusion

- Summary
  - This paper studies the causal impact of sleep on the fatal vehicle crashes in the United States using IV and RDD method and data from ATUS and FARS
- Findings
  - One-hour augment in monthly sleep could result in a reduction of 2.4% in fatalities in the short run (IV)
  - The results are not stable in the long run using different periods of time (RDD)
- Policy Implications
  - Help to form a better policy solution such as whether to keep DST and end clock changes
  - Help to construct social schedules such as work schedules and school start times in ways that promote sleeping, which is related to health and productivity

Chapter 2

Employment and Sleep Patterns

Jingyan (Jean) Guo

University of California, Riverside

Final Defense of Dissertation

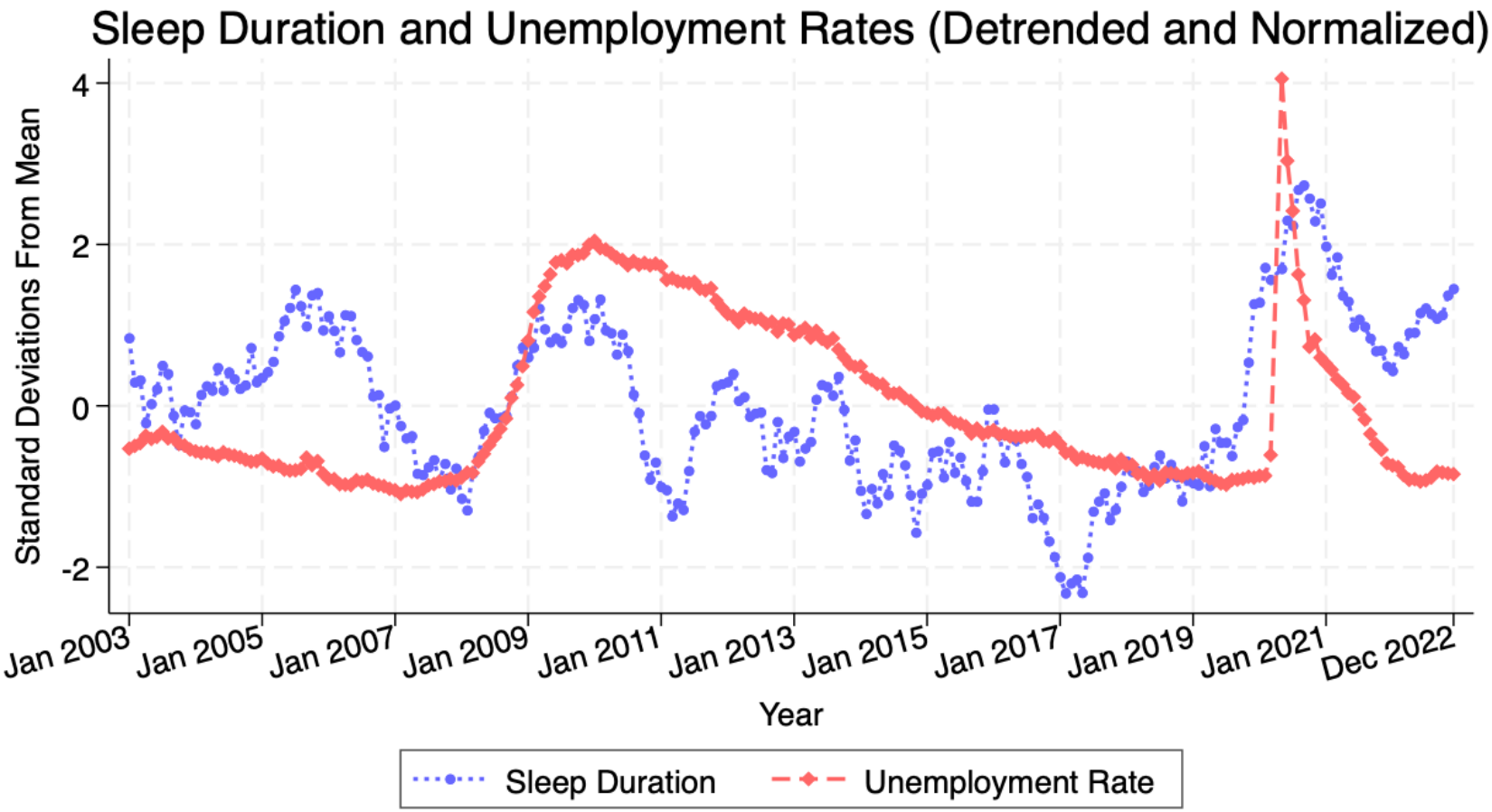
May 29th, 2024

Table of Contents

Motivation

- Sleep is essential for maintaining good health and productivity
  - Sleeping fewer than 7 hours per night can increase the risk of developing negative health conditions (Liu, 2016)
  - High blood pressure, heart disease, stroke, diabetes, obesity, and frequent mental distress
- The business cycle has a significant impact on sleep pattern
  - Sleep duration tends to be countercyclical (Colman and Dave, 2003; Niekamp, 2019)
  - During the Great Recession, people spend more time sleeping (Aguiar et al., 2013)

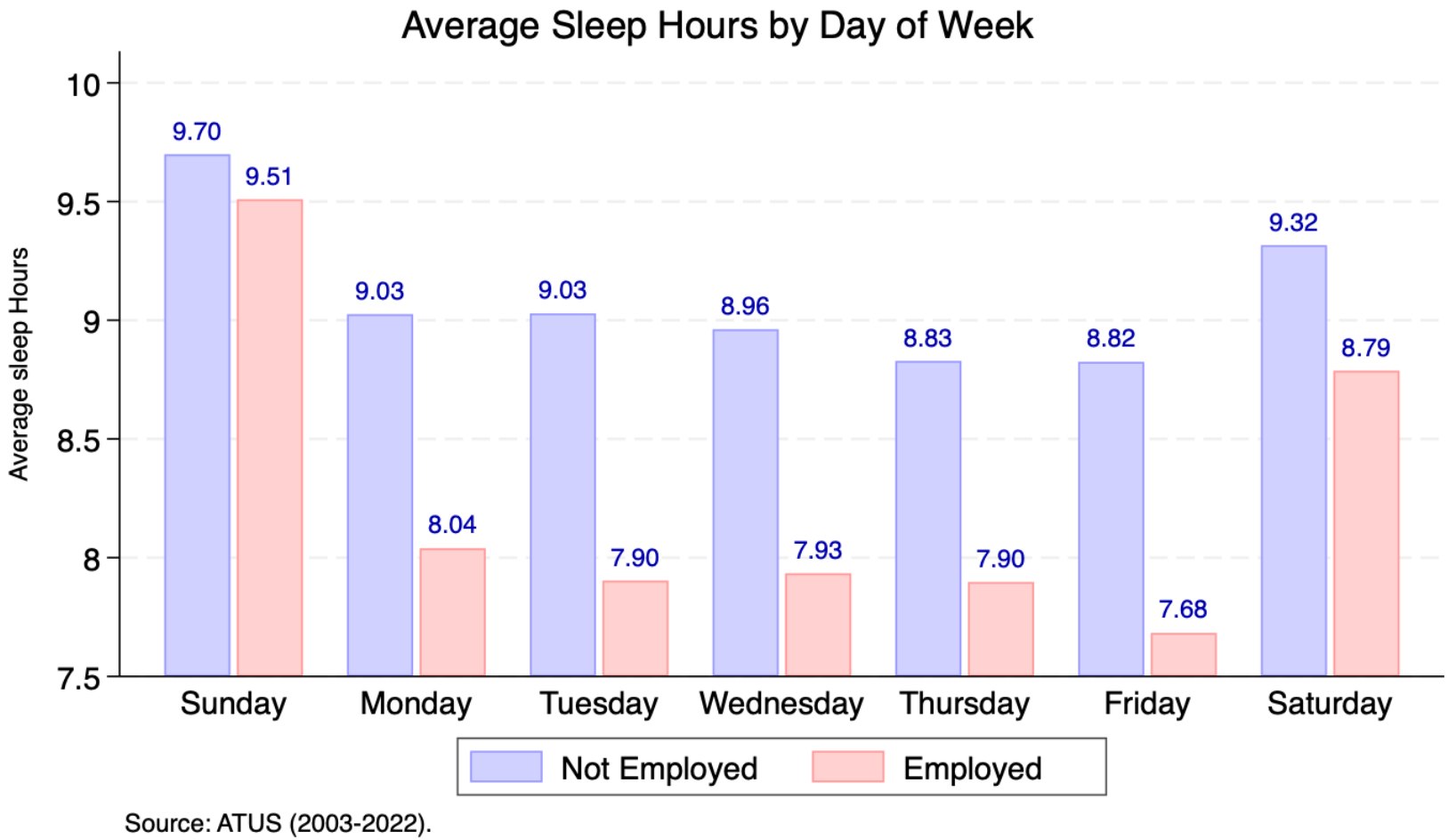
Motivation



Source: ATUS and BLS LAUS (Jan 2003 - Dec 2022).

Unemployment rates are seasonally adjusted and sleep durations are smoothed by applying a moving average. The variables have been detrended by removing a linear trend and normalized by subtracting the mean of the detrended variables and dividing by their standard deviation. The figure shows a positive relationship between unemployment rates and sleep duration.

Motivation



Source: ATUS (2003-2022).

The sample averages only include data from individuals aged 25 to 55 who have reported at least 23 hours of time use. Not employed refers to those who are unemployed and not in the labor force.

Related Literature

- How economic conditions influence health outcomes
  - Mortality rates and economic condition (Ruhm, 2000; Ruhm, 2005; Miller et al., 2009; Stevens et al., 2015)
  - Labor market conditions and health (Charles and DeCicca, 2008; Colman and Dave, 2013)
  - Sleep duration is countercyclical on weekdays but procyclical on weekends (2003 to 2015) (Niekamp, 2019)
- Impact of the COVID-19 pandemic on the labor market
  - Teleworkers spent more time to sleep (Pabilonia and Vernon, 2020)
  - Persistent rise in work from home (WFH) (14.4% to 39.6%) (Bick et al., 2023, Barrero et al., 2021)
- **Contribution**
  - This paper includes analysis from 2003 to 2022
  - How economic conditions affect sleep patterns across different industries by concentration of telework

Table of Contents

Sources of Data

- **Data to measure individual sleep duration**
  - American Time Use Survey (ATUS)
- **Data to measure employment rates**
  - Local Area Unemployment Statistics (LAUS)

▸ Summary Statistics

Table of Contents

Estimation Equation

$$Sleep_{isdmt} = \beta_0 + \beta_1 E_{smt} + X'_{ismt}\beta_2 + \gamma_s + \delta_m + \lambda_d + \theta_t + u_{isdmt} \quad (5)$$

- $Sleep_{isdmt}$: daily sleep duration in minutes for individual i in state s on day of week d in month m in year t
- X_{ismt}: a vector of social demographic and interview controls
- E_{smt} : civilian employment-population ratio for state s in month m of year t
- γ_s: state fixed effects
- δ_m: month fixed effects
- λ_d: day of the week fixed effects

Table of Contents

Effects of Employment Rate on Sleep by Subgroups (2003-2022)

	(1) All	(2) $<$ Bachelor's	(3) $\geq$ Bachelor's	(4) White	(5) BHAI	(6) Female	(7) Male	(8) WM $<$ Bachelor's
All	-1.019^{**}	-1.155^{*}	-0.965	-0.529	-1.777^{*}	-1.183^{*}	-0.822	-1.113
	(0.352)	(0.541)	(0.515)	(0.507)	(0.828)	(0.564)	(0.468)	(0.809)
Observations	120743	72010	48733	80057	33617	65868	54875	20985
R^2	0.11	0.10	0.13	0.11	0.10	0.11	0.12	0.10
Weekday	-2.308^{***}	-2.900^{***}	-1.699^{*}	-1.496^{*}	-4.455^{***}	-2.753^{***}	-1.824^{**}	-2.785^{*}
	(0.441)	(0.625)	(0.675)	(0.639)	(1.136)	(0.699)	(0.595)	(1.249)
Observations	60250	35690	24560	40287	16411	32706	27544	10625
R^2	0.07	0.07	0.05	0.06	0.07	0.07	0.07	0.07
Weekend	0.259	0.512	-0.201	0.461	0.658	0.353	0.128	0.471
	(0.532)	(0.825)	(0.628)	(0.715)	(1.153)	(0.712)	(0.766)	(1.516)
Observations	60493	36320	24173	39770	17206	33162	27331	10360
R^2	0.05	0.05	0.06	0.05	0.05	0.06	0.06	0.05

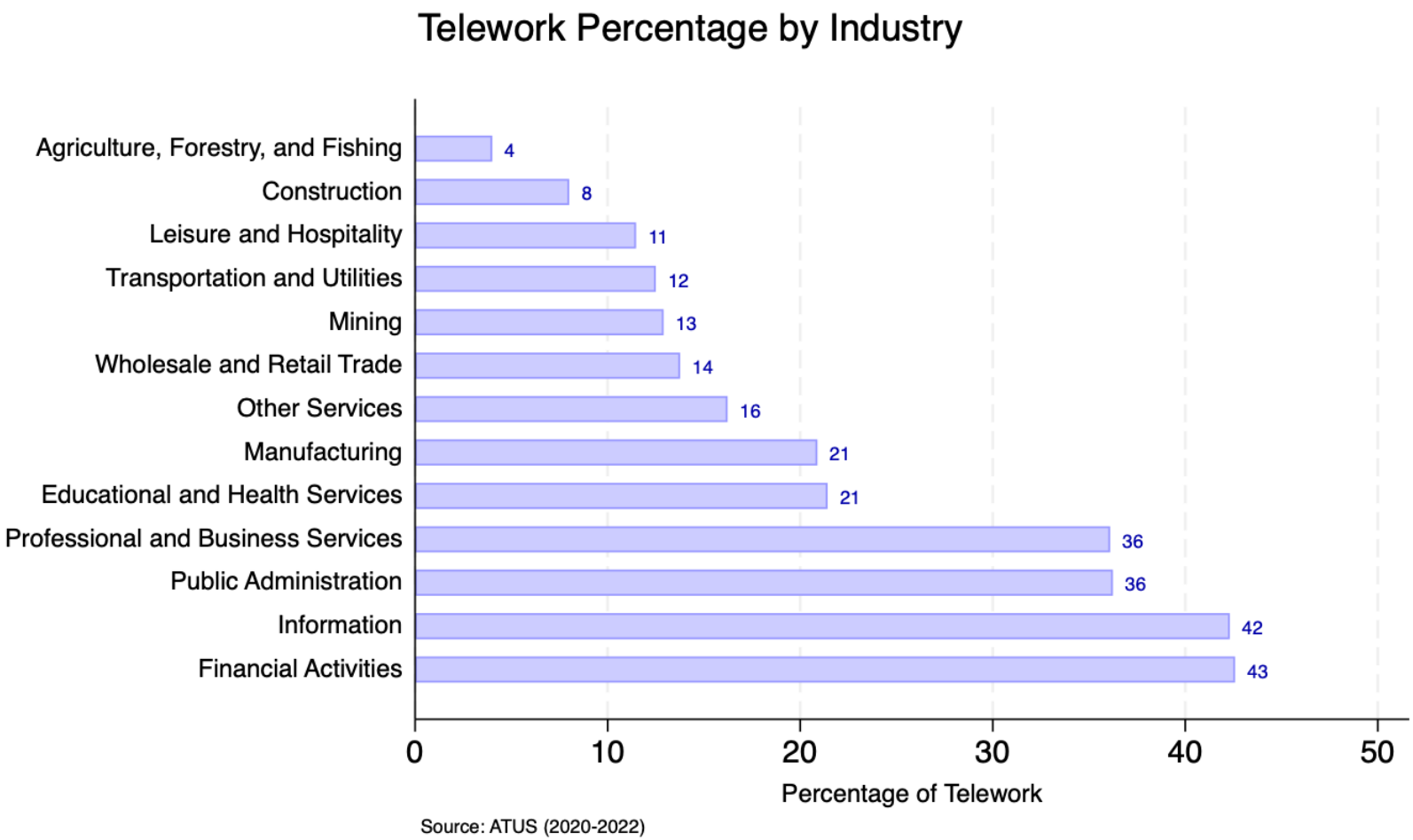
Data are from ATUS and BLS LAUS (2003-2022). The dependent variable is daily sleep in minutes for respondents aged 25-55. Each cell represents the estimates of employment to population rate on sleep. Controls include socio-demographics (age, gender, race, education, race interacted with education, marital status, number of children, indicator for having a child under 3, and industry codes) and interview characteristics (indicators for holiday and incomplete diary). The standard errors are robust to heteroscedasticity and clustered at state level (reported in parentheses). Column 5 refers to (BAHI) Black, Hispanic, or American Indian. Column 8 restricts to white males with education less than a Bachelor's degree.

Effects of Employment Rate on Sleep by Marital Status (2003-2022)

	(1) Single	(2) Married	(3) Single Parent	(4) Married Parent	(5) MF(sleep)	(6) MF(work)
All	-0.791	−1.141∗	−3.081∗∗	-0.958	-1.795	2.102
	(0.643)	(0.504)	(0.974)	(0.524)	(0.972)	(1.109)
Observations	48062	72681	18122	64968	21066	21066
R^2	0.10	0.13	0.11	0.10	0.11	0.31
Weekday	−3.234∗∗∗	−1.692∗	−5.507∗∗∗	−2.866∗∗∗	−2.814∗	2.080
	(0.909)	(0.681)	(1.261)	(0.759)	(1.395)	(1.854)
Observations	23969	36281	8959	32453	10325	10325
R^2	0.07	0.06	0.08	0.07	0.06	0.34
Weekend	1.447	-0.438	-0.765	0.848	-0.610	2.138
	(0.898)	(0.654)	(1.237)	(0.682)	(1.137)	(1.737)
Observations	24093	36400	9163	32515	10741	10741
R^2	0.05	0.06	0.05	0.05	0.06	0.09

Data are from ATUS and BLS LAUS (2003-2022). The dependent variable is daily sleep in minutes for respondents aged 25-55. Each cell represents the estimates of employment to population rate on sleep. Column 5 refers to married females without college degree for sleep in minutes. Column 6 refers married females without college degree for work in minutes.

Telework Concentration by Industry



Data are from ATUS (2020-2022). Starting July 2020, a new survey question has been introduced asking respondents, "At any time in the last 4 weeks, did you telework or work at home for pay?" This graph depicts the percentage of telework by industry.

Effects of Employment Rate on Sleep by Telework

	(1) All 2003-2022	(2) > Mean Telework 2003-2022	(3) ≤ Mean Telework 2003-2022	(4) > Mean Telework 2003-2019	(5) ≤ Mean Telework 2003-2019
All	−1.019∗∗	−0.964∗	−1.201∗	-0.754	−1.317∗
	(0.352)	(0.465)	(0.533)	(0.702)	(0.630)
Observations	120743	54957	48590	49436	44410
R^2	0.11	0.13	0.12	0.12	0.12
Weekday	−2.308∗∗∗	−1.502∗	−2.583∗∗∗	-1.461	−2.802∗∗∗
	(0.441)	(0.590)	(0.626)	(0.765)	(0.711)
Observations	60250	27631	24178	24774	22057
R^2	0.07	0.04	0.05	0.04	0.05
Weekend	0.259	-0.515	0.162	-0.111	0.282
	(0.532)	(0.797)	(0.833)	(1.001)	(0.946)
Observations	60493	27326	24412	24662	22353
R^2	0.05	0.06	0.06	0.05	0.06

Data are from ATUS and BLS LAUS (2003-2022). The dependent variable is daily sleep in minutes for respondents aged 25-55. Column 2-5 represents individuals employed in industries where the percentage of telework exceeds the mean. Industries with a higher concentration of telework experience smaller decreases in sleep with increasing employment.

◂ Periods ◂ Weekday Figure ◂ Weekday Table

Table of Contents

Conclusion

- Weekday sleep is countercyclical
  - A 1% increase in employment rate reduces weekday sleep by around 2 minutes
- Increase in telework following the COVID-19 pandemic
  - Industries with a higher concentration of telework experience smaller decreases in sleep with increasing employment
- Heterogeneous analysis
  - Minorities, less educated individuals, women, and single adults face greater reductions in sleep during weekdays

◂ Periods

Chapter 3

Caste Differences in Child Growth: Disentangling Endowment and Investment Effects

Joseph Cummins[1] Jingyan Guo[1] Neha Agarwal[3] Anaka Aiyar[2]
Vaishali Jain[1] Andrew Bergmann[4]

[1]University of California, Riverside
[2]University of Vermont
[3]University of Otago
[4]Ohio State University

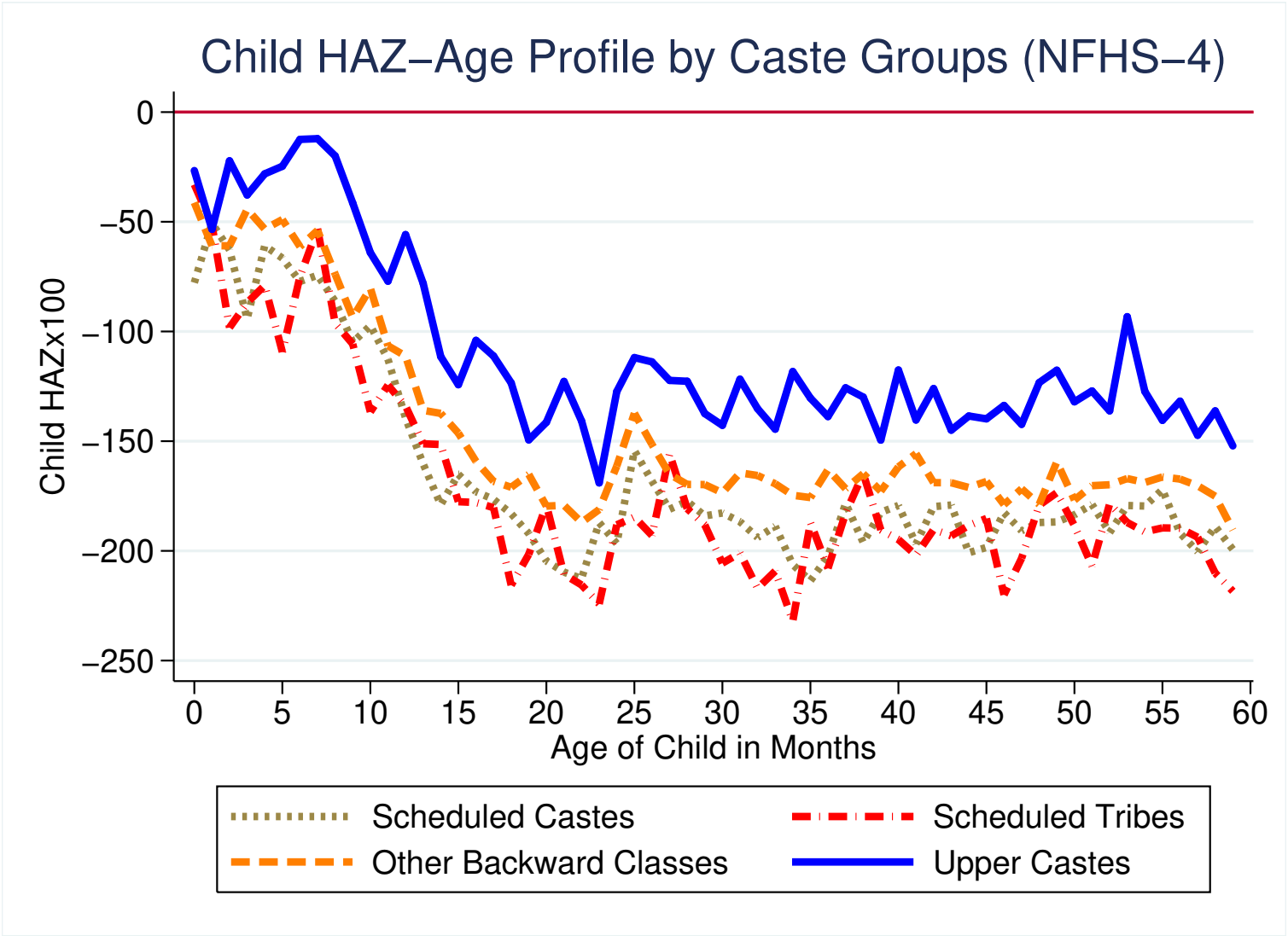
Final Defense of Dissertation

May 29th, 2024

Table of Contents

Motivation and Background

- The child height-for-age z-score (HAZ) differences across Indian caste groups are large
  - (Van de Poel & Speybroeck, 2009; Coffey et al., 2019; Ramachandran & Deshpande, 2021)
- India's caste system
  - Scheduled Caste (SC), Scheduled Tribe (ST), and Other Backwards Castes (OBC)
  - Upper Caste (UC)
- The lower caste children (SC, ST, and OBC) have lower HAZ scores compared with upper caste (UC) children



Median child HAZ for WHO reference population children is 0. The mean child HAZ in this sample is -1.49. A child with HAZ between -2 and -3 indicates moderate chronic malnutrition (stunted), and HAZ below -3 indicates severe chronic malnutrition. The results are weighted by sample weights.

Research Questions

- What are the biological age-dynamics of disparities in child growth patterns across castes?
  - Are the gaps constant as children grow?
  - Do different variables matter at different points in time?
- Can we use health capital theory to disentangle the endowment and investment effects?
  - Based on age-dynamics of HAZ disparities and coefficient estimates

Literature of Decompositions in HAZ Caste Gaps

- Decompositions in explaining caste gaps in HAZ
  - Van de Poel and Speybroeck (2009)
    - ⋆ 1992-3 National Family Health Survey (NFHS-1)
  - Coffey et al., (2019); Ramachandran & Deshpande (2021)
    - ⋆ 2005-6 NFHS-3
  - The gaps are explainable by observed socioeconomic differences such as household assets and parental education

- Contribution
  - Use recent datasets: (NFHS-4, 2015-16) and (NFHS-5, 2019-2021)
  - Provide insights into the economic-biological as opposed to socioeconomic processes behind child growth differentials
  - Include age dynamic explanation for differential processes
  - Speak to timing of policies and interventions

Literature of HAZ-Age Profile

- Documenting HAZ-Age Profile
  - Victora et al., (2010): global pattern in child growth
  - Rieger & Trommlerová (2016): pattern persists
  - Headey et al., (2019): prenatal v.s. postnatal
- Bias and Model Mis-specification with Cohort variation
  - Cummins (2014): birth cohort and HAZ
  - Agarwal et al., (2017): month of birth and HAZ
- Age-dynamic Correlations with Single ("Treatment") Variable
  - Aiyar & Cummins (2019): GDP and HAZ
  - Aiyar & Cummins (2021): add the Indian Enigma context, which allows for health capital interpretation

Table of Contents

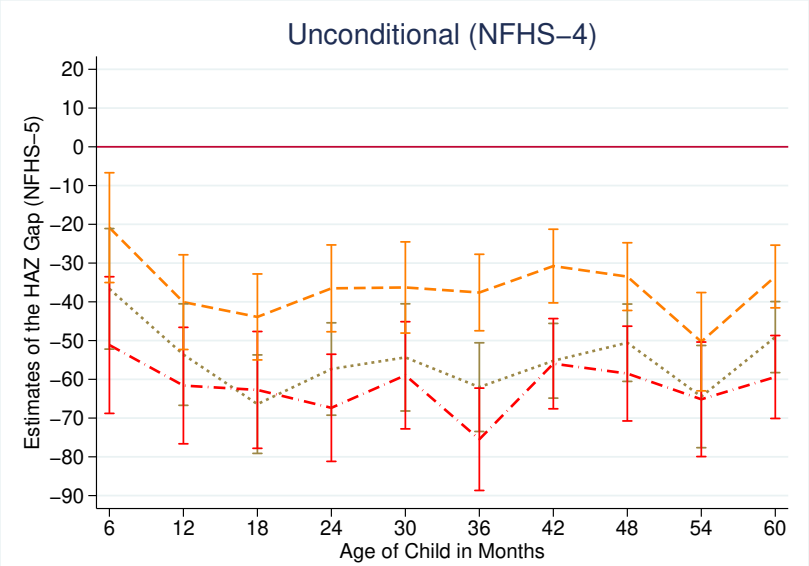
Standard Linear Regression Model

We estimate age-specific mean HAZ differences across caste group (relative to UC children) using a standard linear regression model:

$$Y^A_{irvg} = Caste'^A_g \delta^A_g + X'^A_{1irvg} \beta^A_1 + X'^A_{2v} \beta^A_2 + \lambda^A_r + \epsilon^A_{irvg} \quad (6)$$

- Y^A_{irvg}: the HAZ of child i, aged A, living in region r in PSU v and belonging to caste g
- $Caste^A_{irvg}$: SC, ST, or OBC (with UC omitted)
- X_1 and X_2: individual and village level explanatory variables
- δ^A_g: age-specific estimates of the HAZ gap
- We estimate this equation separately for each 6-month child age group, allowing δ, β, and λ to vary by age

Standard Linear Regression Results



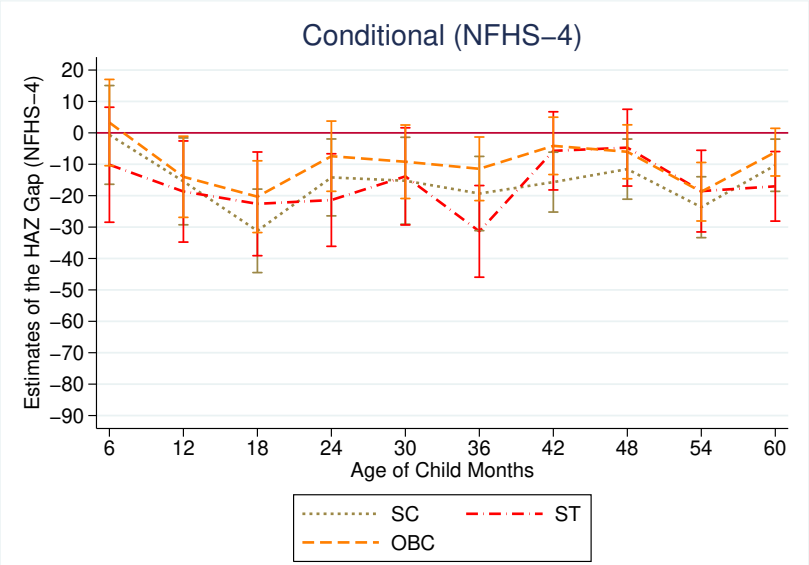


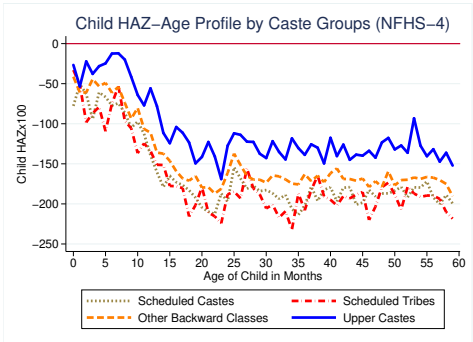
Table of Contents

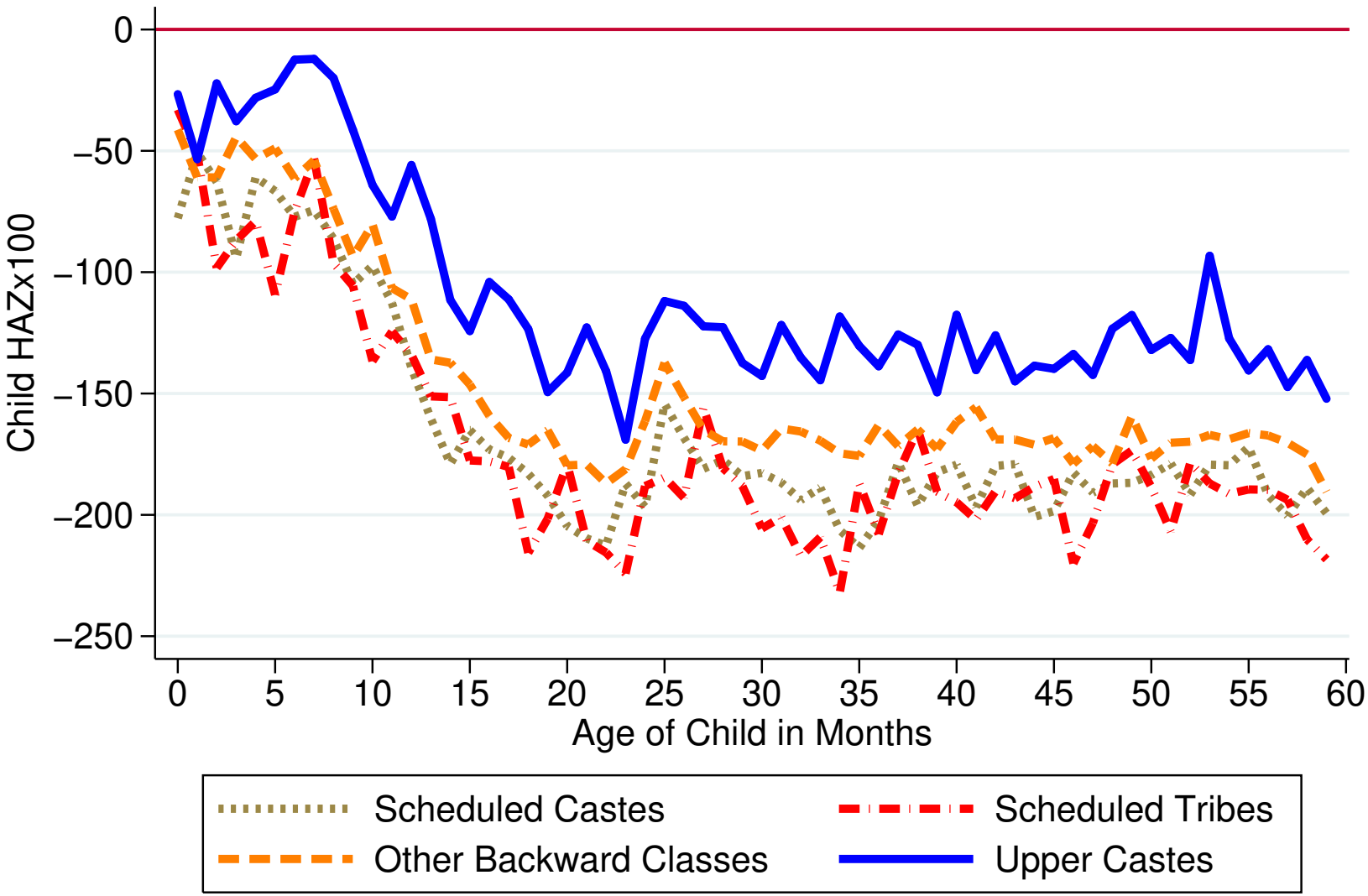
α / β Regression Model (Stage 1)

Consider a region-by-caste level HAZ-age profile (cell), for region r and caste group g:

$$Y_i^{rg} = \alpha^{rg} + \beta^{rg} * Age_i^{rg} + \epsilon_i^{rg} \tag{7}$$

- Y_i^{rg} is the HAZ of child i from region r and caste-group g
- $\hat{\alpha}^{rg}$ is our measure of the health endowment for this group
  - The implied birth HAZ
  - The intercept of the HAZ-age profile
- $\hat{\beta}^{rg}$ (conditional on $\hat{\alpha}^{rg}$) is our measure of child investment
  - The rate of loss of HAZ
  - The slope of the HAZ-age profile
- Focus on the children between 0 to 24 months




Child HAZ–Age Profile by Caste Groups (NFHS–4)
Child HAZx100
0
−50
−100
−150
−200
−250
0
5
10
15
20
25
30
35
40
45
50
55
60
Age of Child in Months
Scheduled Castes
Scheduled Tribes
Other Backward Classes
Upper Castes

α / β Regression Model (Stage 2)

Then, with R regions and G caste groups, we have R*G observations to estimate:

$$\hat{Z}_{rg} = \delta^g * Caste_g + Endowment'_{rg}\beta_1 + Private'_{rg}\beta_2 + Public'_{rg}\beta_3 + \epsilon_{rg} \quad (8)$$

- $\hat{Z}_{rg}$: health endowment ($\hat{\alpha}_{rg}$) or child investment ($\hat{\beta}_{rg}$)
- δ^g: estimated caste gap for health endowment or child investment
- Weighted by cell size of each district-urban-caste group
- Clustered by district-urban group

▸ Summary Statistics

α / β Regression Results

	(1) α b/se	(2) α(| X) b/se	(3) β b/se	(4) β(| X) b/se	(5) β(| α) b/se	(6) β(| X,α) b/se
SC	-46.65***	-4.72	-1.46**	-0.47	-3.90***	-0.74**
	(9.88)	(13.45)	(0.60)	(0.90)	(0.41)	(0.35)
ST	-54.80***	-8.68	-0.92	-0.83	-3.79***	-1.32***
	(15.23)	(16.31)	(0.86)	(1.03)	(0.49)	(0.39)
OBC	-30.67***	-1.53	-0.72	-0.36	-2.32***	-0.45*
	(8.32)	(10.63)	(0.57)	(0.70)	(0.35)	(0.24)
Mean	-37.3	-37.3	-7.8	-7.8	-7.8	-7.8
R Square	0.04	0.25	0.01	0.18	0.68	0.81
Weighted N	37959	37959	37959	37959	37959	37959
Real N	920	920	920	920	920	920

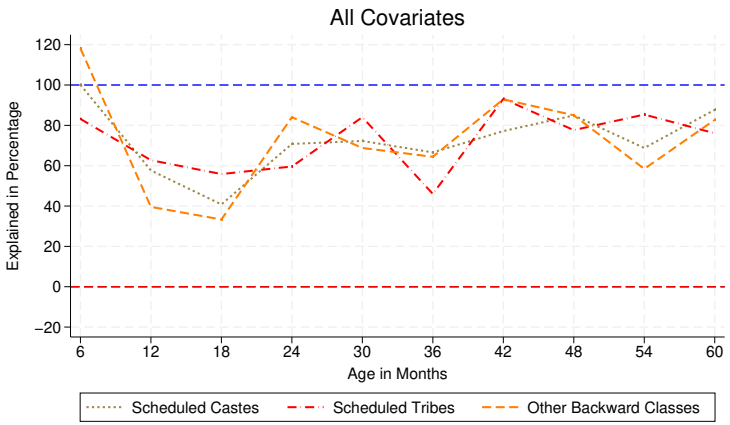
The results are weighted by numbers of individuals in each district-caste-urban cell and clustered at state-urban level. The covariates used include endowment, private investment, and public health variables. The state-urban fixed effects are included in the public health variables. Age Cutoff = Months. p-values: * 0.10, ** 0.05, *** 0.01.

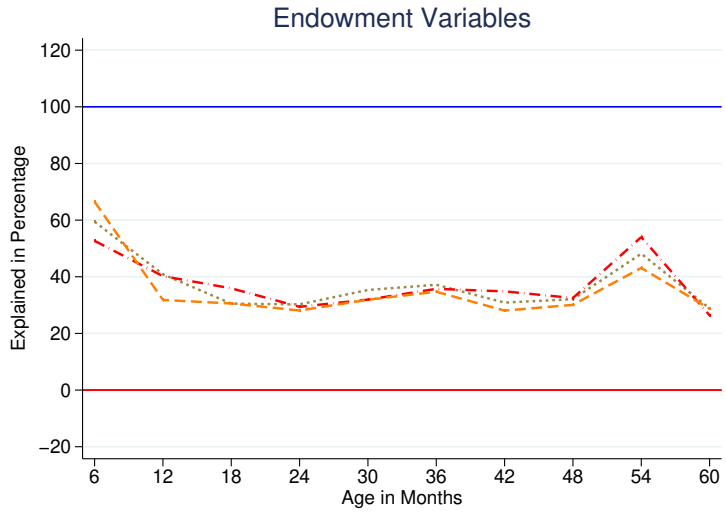
Table of Contents

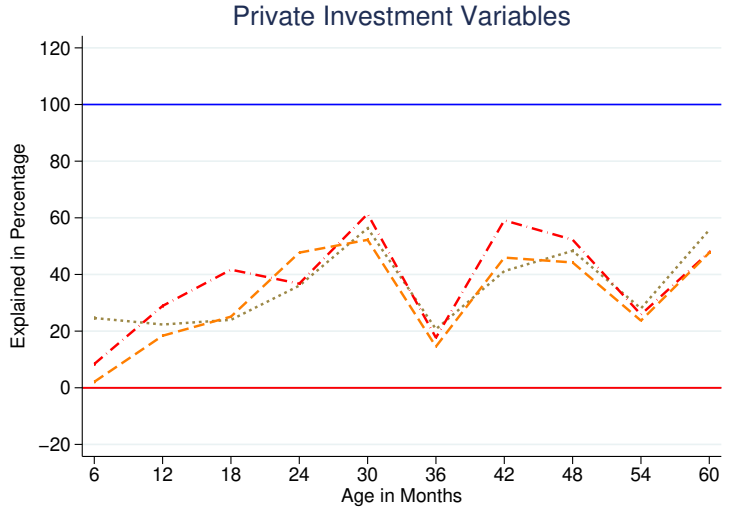
Disentangle Endowments and Investments Effects

- Oaxaca-Blinder decomposition
  - What fraction of the gaps in the unconditional regressions are explainable by what group of variables?
- If endowment variables explain the gap at each age
  - Maybe this is purely intergenerational transmission of health
- If investment variables begin to gain explanatory power after birth
  - Maybe investment-related effects explain part of the gap

Oaxaca-Blinder Decomposition Results in Percentage







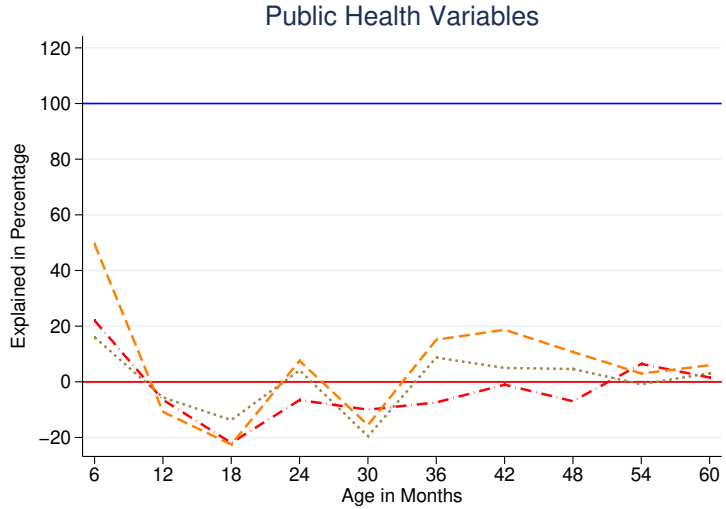


Table of Contents

Conclusion

- HAZ gaps across Indian castes are large
  - Present at birth
  - Increasing in size over the first 5 years
- HAZ gaps are largely explainable
  - Adjusted regression coefficients are close to 0 (all ages)
- Determinants of gaps change over age
  - Endowment variables remain persistent
  - Private investments become more important as children grow
  - By age 5, each explains half of the gap
  - A multi-generational process to remediate HAZ caste gaps

Thank You!

Appendix

Table of Contents

Short Run IV Estimates of the Effects of Sleep on Log of Crashes (Seasonality)

	(1) Crashes b/se	(2) Crashes b/se	(3) Crashes b/se	(4) Crashes b/se	(5) Crashes b/se
Average Monthly Sleep	−0.044***	−0.024***	−0.025***	−0.023***	−0.024***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Mean	-0.41	-0.41	-0.41	-0.41	-0.41
Controls	No	Yes	Yes	Yes	Yes
County FEs	No	Yes	Yes	Yes	Yes
Year FE	No	No	Yes	No	No
Year-Month FE	No	No	No	Yes	No
County-Month FE	No	No	No	No	Yes
Observations	36296	36296	36296	36296	36296

Note: Sleep and sunset time are measured in hours at county-year-month level. The dependent variable of sleep is monthly average sleep hours. The dependent variable of crashes refers to fatal crashes per 100,000 population at state-county-month level. Controls include socio-demographics (age, race, sex, education, marital status, nativity status, and number of children), geographic characteristics (latitude, longitude, and indicator for large counties), and interview characteristics (indicators for holiday and weekend). Seasonality are captured by adding time fixed effects. The standard errors are robust to heteroscedasticity and clustered at state-county level (reported in parentheses). F test on the excluded instrument. Significance levels: * 0.10, ** 0.05, *** 0.01

◂ IV Short Run

Short Run Effects of Sleep on Log of Crashes by Types of Roads

	(1) Crashes (All Roads) b/se	(2) Crashes (Highway) b/se	(3) Crashes (County Road) b/se	(4) Crashes (Local Street) b/se
Average Monthly Sleep	−0.024***	−0.019**	-0.006	-0.002
	(0.01)	(0.01)	(0.00)	(0.01)
Mean	-0.41	-0.92	-1.33	-1.27
Controls	Yes	Yes	Yes	Yes
County FEs	Yes	Yes	Yes	Yes
Observations	36296	30623	16172	23986

Notes: Sleep is measured in hours at county-year-month level. The dependent variable of sleep is monthly average sleep hours. The dependent variable of crashes refers to fatal crashes per 100,000 population at county-year-month level by different types of roads. Controls include socio-demographics (age, race, sex, education, marital status, nativity status, and number of children), geographic characteristics (latitude, longitude, and indicator for large counties), and interview characteristics (indicators for holiday and weekend). The standard errors are robust to heteroscedasticity and clustered at state-county level (reported in parentheses). F test on the excluded instrument. Significance levels: * 0.10, ** 0.05, *** 0.01.

◂ IV Short Run

Short Run Effects of Sleep on Log of Crashes by Light Condition

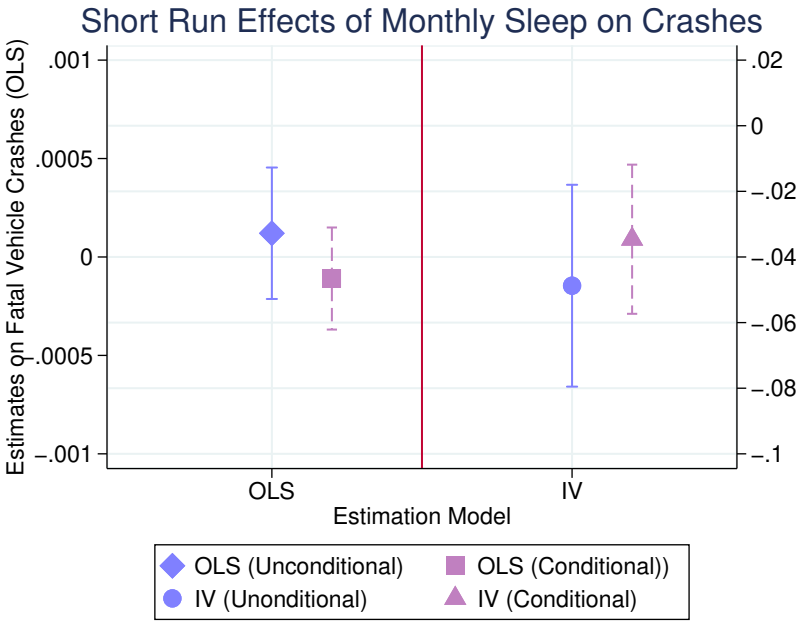
	(1) Crashes (All Hours) b/se	(2) Crashes (Morning) b/se	(3) Crashes (Evening) b/se	(4) Crashes (Daytime) b/se	(5) Crashes (Nighttime) b/se
Average Monthly Sleep	−0.024∗∗∗	0.025	0.155	−0.021∗∗∗	−0.000∗∗∗
	(0.01)	(0.02)	(0.12)	(0.01)	(0.00)
Mean	-0.41	-1.71	-1.43	-1.28	-1.88
Controls	Yes	Yes	Yes	Yes	Yes
County FEs	Yes	Yes	Yes	Yes	Yes
Observations	36296	18075	23521	27022	28100

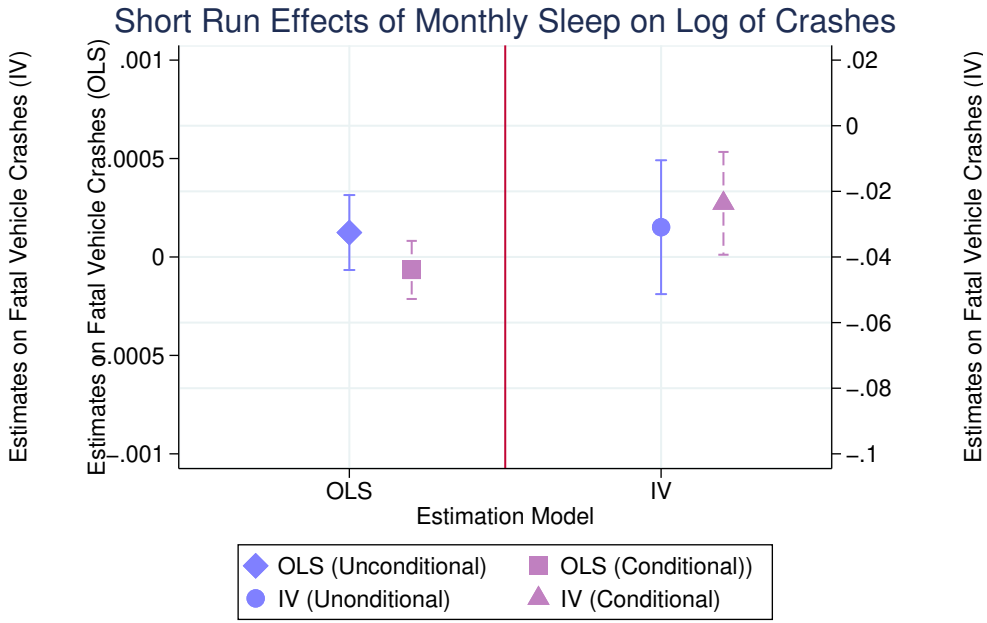
Notes: Sleep is measured in hours at county-year-month level. The dependent variable of sleep is monthly average sleep hours. The dependent variable of crashes refers to fatal crashes per 100,000 population at county-year-month level by different types of roads. Controls include socio-demographics (age, race, sex, education, marital status, nativity status, and number of children), geographic characteristics (latitude, longitude, and dummy for large counties), and interview characteristics (indicators for holiday and weekend). The standard errors are robust to heteroscedasticity and clustered at state-county level (reported in parentheses). "Morning" is defined as +/- two hours from the average sunrise time in that county. "Evening" is defined as +/- two hours from the average sunset time in that county. Significance levels: * 0.10, ** 0.05, *** 0.01.

◂ IV Short Run

Table of Contents

Short Run Effects of Monthly Sleep on Crashes (IV) (v1)





◂ Graph (Version 2)

Notes: This graph shows estimates of the short run effects of monthly sleep on crashes using OLS and IV. The error bars are at 95% confidence intervals for the mean. Sleep denotes monthly average sleep hours. The dependent variable of crashes refers to fatal crashes per 100 millions VMT at county-year-month level. Controls include socio-demographics (age, race, sex, education, marital status, nativity status, and number of children), geographic characteristics (latitude, longitude, and indicator for large counties), and interview characteristics (indicators for holiday and weekend).

Effects of Locating on the Late Sunset Side on Sleep for Employed (2004-2019)

	(1) Sleep Hours b/se	(2) Sleep Hours b/se	(3) Sleep Hours b/se	(4) Sleep Hours b/se	(5) Sleep Hours b/se	(6) Sleep Hours b/se	(7) Sleep $\geq 8hrs$ b/se
Late Sunset Side=1	-0.158	0.035	−0.241∗	0.119	0.354∗∗	-0.009	-0.009
	(0.12)	(0.11)	(0.14)	(0.17)	(0.16)	(0.18)	(0.03)
Mean	8.59	8.59	8.59	8.59	8.59	8.59	8.59
Controls	No	Yes	Yes	No	Yes	Yes	Yes
County FEs	No	No	Yes	No	No	Yes	No
Bandwidth (miles)	250	250	250	100	100	100	250
Observations	27542	27542	27542	7172	7172	7172	27542

Data are from ATUS (2004-2019). Estimates include the distance to the time-zone boundary and the interaction with the late sunset border, socio-demographics (age, race, sex, education, marital status, nativity status, and number of children), geographic characteristics (latitude, longitude, and indicator for large counties), and interview characteristics (indicators for holiday and weekend). The standard errors are robust to heteroscedasticity and clustered at state-county level (reported in parentheses). Significance levels: * 0.10, ** 0.05, *** 0.01.

◂ Figures

Effects of Locating on the Late Sunset Side on Log of Fatal Crashes for Employed (2004-2019)

	(1) Crashes b/se	(2) Crashes b/se	(3) Crashes b/se	(4) Crashes b/se	(5) Crashes b/se	(6) Crashes b/se
Late Sunset Side=1	-0.198	-0.347**	-0.260*	0.246	0.185	-0.033
	(0.18)	(0.14)	(0.16)	(0.21)	(0.14)	(0.26)
Mean	8.59	8.59	8.59	8.59	8.59	8.59
Controls	No	Yes	Yes	No	Yes	Yes
County FEs	No	No	Yes	No	No	Yes
Bandwidth (miles)	250	250	250	100	100	100
Observations	27542	27542	27542	7172	7172	7172

Data are from FARS and ATUS (2004-2019). Estimates include the distance to the time-zone boundary and the interaction with the late sunset border, socio-demographics (age, race, sex, education, marital status, nativity status, and number of children), geographic characteristics (latitude, longitude, and indicator for large counties), and interview characteristics (indicators for holiday and weekend). The standard errors are robust to heteroscedasticity and clustered at state-county level (reported in parentheses). Significance levels: * 0.10, ** 0.05, *** 0.01.

◂ Figures

Effects of Locating on the Late Sunset Side on Sleep for Employed (2004-2013)

	(1) Sleep Hours b/se	(2) Sleep Hours b/se	(3) Sleep Hours b/se	(4) Sleep Hours b/se	(5) Sleep Hours b/se	(6) Sleep Hours b/se	(7) Sleep $\geq 8hrs$ b/se
Late Sunset Side=1	-0.443***	-0.273***	-0.229	-0.363*	-0.066	-0.356**	-0.077**
	(0.10)	(0.10)	(0.15)	(0.19)	(0.10)	(0.17)	(0.03)
Mean	8.49	8.49	8.49	8.53	8.53	8.53	0.61
Controls	No	Yes	Yes	No	Yes	Yes	Yes
State FEs	No	No	Yes	No	No	Yes	No
Bandwidth (miles)	250	250	250	100	100	100	250
Observations	8305	8305	8305	2598	2598	2598	8305

Data are from ATUS (2004-2019). Estimates include the distance to the time-zone boundary and the interaction with the late sunset border, socio-demographics (age, race, sex, education, marital status, nativity status, and number of children), geographic characteristics (latitude, longitude, and indicator for large counties), and interview characteristics (indicators for holiday and weekend). The standard errors are robust to heteroscedasticity and clustered at state-county level (reported in parentheses). Significance levels: * 0.10, ** 0.05, *** 0.01.

◂ Figures

Effects of Locating on the Late Sunset Side on Log of Fatal Crashes for Employed (2004-2013)

	(1) Crashes b/se	(2) Crashes b/se	(3) Crashes b/se	(4) Crashes b/se	(5) Crashes b/se	(6) Crashes b/se
Late Sunset Side=1	-0.259	-0.384**	-0.220	0.249	0.233	0.033
	(0.21)	(0.18)	(0.18)	(0.19)	(0.17)	(0.27)
Mean	8.55	8.55	8.55	8.55	8.55	8.55
Controls	No	Yes	Yes	No	Yes	Yes
County FEs	No	No	Yes	No	No	Yes
Bandwidth (miles)	250	250	250	100	100	100
Observations	20027	20027	20027	5160	5160	5160

Data are from FARS and ATUS (2004-2013). Estimates include the distance to the time-zone boundary and the interaction with the late sunset border, socio-demographics (age, race, sex, education, marital status, nativity status, and number of children), geographic characteristics (latitude, longitude, and indicator for large counties), and interview characteristics (indicators for holiday and weekend). The standard errors are robust to heteroscedasticity and clustered at state-county level (reported in parentheses). Significance levels: * 0.10, ** 0.05, *** 0.01.

◂ Figures

Effects of Locating on the Late Sunset Side on Sleep for Employed (2014-2019)

	(1) Sleep Hours b/se	(2) Sleep Hours b/se	(3) Sleep Hours b/se	(4) Sleep Hours b/se	(5) Sleep Hours b/se	(6) Sleep Hours b/se	(7) Sleep $\geq 8hrs$ b/se
Late Sunset Side=1	0.385**	0.657***	0.023	0.405*	0.597***	0.404	0.090**
	(0.19)	(0.14)	(0.19)	(0.21)	(0.19)	(0.26)	(0.04)
Mean	8.63	8.63	8.63	8.64	8.64	8.64	0.65
Controls	No	Yes	Yes	No	Yes	Yes	Yes
State FEs	No	No	Yes	No	No	Yes	No
Bandwidth (miles)	250	250	250	100	100	100	250
Observations	3605	3605	3605	1108	1108	1108	3605

Data are from ATUS (2004-2019). Estimates include the distance to the time-zone boundary and the interaction with the late sunset border, socio-demographics (age, race, sex, education, marital status, nativity status, and number of children), geographic characteristics (latitude, longitude, and indicator for large counties), and interview characteristics (indicators for holiday and weekend). The standard errors are robust to heteroscedasticity and clustered at state-county level (reported in parentheses). Significance levels: * 0.10, ** 0.05, *** 0.01.

◂ Figures

Effects of Locating on the Late Sunset Side on Log of Fatal Crashes for Employed (2014-2019)

	(1) Crashes b/se	(2) Crashes b/se	(3) Crashes b/se	(4) Crashes b/se	(5) Crashes b/se	(6) Crashes b/se
Late Sunset Side=1	-0.083	-0.270**	-0.319*	0.184	0.009	-0.556**
	(0.19)	(0.13)	(0.17)	(0.29)	(0.15)	(0.25)
Mean	8.69	8.69	8.69	8.69	8.69	8.69
Controls	No	Yes	Yes	No	Yes	Yes
County FEs	No	No	Yes	No	No	Yes
Bandwidth (miles)	250	250	250	100	100	100
Observations	7515	7515	7515	2012	2012	2012

Data are from FARS and ATUS (2014-2019). Estimates include the distance to the time-zone boundary and the interaction with the late sunset border, socio-demographics (age, race, sex, education, marital status, nativity status, and number of children), geographic characteristics (latitude, longitude, and indicator for large counties), and interview characteristics (indicators for holiday and weekend). The standard errors are robust to heteroscedasticity and clustered at state-county level (reported in parentheses). Significance levels: * 0.10, ** 0.05, *** 0.01.

◂ Figures

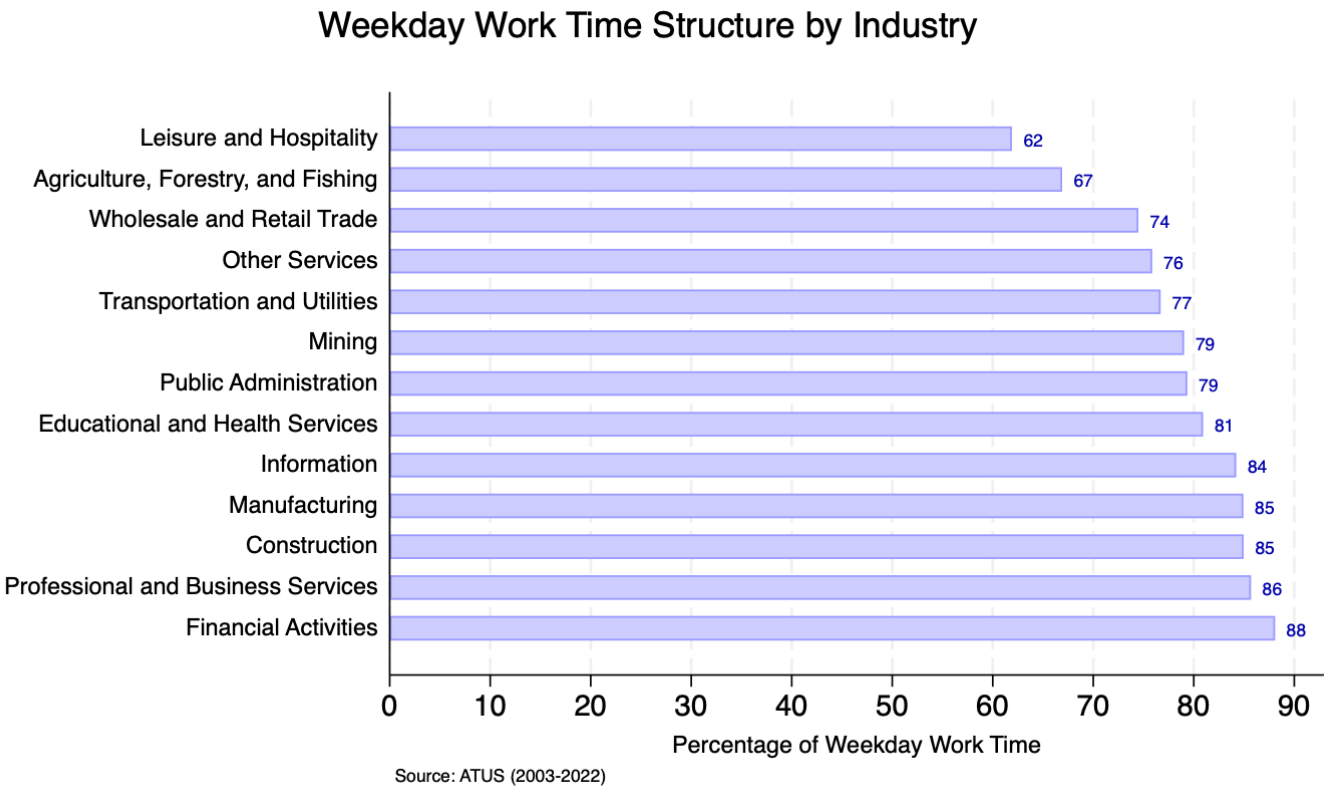
Effects of Employment Rate on Sleep by Periods

	(1) 2003-2022	(2) 2003-2015	(3) 2016-2022	(4) 2003-2007	(5) 2008-2010	(6) 2011-2015	(7) 2016-2019	(8) 2020-2022
All	−1.019**	−1.133*	-0.935	-1.020	-2.399	−2.889*	2.383	−1.688*
	(0.352)	(0.444)	(0.593)	(0.921)	(1.604)	(1.196)	(2.162)	(0.705)
Observations	120743	91184	29559	40772	20772	29640	18404	11155
R^2	0.11	0.11	0.12	0.11	0.12	0.11	0.12	0.12
Weekday	−2.308***	−2.875***	-1.536	-2.075	-2.694	−5.420**	4.251	−3.335**
	(0.441)	(0.512)	(0.938)	(1.118)	(1.441)	(2.021)	(2.475)	(1.052)
Observations	60250	45346	14904	20278	10230	14838	9213	5691
R^2	0.07	0.07	0.07	0.06	0.06	0.08	0.08	0.07
Weekend	0.259	0.549	-0.452	-0.063	-1.790	-1.247	0.498	-0.379
	(0.532)	(0.673)	(1.045)	(1.318)	(2.573)	(1.558)	(3.096)	(1.277)
Observations	60493	45838	14655	20494	10542	14802	9191	5464
R^2	0.05	0.05	0.06	0.06	0.06	0.06	0.05	0.06

Data are from ATUS and BLS LAUS (2003-2022). The dependent variable is daily sleep in minutes for respondents aged 25-55. Each cell represents the estimates of employment to population rate on sleep.

◂ Telework

Weekday Work Time Structure



$$\text{Percentage of Weekday Work Time} = \frac{W_{\text{weekday}}}{W_{\text{weekday}} + W_{\text{weekend}}}$$

◂ Telework

Effects of Employment Rate on Sleep by Work Time Structure and Industry (2003-2022)

	(1) Employed	(2) Not Employed	(3) > Med Weekday	(4) ≤ Med Weekday	(5) Blue-collar	(6) White-collar
All	−1.140∗∗	0.010	−1.010∗	-1.178	-1.279	-0.351
	(0.392)	(0.900)	(0.381)	(0.624)	(1.110)	(0.661)
Observations	97713	23030	73864	29712	18907	20764
R^2	0.13	0.06	0.14	0.10	0.16	0.16
Weekday	−1.992∗∗∗	-2.642	−1.979∗∗∗	−2.120∗	-1.580	-0.313
	(0.445)	(1.391)	(0.458)	(0.804)	(1.011)	(0.844)
Observations	49045	11205	37106	14715	12859	14405
R^2	0.04	0.06	0.04	0.05	0.15	0.15
Weekend	-0.267	2.415	-0.148	-0.169	-0.111	-0.501
	(0.605)	(1.363)	(0.682)	(1.007)	(1.447)	(0.899)
Observations	48668	11825	36758	14997	12788	14291
R^2	0.06	0.05	0.06	0.05	0.14	0.15

Data are from ATUS and BLS LAUS (2003-2022). The dependent variable is daily sleep in minutes for respondents aged 25-55. Each cell represents the estimates of employment to population rate on sleep. Column 1 limits the sample to employed population and column 2 limites to unemployed and population that are not in the labor force. Column 3 limits to below median weekday work time percentage. Column 5 refers to blue-collar workers: construction and manufaturing. Column 6 refers to white-collar workers: financial activities and professional and business services.

◂ Telework

Table of Contents

Summary Statistics by Caste Group

	SC	ST	OBC	UC	All
HAZ (x 100)	-166.79	-171.84	-148.69	-112.61	-149.29
	(164.41)	(169.98)	(166.18)	(160.43)	(166.28)
Endowment Variables					
Mother's Age at Child's Birth	24.21	24.03	24.21	24.44	24.23
	(4.78)	(4.95)	(4.53)	(4.52)	(4.64)
Mother's HAZ (x 100)	-217.65	-211.85	-201.27	-182.60	-203.18
	(93.00)	(89.77)	(95.68)	(94.38)	(94.87)
Birth Order	2.29	2.29	2.12	1.86	2.14
	(1.46)	(1.43)	(1.31)	(1.04)	(1.33)
Delivery Care	0.51	0.47	0.56	0.71	0.56
	(0.50)	(0.50)	(0.50)	(0.45)	(0.50)
Private Investment Variables					
Poorest	0.33	0.50	0.23	0.08	0.26
	(0.47)	(0.50)	(0.42)	(0.28)	(0.44)
Maternal Educ at least Primary	0.65	0.55	0.71	0.89	0.71
	(0.48)	(0.50)	(0.45)	(0.31)	(0.45)
Treat Drinking Water	0.22	0.38	0.31	0.44	0.32
	(0.41)	(0.49)	(0.46)	(0.50)	(0.47)
Owns Agricultural Land	0.33	0.55	0.48	0.49	0.45
	(0.47)	(0.50)	(0.50)	(0.50)	(0.50)
No Toilet Facility	0.60	0.74	0.51	0.25	0.51
	(0.49)	(0.44)	(0.50)	(0.43)	(0.50)
Public Health Variables					
PSU has sewer system access	0.10	0.07	0.12	0.23	0.13
	(0.31)	(0.26)	(0.33)	(0.42)	(0.34)
Urban = 1	0.23	0.11	0.25	0.39	0.26
	(0.42)	(0.32)	(0.43)	(0.49)	(0.44)
N	34913	21441	64206	26218	146778

Summary Statistics

	(1) All mean/sd	(2) Female mean/sd	(3) Male mean/sd	(4) Employed mean/sd	(5) Unemployed mean/sd
Crashes per 100,000 Population	0.74	0.74	0.73	0.74	0.72
	(0.46)	(0.47)	(0.46)	(0.47)	(0.45)
Sleep Duration (Hours)	8.61	8.69	8.52	8.56	9.25
	(2.05)	(2.04)	(2.06)	(2.04)	(2.13)
Age	38.75	38.61	38.90	39.01	35.28
	(9.71)	(9.74)	(9.67)	(9.54)	(11.13)
White	0.78	0.75	0.80	0.78	0.69
	(0.42)	(0.43)	(0.40)	(0.41)	(0.46)
Black	0.14	0.17	0.12	0.14	0.23
	(0.35)	(0.37)	(0.32)	(0.34)	(0.42)
High School	0.48	0.48	0.48	0.47	0.60
	(0.50)	(0.50)	(0.50)	(0.50)	(0.49)
College	0.44	0.45	0.43	0.45	0.21
	(0.50)	(0.50)	(0.49)	(0.50)	(0.41)
Married	0.54	0.50	0.57	0.55	0.39
	(0.50)	(0.50)	(0.49)	(0.50)	(0.49)
Nativity Status	0.77	0.79	0.76	0.77	0.75
	(0.42)	(0.41)	(0.43)	(0.42)	(0.43)
Number of Children	1.09	1.11	1.06	1.08	1.14
	(1.15)	(1.12)	(1.18)	(1.14)	(1.23)
Holiday	0.02	0.02	0.02	0.02	0.02
	(0.13)	(0.13)	(0.12)	(0.13)	(0.13)
Latitude	37.25	37.20	37.29	37.26	37.02
	(5.05)	(5.07)	(5.04)	(5.05)	(5.07)
Large County	0.68	0.68	0.68	0.68	0.72
	(0.47)	(0.47)	(0.46)	(0.47)	(0.45)
Weekend	0.51	0.51	0.50	0.51	0.50
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
N	36296	18569	17727	33811	2485

Data are from ATUS (2004-2019) and FARS (2004-2019). The sample is restricted to people who are in the labor force and aged between 18 and 55. The crashes data from FARS are matched to the ATUS at county-year-month level.

▸ Sleep Duration

Summary Statistics

	(1) All mean/sd	(2) Female mean/sd	(3) Male mean/sd	(4) Employed mean/sd	(5) Not Employed mean/sd	(6) $<$ Bachelor's mean/sd	(7) $\geq$ Bachelor's mean/sd
Sleep Time (Hours)	8.66	8.76	8.53	8.52	9.23	8.45	8.79
	(2.23)	(2.21)	(2.25)	(2.15)	(2.49)	(1.94)	(2.40)
Age	40.42	40.19	40.68	40.38	40.59	40.21	40.55
	(8.49)	(8.55)	(8.41)	(8.39)	(8.89)	(8.15)	(8.71)
Married	0.60	0.59	0.62	0.61	0.59	0.67	0.56
	(0.49)	(0.49)	(0.49)	(0.49)	(0.49)	(0.47)	(0.50)
Number of Children	1.17	1.24	1.08	1.13	1.33	1.16	1.18
	(1.19)	(1.19)	(1.19)	(1.15)	(1.32)	(1.14)	(1.23)
Children Under 3	0.16	0.16	0.15	0.15	0.20	0.18	0.14
	(0.36)	(0.37)	(0.36)	(0.35)	(0.40)	(0.38)	(0.35)
White	0.66	0.65	0.68	0.68	0.58	0.74	0.61
	(0.47)	(0.48)	(0.47)	(0.47)	(0.49)	(0.44)	(0.49)
Black	0.13	0.14	0.11	0.12	0.16	0.09	0.15
	(0.33)	(0.35)	(0.31)	(0.32)	(0.37)	(0.28)	(0.36)
Hispanic	0.15	0.15	0.15	0.14	0.19	0.08	0.20
	(0.36)	(0.36)	(0.36)	(0.35)	(0.40)	(0.27)	(0.40)
American Indian	0.01	0.01	0.01	0.01	0.01	0.00	0.01
	(0.09)	(0.09)	(0.09)	(0.08)	(0.11)	(0.06)	(0.10)
Holiday	0.02	0.02	0.02	0.02	0.02	0.02	0.02
	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)
Incomplete Diary	0.08	0.09	0.07	0.08	0.08	0.09	0.07
	(0.27)	(0.28)	(0.25)	(0.27)	(0.27)	(0.29)	(0.26)
N	120743	65868	54875	97713	23030	48733	72010

Data are from ATUS and BLS LAUS (2003-2022). The sample is restricted to respondents aged 25-55.

▸ Sources of Data