On the Effects of Sunshine on Economic Productivity

Justin Jung Justin Jones Joey Farrell

July 14, 2023

1 Introduction

Sunshine is often perceived to have an impact on productivity in the work place. Anecdotally, we may believe that bright and sunny days are energizing and more conducive to work, while dark and gloomy days may be demotivating. This intuition prompts us to explore the potential relationship between sunshine levels and economic productivity.

We can look to documented literature surrounding the psychological and physiological effects of sunshine to further motivate the investigation of this anecdotal claim. Drawing from An et al. (2016), which establishes a link between sunlight exposure, improved mental health, and work attitudes, we might posit that the psychological benefits of sunshine can increase worker productivity. Moreover, we can look to literature regarding Daylight Saving Time (DST). Daylight Saving Time is a policy enacted to better align sleep and waking schedules with the daylight, possibly to improve productivity and overall well-being given the physiological effects that sunshine has via Vitamin D exposure and modulation of the circadian rhythm. Herber et al. (2017) specifically investigates the effect of daylight (by considering changes in daylight saving time) on student school performance. We wish to generalize this investigation and consider the effect that sunshine has on overall worker productivity.

The notion that sunlight has an effect on economic productivity in general is well-supported by preliminary back-of-the-envelope analysis. Specifically, the existing U.S. sunshine and GDP data demonstrates a promising trend for our hypothesis; Figure 1 below illustrates that real, annual U.S. GDP (in hundreds of millions of dollars) and annual average sunlight exposure (measured by solar irradiance, in kJ/m^2) are positively correlated over the time period spanning from 1991 to 2012. Note that both quantities were de-meaned for purposes of visualizing the correlation.

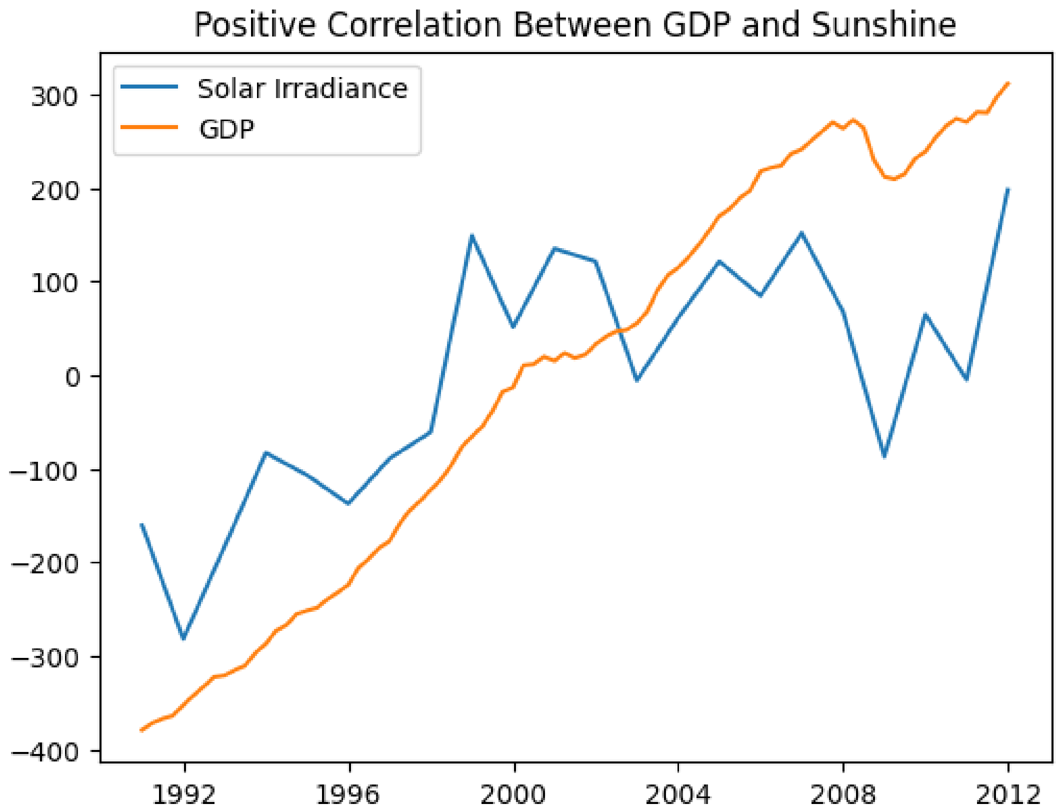

Figure 1: Positive correlation between GDP and sunshine

It is evident that this positive correlation is only proof of the fact that both of these quantities have been generally increasing over time. In order to get a better understanding of a potential causal link between these time series, we will conduct industry-level analyses, account for confounding effects using a collection of natural covariates, adjust for state-level and time-level variations using fixed effects, and consider instrumental regressions. We will additionally construct a "worker productivity" metric (see Section 3) and regress it on solar irradiance levels and additional covariate controls. In this way, we attempt to discern any causal effects that sunshine may have on the average worker.

Overall, we hope that this paper provides insight into the marginal effect that additional exposure to a unit of sunlight has on the productivity of the U.S. economy. We implement many regression techniques in an effort to establish tenable causal links, but note that there are endogeneity and confounding risks associated with sunlight levels and GDP. In observance of this fact, we will pay close attention to our regression assumptions and provide substantial economic interpretations of our results.

2 Literature Review

Prior researchers have explored the impact of environmental factors on workplace productivity, including the effect of sunshine. For example, Boubekri et al. (2014) investigated the relationship between light exposure at work and sleep quality outside the workplace. Their study revealed that workers with light exposure in their offices averaged 46 more minutes of sleep per night compared to those in windowless offices. This finding suggests that sunshine exposure at work may positively affect individuals' well-being and potentially enhance productivity. Additionally, Rosekind et al. (2010) found that poor sleep results in lost productivity that costs an estimated 1,967 USD per employee, measured in terms of 2007 dollars. This indirectly furthers the idea that light exposure has a positive relationship with productivity.

In contrast to the positive effect of sunshine, Lee et al. (2014) found an interesting relationship between bad weather and productivity. Using the concept of limited attention, they observed that adverse weather conditions, by limiting outdoor distractions, led to greater speed and accuracy in data-entry roles. This study highlights the complex interplay between environmental factors and productivity, suggesting that the impact of weather on productivity may vary depending on tasks and their cognitive demands.

Furthermore, environmental factors such as air pollution and temperature have also been studied in relation to productivity. Kahn and Li (2020) examined the effect of air pollution on productivity and found that it leads to a decrease in overall productivity levels of highly skilled judges; there was a reported 0.182% increase in case handling time per 1% increase in $PM_{2.5}$. LoPalo (2023) found that temperature also has a negative effect on productivity. They concluded that hot and humid regions may have largest potential impacts on productivity due to climate change. While these two studies focused on air pollution and temperature, it emphasizes the importance of understanding how various environmental factors, including sunshine, can influence productivity outcomes.

However, there is a gap in the existing literature concerning the relationship between sunshine and workplace productivity, particularly regarding worker productivity as measured by GDP or its derivatives. Our research aims to fill this gap by investigating the direct impact of sunlight exposure on worker productivity.

It is worth noting that previous empirical studies on productivity have predominantly employed fixed effects regression models. Following this, we utilize fixed effects regression and consider many ablations with different covariates, outcome variables, and controls to solidify our analysis. Moreover, to extend our analysis we also employ instrumental variables, which is not commonly found in existing literature for this subject, to examine the relationship between sunshine and productivity more comprehensively.

By incorporating instrumental variables into our analysis, we can address endogeneity issues that may arise when examining the effect of sunshine on productivity. In addition, this approach allows for a more robust assessment of the causal relationship between sunlight exposure and productivity of workers.

Overall, through our research, we contribute to the existing literature by examining the specific impact of sunshine on productivity, considering the broader economic implications, and employing instrumental variable analysis to address methodological challenges associated with our research context.

3 Data

In this section, we describe the various sources of data used to investigate the connection between sunlight and productivity. The sources include governmental industry and output data, as well as weather data.

- Productivity

  To gather relevant data for our study on productivity and output, we utilized several reputable sources, including the Bureau of Economic Analysis (BEA), the Bureau of Labor Statistics (BLS), and St. Louis Federal Reserve Economic Data (FRED).

  We obtained current state annual GDP and current state quarterly GDP data from the BEA. State annual GDP figures were measured in millions of chained 2012 dollars, covering the period from 1997 to 2022. The calculations were performed on unrounded data, using the chain-type quantity index and the 2012 current value of the series divided by 100. On the other hand, current quarterly GDP data was measured in millions of current dollars for each quarter from 2005 to 2022. The data was seasonally adjusted at annual rates to ensure consistency.

  For industry-specific GDP data, we relied on FRED. We obtained industry GDP figures for twelve high-level two-digit NAICS industry groupings and two specific NAICS groupings. Data for each state were combined into one dataset for analysis. Specifically, we examined industry codes 11, 21, 22, 23, 31-33, 44-45, 52, 53, 54, 62, 71, 5415, and 111-112 by state. These industries were chosen based on regression analysis that demonstrated their having a potentially interesting relationship with sunlight. However, it's worth noting that the availability of historical data for NAICS codes is limited. In 1997, industry designations were transitioned from Standard Industrial Classification (SIC) to NAICS. This change led to some categories being redefined or newly introduced, and because this would potentially impact the outcomes of our research, we constrained our analysis to years after and including 1997.

  In order to calculate our metric for individual worker productivity, in contrast to overall state-level industry productivity rates which were measured via pure real GDP, we computed the ratio between the total wages paid to laborers in a year within a fixed state and that same state's annual GDP. The wage data was drawn from the BLS, in particular BLS's Quarterly Census of Employment and Wages (QCEW). This census data is gathered by polling employers, and accounts for 95% of U.S. jobs at the county, MSA, state, and national levels by industry. However, because employer reports of wages in census data will always be current values, this necessitated using current (i.e., not inflation-adjusted) GDP values, which were accessed from the BEA via the data table SQGDP2 "Gross domestic product (GDP) by state." With both of these metrics in hand, it was then trivial to compute the worker productivity estimate by dividing the former quantity by the latter in every year and for every state.

- Environmental

  The data on solar irradiance, natural disaster occurrences, and climatic variables were obtained from reputable datasets, including the Centers for Disease Control and Prevention (CDC)'s solar irradiance data, the National Centers for Environmental Information (NCEI)'s state-level natural disaster data, and the Climatic Research Unit (CRU) High-resolution gridded time-series datasets for precipitation and cloud cover.

  Sunshine data was sourced from the CDC's annual and quarterly solar irradiance datasets, covering the period from 1991 to 2012. To calculate the annual averages, the data was grouped by state, county, and year, and the mean global solar irradiance (GHI) value was calculated for each group. GHI represents the shortwave irradiance from the sun on the Earth's horizontal surface, which is the sum of the direct irradiance and the diffuse horizontal irradiance (Maxwell, Wilcox, and Rymes 1994). The mean values were initially measured in watt hours per square meter (Wh/m^2) and were converted to kJ/m^2 by multiplying GHI by a conversion factor of 3.6. The monthly averages were calculated similarly, with the data grouped by county, year, and month while calculating the mean. In order to compare our data with the quarterly GDPs, we took the sum of each relevant three month period.

  The NCEI United States state-level natural disaster data spanned from 1980 to 2023 and consisted of a time series of the number and approximate cost of billion-dollar disaster events, CPI-adjusted. The cost calculations included insured and uninsured direct costs, such as building damage, material assets within buildings, and time lost from business interruption. To account for uninsured and under-insured losses, scaling factors were applied. However, losses to natural capital or assets, such as loss of life and healthcare-related loss, were not included in the analysis.

Lastly, CRU high-resolution gridded time-series datasets were utilized to gather information on precipitation and cloud cover from 1901 to 2022. This dataset includes variables for latitude, longitude, month, precipitation, and cloud cover. The data was derived from observations collected at weather stations worldwide. Precipitation is represented as a percentage anomaly ranging from -100 to 0, based on each station's 1961-1990 monthly averages. Cloud cover is derived from observations of sun hours as described in Harris et al. (2014). These datasets were linked with data on the center coordinate locations of the 48 continental United States and the District of Columbia to calculate the averages.

4 Methods

4.1 Fixed Effects

We used fixed effects regressions to isolate the effect of sunshine on worker productivity. We recognize that there may be confounding variables in our regression of worker productivity (as measured by GDP or labor share) on the levels of sunshine irradiance. Thus, we employed fixed effects regressions to control for factors that would otherwise impact our interpretation of the causal effect of sunshine. In particular, we incorporated both entity effects and time effects into our model, which enables us to account for both state-level heterogeneity and macroeconomic shocks that impact the entire U.S. economy in a given year (e.g., COVID-19).

4.1.1 Entity Effects

Entity effects, specifically state-specific effects, were included in the regression model to control for factors that may affect the statistical inference of the regression. For instance, states with strong agricultural sectors would be positively impacted in terms of productive output (GDP) with increased levels of sunshine as a result of higher crop yields. Similarly, states that have a focus on tourism and leisure (such as Florida) would see variations in worker productivity due to the increase or decrease of sunshine. For our analysis, we want to isolate the effect of increases of worker productivity as induced by the effects of sunshine on the worker rather than overall productivity changes resulting from factors such as crop yields or tourism demand. Thus, to control for this extraneous state or "entity" level factor, we add entity effect controls in our regression, one for each of the 48 contiguous states plus the District of Columbia.

Our model with only entity effects is precisely

$$y_{it} = \alpha_i + \beta' x_{it} + \epsilon_{it},$$

where α_i corresponds to the entity level control and $i \in [1, 2, ..., 49]$.

4.1.2 Time Effects

We also consider time effects in our regression model. Because our outcome variable is either GDP (a direct measure of productive output) or labor share of GDP (which directly depends on GDP), we recognize that productivity shocks at certain time periods can corrupt our causal analysis of the effect of sunshine levels. In particular, we claim that productivity shocks, such as global economic crises, geopolitical shocks, macroeconomic shocks (e.g trade wars), are unrelated to the effect sunshine levels have on worker productivity and thus want to control and partial out the effect of productivity shocks from our model. In doing so we can prevent any unwanted correlation between sunshine levels and worker productivity, which might come about for example in periods such as the aftermath of the financial crisis of 2008, when there happen to be low sunshine levels and also low overall productivity. Thus we include a time effect variable, one per year, to account for these effects.

Our model with only time effects is precisely

$$y_{it} = \gamma_t + \beta' x_{it} + \epsilon_{it},$$

where α_t corresponds to the time effect control for time period t.

4.2 Covariate Controls

To address significant confounding variables in the regressions of industry-level GDP and sunshine levels, particularly considering intra-industry correlations, certain covariate controls were included. The agriculture industry posts a clear story: the amount of sunshine directly impacts animal and plant vitality in farms, thereby affecting output yields; this is a fact which will be made more or less obvious via preliminary regressions in Section 5. As such, one of the covariates that we include in our regressions is the GDP of the US agriculture industry. In this way, we will be better able to isolate the causal effects of sunlight on distinct industries with less obvious dependencies on the weather.

The effect of natural disasters is generally quite local, and surely there are very few if any natural disasters which have impacted the entirety of the United States. Such effects are not already accounted for via our fixed time effects, so if we wish to discount another relatively obvious consequence of bad weather from our regression analysis we must manually project it out by making it an additional covariate. Thus, we include the annual estimated cost due to natural disasters in a given state in our regression. As stated in the data section, we use an estimation here because the provided data consists of categorical ranges of natural disaster costs. Specifically, we choose to take the mean of the range as our covariate value.

4.3 IV

Even with covariate controls, entity effects and time effects, we might be concerned about endogeneity in our model. Thus, we implement instrumental variable regressions. Here, our instrument for sunshine is the level of cloud cover.

We claim that cloud cover is a valid instrument. We will first demonstrate that it satisfies the relevance requirement. Intuitively, it ought to seem true that cloud cover is (negatively) correlated with sunshine levels, as a greater presence of clouds in the sky should result in more sunlight being scattered and/or blocked on its way to Earth. To verify this, consider the scatter plot below in Figure 2, which visually proves the negative correlation between the two quantities. Quantitatively, we have a large Pearson's coefficient of -0.84.

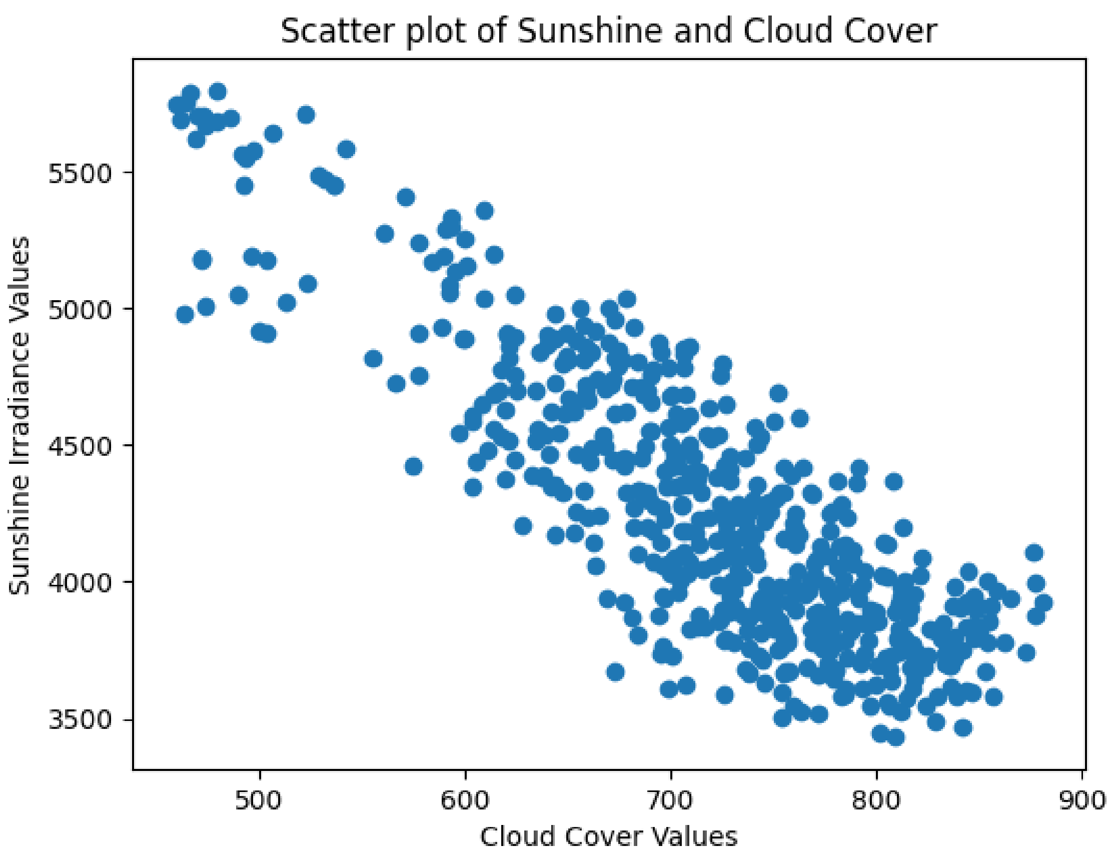

Figure 2: Negative Correlation between Sunshine and Cloud Cover

We also claim that cloud cover satisfies the exclusion restriction assumption. We believe that this is a valid because we find it unlikely that the percentage of the sky covered by clouds has a direct effect on productivity levels that does not factor through the amount of sunlight. In particular, cloud cover is primarily driven by atmospheric conditions, moisture levels, and weather patterns, which we believe are orthogonal to significant economic variables within the United States after controlling for some form of geographic region.

The IV model is characterized by the assumption that sunlight data (in x) is endogenous, with the additional covariates being in t and the exogenous cloud cover values being in z. It is assumed that the variables satisfy the following relationships:

$$x = z'\gamma + v_i \text{ and}$$

$$y = \beta_0 + \beta_1' x + \beta_2' t + u_i.$$

Note that the formulas above are a mathematical formalization of the relevance (correlation between x and z) and exclusion (exogeneity of z) conditions already mentioned above.

5 Results

5.1 Fixed Effect Regression without Covariates

The relevant table here is Table 1. Immediately, it is evident that there is a select group of economic industries within the United States whose output is significantly correlated with the amount of sunshine exposure. Some of these industries are only significant ($p < 0.05$) when no fixed effects are accounted for. This would include Mining; Construction; Retail Trade; Real Estate and Rental and Leasing; Arts, Entertainment, and Recreation; Computer Systems Design and Related Services; and Professional, Scientific, and Technical Services. For these such industries, our natural economic interpretation is that there is a latent confounding factor. In particular, it is logical that industries like Real Estate and Rental and Leasing or Arts, Entertainment, and Recreation would be correlated to sunshine through geographic location in the U.S., as the states with sunnier and more pleasant weather have an easier time attracting vacation home buyers and tourists–generally those seeking a warm-weather escape. As additional examples, states like California and Texas are known for the size of their technology industries, and therefore attribute heavy weight to the correlation with Sun exposure, and the industrial Rust Belt region and its eventual decline driving immigration into the Sun Belt is one likely confounding factor for the relationship observed with respect to the Mining and Construction industries.

However, there are a select few industries for which the productivity's significant correlation with sunshine is more robust than the ones mentioned above. These industries include Agriculture, Forestry, Fishing and Hunting; Utilities; and Manufacturing. The economic explanation for the first two is quite clear, as both of these industries' success in a given year is dependent on the weather. In particular, the yield of a farm depends on the weather, and warmer weather (correlated with the amount of sunshine) results in substantially greater electricity use in the form of A/C and other cooling appliances. Manufacturing is an industry for which the explanation for the observed trend is not immediately evident. The argument above would suggest that industrial jobs or those dependent on agriculture like manufacturing would be correlated to sunshine only due to the confounding variable of geographic region, but this is not the case here given that our regression with (only) entity effects yields a significant $p = 0.0404$. Perhaps this p-value is early evidence of the conjectured mental effect that exposure to sunshine has on an individual laborer's productivity level. More investigation and analysis is required to arrive at the aforementioned conclusion with any certainty.

5.2 Fixed Effect Regression with Covariates

The results of this regression are contained in Table 2. Most of the industries that had significant correlations with sunshine in Table 1 still did after adding our two additional covariates, and in all instances where significance was preserved the sign of the correlation

Dependent Variable	β	p-**value**	**EE**	**TE**
State GDP	112.27	**0.0002**	No	Yes
State GDP	21.418	0.0523	Yes	No
State GDP	11.921	0.3296	Yes	Yes
Agriculture, Forestry, Fishing and Hunting	2.1279	**0.0000**	No	Yes
Agriculture, Forestry, Fishing and Hunting	-1.3831	**0.0000**	Yes	No
Agriculture, Forestry, Fishing and Hunting	-1.6701	**0.0000**	Yes	Yes
Mining	10.817	**0.0000**	No	Yes
Mining	0.6654	0.5245	Yes	No
Mining	0.9741	0.4609	Yes	Yes
Utilities	2.0617	**0.0000**	No	Yes
Utilities	0.5786	**0.0120**	Yes	No
Utilities	0.4829	**0.0264**	Yes	Yes
Construction	8.3369	**0.0000**	No	Yes
Construction	1.6351	0.2648	Yes	No
Construction	-1.4353	0.3823	Yes	Yes
Manufacturing	9.9982	**0.0084**	No	Yes
Manufacturing	5.4828	**0.0404**	Yes	No
Manufacturing	3.4485	0.2882	Yes	Yes
Retail Trade	9.0335	**0.0000**	No	Yes
Retail Trade	0.8780	0.1563	Yes	No
Retail Trade	-0.6638	0.3627	Yes	Yes
Finance and Insurance	-3.0191	0.2578	No	Yes
Finance and Insurance	0.4405	0.8066	Yes	No
Finance and Insurance	-0.5699	0.7955	Yes	Yes
Real Estate and Rental and Leasing	14.735	**0.0004**	No	Yes
Real Estate and Rental and Leasing	1.9987	0.3633	Yes	No
Real Estate and Rental and Leasing	1.7779	0.4579	Yes	Yes
Health Care and Social Assistance	3.4441	0.0831	No	Yes
Health Care and Social Assistance	0.0518	0.9634	Yes	No
Health Care and Social Assistance	1.5054	0.1754	Yes	Yes
Arts, Entertainment, and Recreation	1.6427	**0.0001**	No	Yes
Arts, Entertainment, and Recreation	0.1730	0.4431	Yes	No
Arts, Entertainment, and Recreation	-0.0018	0.9947	Yes	Yes
Computer Systems Design and Related Services	1.2868	**0.0030**	No	Yes
Computer Systems Design and Related Services	0.4257	0.5589	Yes	No
Computer Systems Design and Related Services	0.0602	0.9406	Yes	Yes
Professional, Scientific, and Technical Services	7.1825	**0.0043**	No	Yes
Professional, Scientific, and Technical Services	2.0646	0.1497	Yes	No
Professional, Scientific, and Technical Services	1.8447	0.2525	Yes	Yes

Table 1: Fixed Effects without Covariates

was the same, indicating that controlling for these factors does not change the inference explained in the earlier section for these particular industries. In particular, it is noteworthy that the utilities industry still has a significant positive correlation with sunlight exposure even after projecting out state-level effects.

There are some dependent variables whose p-values increased beyond the significance threshold after adding these additional controls. The first of these is the all-industry GDP of the state, suggesting that the earlier positive correlations we had observed between a state's overall productivity and its amount of Sun exposure was occluded by some combination of agricultural industry strength and damages faced due to natural disasters. We claim the same for the manufacturing industry, whose p-values are now inflated to be at least 0.1 regardless of the fixed effects used. The real estate and entertainment industries also saw losses of significance when compared to the uncontrolled regression, but the original significance for both of these only existed when entity effects were excluded, so these results were not of interest given we are not curious about using Sun as an indicator for certain U.S. states or regions. Lastly, we can of course no longer perform the regression that had agricultural industry output as the dependent variable, as we are now controlling for it as a covariate.

The only industry of note that became newly significant after controlling for our covariates is Finance and Insurance, and this one is of interest given its only significant correlation with sunshine is negative (with time effects enabled but entity effects disabled). We should therefore note that there is no significant relationship after controlling for geographic region via state-level effects. Despite this fact, it is interesting to observe the phenomenon, as it may be evidence of the claim put forth by Lee et al. (2014), which was that better weather conditions can cause greater distractions particularly in workers with routine desk jobs, and thereby lead to a decrease in their productivity.

Finally, we would like to make note of the fact that the Retail Trade industry (while significant originally with only temporal fixed effects) gained new significance when controlling for region via state fixed effects. This is substantial, as it means that sunlight is not being used as a mere indicator for certain states which happen to also have strong retail industries, but rather that sunlight has a deeper–potentially causal–effect on the industry's success. We interpret this as owing to the fact that sunlight can influence consumer behavior, attracting people to engage in outdoor activities and potentially increasing foot traffic in retail areas–thereby leading to higher sales and economic output.

5.3 IV Regression

The results from our IV regression (with covariates) are provided in Table 3. With our cloud cover instrument, we find many significant estimates for β, the coefficient on the solar irradiance. We will analyze a few of these which are noteworthy. For Agriculture, Forestry, Fishing and Hunting, we find a positive coefficient ($\beta = 1.4188$) with extreme significance ($p = 0.0000$). As discussed earlier, it is tenable to claim that more sunlight provides better conditions for crop growth, forestry operations, and hunting, which leads to the higher productivity and economic output of this sector on average. Given that this strong correlation stands after using an exogenous instrument, the intuition that this relationship is causal and un-confounded is confirmed.

The only other industry with significant regression results and $\beta > 0$ is Mining. For the mining industry, we found $\beta = 8.7383$ and $p = 0.0000$. An intuitive explanation for how sunlight may cause improved efficiency for this sector is via the better working conditions fair weather provides, as mining is a job that is done outdoors and in the elements, and therefore the ability of workers in this industry to complete certain tasks depends on the weather being sufficiently good to attempt said tasks safely. This is confirmation of the fact that mining benefits from additional Sun exposure, which was not clear when we were using entity level effects to control for geography correlations due to the loss of statistical power that this decision induced. With our region-based controls, which increases statistical power by grouping states into regional bins, we are able to maintain the positive coefficient that

Dependent Variable	β	p-**value**	**EE**	**TE**
State GDP	-0.1695	0.9936	No	Yes
State GDP	15.159	0.1819	Yes	No
State GDP	-1.4785	0.9046	Yes	Yes
Mining	8.9291	**0.0000**	No	Yes
Mining	-0.9060	0.3821	Yes	No
Mining	-1.1096	0.3899	Yes	Yes
Utilities	0.4732	0.1751	No	Yes
Utilities	0.4789	**0.0437**	Yes	No
Utilities	0.3023	0.1708	Yes	Yes
Construction	3.2961	**0.0005**	No	Yes
Construction	2.1730	0.1506	Yes	No
Construction	-0.5750	0.7318	Yes	Yes
Manufacturing	-4.0341	0.1307	No	Yes
Manufacturing	4.3586	0.1131	Yes	No
Manufacturing	1.1641	0.7249	Yes	Yes
Retail Trade	1.8058	0.1353	No	Yes
Retail Trade	1.2746	**0.0454**	Yes	No
Retail Trade	-0.3844	0.6082	Yes	Yes
Finance and Insurance	-9.2685	**0.0002**	No	Yes
Finance and Insurance	0.7137	0.6995	Yes	No
Finance and Insurance	-0.7811	0.7289	Yes	Yes
Real Estate and Rental and Leasing	-0.1097	0.9701	No	Yes
Real Estate and Rental and Leasing	0.9725	0.6683	Yes	No
Real Estate and Rental and Leasing	-0.2645	0.9136	Yes	Yes
Professional, Scientific, and Technical Services	-1.5034	0.4202	No	Yes
Professional, Scientific, and Technical Services	0.7959	0.5872	Yes	No
Professional, Scientific, and Technical Services	-0.1171	0.9423	Yes	Yes
Health Care and Social Assistance	-0.1097	0.9701	No	Yes
Health Care and Social Assistance	-0.2565	0.8259	Yes	No
Health Care and Social Assistance	0.5526	0.6241	Yes	Yes
Arts, Entertainment, and Recreation	0.0671	0.7733	Yes	No
Arts, Entertainment, and Recreation	0.1245	0.6652	No	Yes
Arts, Entertainment, and Recreation	-0.1824	0.5195	Yes	Yes

Table 2: Fixed Effects with Covariate Controls

Dependent Variable	β	p**-value**
State GDP	-47.798	0.1142
Agriculture, Forestry, Fishing and Hunting	1.4188	**0.0000**
Mining	8.7383	**0.0000**
Utilities	-0.0969	0.8488
Construction	1.5568	0.2490
Manufacturing	-9.6076	**0.0118**
Retail Trade	-1.0504	0.5421
Finance and Insurance	-13.708	**0.0003**
Professional, Scientific, and Technical Services	-5.0639	**0.0333**
Real Estate and Rental and Leasing	-5.7603	0.1392
Health Care and Social Assistance	-7.3216	**0.0006**
Arts, Entertainment, and Recreation	-0.4419	0.2181

Table 3: IV Regression with Covariates

was seen on mining in Table 2 when no fixed effects were enabled.

In this IV regression, we find many significant regression results with $\beta < 0$. Specifically, we find a significant negative causal effect of sunlight on the industries of Manufacturing; Finance and Trade; Professional, Scientific, and Technical Services; and Health Care and Social Assistance. We note that, because we control for the regional indicator in this regression, this result cannot be because of drastic geographic region based correlations. Instead, this seems to speak in favor of the hypothesis of Lee et al. (2014), who believed that better weather had a negative, distracting effect, especially on white-collar workers.

5.4 Regression with Quarterly Data

We acknowledge the potential limitation that our insufficient sample size may occlude statistically meaningful results. With an annual time period of 11 years (2001-2012) and 49 continental states, we have a total of 539 data points. However, one must not neglect to consider our entity and time effects, which take our model up to having well over 50 covariates. Given the general rule of thumb that one should have more than 10 observations per variable, we can see that our data sample size of 539 might be costing us the statistical power necessary to find additional statistically significant results.

Dependent Variable	β	p**-value**	**EE**	**TE**
State GDP (Millions)	5.2562	0.0575	No	No
State GDP (Millions)	75.184	0.0000	No	Yes
State GDP (Millions)	0.3775	0.1735	Yes	No

Table 4: Fixed Effects Regression using Quarterly Periods

To address this problem, we utilize a finer time scale of quarterly data instead of annual data. The quarterly data comes from a narrower time period of 2005 to 2012, resulting in a total of 1536 data points, which is a three-fold increase in the sample size.

We also might think that we would be able to pick up a causal relationship between worker productivity and sunshine levels on a finer time scale period such as a quarterly time period, which we might miss for a coarser annual time period. A coarser time scale may fail to capture shorter term changes in sunshine levels and the impact on worker productivity–such as the possibility of seasonal differences in the effects of sunshine levels on worker productivity–by averaging out such finer changes in a yearly time period analysis.

5.5 Individual Productivity (Wage Share of GDP)

In our previous analysis, we have been using GDP as the dependent variable in our regressions as a proxy for the productivity of the United States (or, potentially, a specific industry). While GDP is a logical metric for overall productivity, one might posit that it is not a good metric for the average individual worker's productivity, as many contributors to GDP (e.g., investments) are orthogonal to individual laborers' productivity levels.

To alleviate this issue, we consider a different metric for our dependent variable: the labor share of GDP. To clarify, this quantity differs from the BLS definition of labor share, measured as $\frac{\text{Employee compensation + Proprietors' labor compensation}}{\text{Output}}$; we define it to be $\frac{\text{total wages paid to laborers}}{\text{GDP}}$. Under the assumption that the labor market is competitive, this metric essentially indicates what percentage of a state's productivity in a given year is attributable to the productivity of its laborers, as in a competitive labor market the workers are compensated in a quantity that is precisely equal to the monetary amount of value they generate. Therefore, this can

Dependent Variable	β	p**-value**	**EE**	**TE**
State Labor Share	-1.817e-06	0.0005	No	No
State Labor Share	-1.161e-06	0.0000	No	Yes
State Labor Share	-3.576e-07	0.1348	Yes	No
State Labor Share	-8.993e-08	0.8508	Yes	Yes

Table 5: Fixed Effects Regression with Labor Share Outcome

be seen as a reflection of average individual-level productivity, because it excludes all factors other than the quality and quantity of work output by employees as reflected by their pay.

Examining the regression results presented in Table 5, we see that, when controlling for state-level effects, there is no significant correlation between an individual's productivity level (as measured by our metric) and the amount of irradiance from the Sun. However, when regions are not controlled for via EEs, we see that, irrespective of whether or not we control for national-level shocks, there is a significant estimate of $\beta \approx 0$. Thus, we find zero correlation between the amount of sunshine exposure in a given state and the productivity of said state's labor force as measured by their laborer's share of GDP. It is worth noting that this may be more of a reflection of the inelasticity in the labor markets, as it is possible that our assumption of competition is too strong to be realistic, and that generally laborers in a given state tend to take home the same relative amount of pay irrespective of their overall performance level.

5.6 Regional Grouping of Entities

Given the size of our sample data, we might posit that having 49 entity effects may be too many and can be leading to a reduction in the statistical significance we find for the coefficient on the sunshine covariate. To alleviate this problem, we instead consider a one-hot encoded variable representing one of the eight BEA designated US regions: New England, Mideast, Great Lakes, Plains, Southeast, Southwest, Rocky Mountain, and Far West.

Table 6 shows the results of the regression of labor share of GDP on sunshine irradiance levels. In this regression, we do not include covariate controls of natural disaster costs and US agriculture GDP, as we were not able to obtain quarterly-precision data for these quantities. Unfortunately, none of these regression coefficients are significant, so even our weakening of state-level entity effects to regional-level entity effects was not enough to garner enough statistical power to recover significant correlations.

Table 7 replicates Table 2 with the addition of the regional control covariate. To be precise, our model is a fixed effects model with natural disaster and agricultural GDP controls in addition to the regional control covariate as a substitute for entity effects.

With a smaller model (a reduction from forty-nine state variables to eight regional variables), sunshine levels become a statistically significant predictor for some regressions. For example, even with time effects, we see that sunshine levels has a statistically significant positive relationship with the GDP of the retail trade industry, whereas in the fixed effects regression with both time effects and entity effects, we are not able to discern a statistically significant relationship between sunshine levels and the GDP of retail trade.

One might posit that regional based controls are not sufficient and that state level effects are causing the presumed positive relationship between sunshine levels and retail trade. For example, the three states with the biggest retail trade industry GDP's are precisely California, Texas, and Florida, which are in geographic regions of Far West, Southwest, and Southeast specifically, but there are many states that are in those regions (e.g., New Mexico, Nevada, Utah in the Far West region) which can bias the regional entity variable (downwards in the case of California) and thus not sufficiently account for such state specific effects.

However, we might consider that regional based controls might be sufficient for industries such as Mining, which geographically are clustered in certain regions, namely the Southwest, Rocky Mountain, and Far West. For context, the states with the highest mining GDP's are precisely Texas, Louisiana, Oklahoma, California, Wyoming, and Colorado.

The interpretation for a negative relationship between sunshine levels and the mining industry GDP is not clear. However, a plausible explanation could be that higher sunshine levels can lead to higher operational costs in mining projects due to the increased costs

Dependent Variable	β	p-value	EE	TE
State Labor Share	-2.625e-07	0.5741	No	No
State Labor Share	9.459e-07	0.5174	No	Yes

Table 6: Fixed Effects Regression with Regional Control

in cooling/temperature management, which would lead to higher energy consumption and operational costs, decreasing the profitability of mining projects and thus decreasing the total GDP. Another interpretation could be that, given we had found significant positive coefficients on Mining with different experimental approaches in the past, it is simply the case that region-based grouping results in fixed effects that do not fully capture the heterogeneity between states in the same area.

6 Discussion and Conclusion

The findings from our empirical investigation provide valuable insights into the marginal effect of additional exposure to sunlight on the productivity of the U.S. economy. Through our analysis, we see that the correlation between industry productivity and sunshine exposure is extremely heterogeneous, and that even the sign of the effect depends on what industry is being considered. Our results using state labor share as a reflection of individual productivity showed either insignificant effect or zero correlation with sunshine. Because we find significant effects at the aggregate industry levels, we interpret the latter findings as suggesting that the metric we used as a proxy for an individual laborer's productivity was ineffective. We believe this could likely be due to wages being more inelastic than we had assumed, since we relied on the notion that wages in the labor market were competitively priced.

In particular, our findings suggest that industries such as agriculture, mining, and utilities have a positive relationship with sunlight due to a direct dependence of these industries' business operations on the weather, as explained in Section 5. However, some industries like finance and research exhibited the inverse relationship, potentially due to the distraction

Dependent Variable	β	p**-value**	**EE**	**TE**
State GDP	20.942	0.3810	No	No
State GDP	15.849	0.5208	No	Yes
Agriculture, Forestry, Fishing and Hunting	0.0435	0.4162	No	No
Agriculture, Forestry, Fishing and Hunting	0.0228	0.6777	No	Yes
Mining	-6.7076	**0.0002**	No	No
Mining	-7.1465	**0.0001**	No	Yes
Utilities	0.4147	0.2997	No	No
Utilities	0.3358	0.4124	No	Yes
Construction	3.7351	**0.0015**	No	No
Construction	3.3905	**0.0040**	No	Yes
Manufacturing	-5.2715	0.0879	No	No
Manufacturing	-6.5244	**0.0404**	No	Yes
Retail Trade	3.4012	**0.0118**	No	No
Retail Trade	3.1295	**0.0248**	No	Yes
Finance and Insurance	-3.2339	0.3043	No	No
Finance and Insurance	-3.7367	0.2519	No	Yes
Real Estate and Rental and Leasing	9.5386	**0.0040**	No	No
Real Estate and Rental and Leasing	9.3735	**0.0060**	No	Yes
Health Care and Social Assistance	-0.5454	0.7544	No	No
Health Care and Social Assistance	-0.6974	0.6971	No	Yes
Arts, Entertainment, and Recreation	1.8303	**0.0000**	No	No
Arts, Entertainment, and Recreation	1.8516	**0.0000**	No	Yes
Computer Systems Design and Related Services	1.2868	**0.0030**	No	No
Computer Systems Design and Related Services	0.4257	0.5589	No	Yes
Professional, Scientific, and Technical Services	3.4170	0.1007	No	No
Professional, Scientific, and Technical Services	3.1141	0.1497	No	Yes

Table 7: Fixed Effects Regression with Covariates and Regional Control

effect that better weather can have, consistent with the findings of Lee et al. (2014).

However, it is crucial to note the endogeneity and confounding risks associated with sunlight levels and GDP. Sunlight exposure may be influenced by various factors, such as geographical location, climate patterns, and regional economic conditions, which can complicate the interpretation of our results. We have taken precautions to address the most evident potential confounding effects and have provided economic interpretations that account for these limitations. Specifically, we have experimented with many regression techniques and control variables that attempt to project out these, such as the state- and region-level fixed effects.

We perceive some potential future research that could be done in order to resolve some unanswered questions. For example, potential ablations include partitioning individuals into two blocks depending on whether they work in offices or outdoors, researching the effect of population density given tall buildings obstruct sunlight, and distinguishing the effect of direct sunlight versus sunshine diffused through windows or other materials. The latter is somewhat explored by Boubekri et al. (2014), but their research is limited to indoor exposure without a comparison to direct light.

In conclusion, our empirical investigation reveals a diverse and industry-specific relationship between sunlight exposure and productivity in the U.S. economy. We are hopeful that, despite the need to resolve further unanswered questions, these results and analyses will have important implications for policymakers and industry stakeholders aiming to harness the potential benefits of sunlight exposure on productivity.

References

An, M., S. M. Colarelli, K. O'Brien, and M. E. Boyajian (2016, May). Why we need more nature at work: Effects of natural elements and sunlight on employee mental health and work attitudes. *PloS One*.

Boubekri, M., I. Cheung, K. Reid, C. Wang, and P. Zee (2014). Impact of windows and daylight exposure on overall health and sleep quality of office workers: A case-control pilot study. *Journal of Clinical Sleep Medicine 10*(6), 603–611.

Harris, I., P. Jones, T. Osborn, and D. Lister (2014). Updated high-resolution grids of monthly climatic observations – the cru ts3.10 dataset. *International Journal of Climatology 34*(3), 623–642.

Herber, S. P., J. S. Quis, and G. Heineck (2017). Does the transition into daylight saving time affect students' performance? *Economics of Education Review 61*, 130–139.

Kahn, M. E. and P. Li (2020, jul). Air pollution lowers high skill public sector worker productivity in china. *Environmental Research Letters 15*(8), 084003.

Lee, J., F. Gino, and B. Staats (2014, 01). Rainmakers: Why bad weather means good productivity. *The Journal of applied psychology 99*.

LoPalo, M. (2023, January). Temperature, worker productivity, and adaptation: Evidence from survey data production. *American Economic Journal: Applied Economics 15*(1), 192–229.

Rosekind, M. R., K. B. Gregory, M. M. Mallis, S. L. Brandt, B. Seal, and D. Lerner (2010). The cost of poor sleep: Workplace productivity loss and associated costs. *Journal of Occupational and Environmental Medicine 52*(1), 91–98.