Climate Factors and Gestational Age

Author: Dheya Madhani *(dheyam@gmail.com)*

Advisor: Dr. Andrii Babii

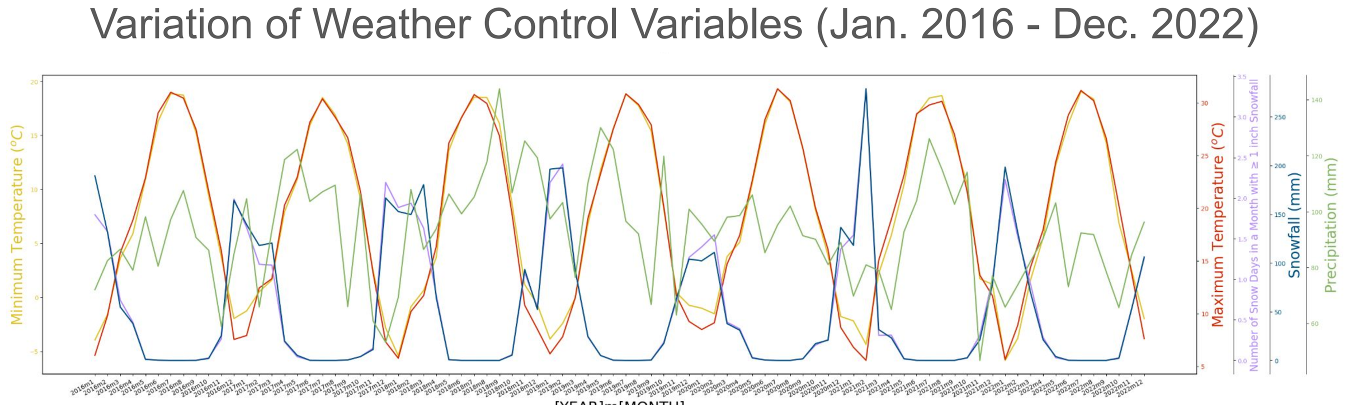
Abstract

Climate change and weather-related variables like snowfall, temperature, rainfall, and humidity have extensively been studied and used to predict health outcomes, labor outcomes, and agricultural outcomes. However, several models do not control for all weather variables. Moreover, the weather dataset is a multi-dimensional panel dataset where the dimensions include time, location, and weather variables like temperature, precipitation, rainfall, and snowfall. Factor modeling is an approach that addresses the high dimensionality of the dataset and captures its essential information. Traditionally, factor modeling is used in macroeconomics and finance analysis. Thus, this thesis aims to study the causal effects of two key weather variables on gestational age. Moreover, it aims to control for several weather variables using a novel method called tensor principal component analysis (PCA) to avoid omitted variable bias. As a result, existing studies could improve predictions of other health, agriculture, and labor outcomes affected by climate through this method and policy-makers can effectively allocate healthcare resources based on accurate predictions.

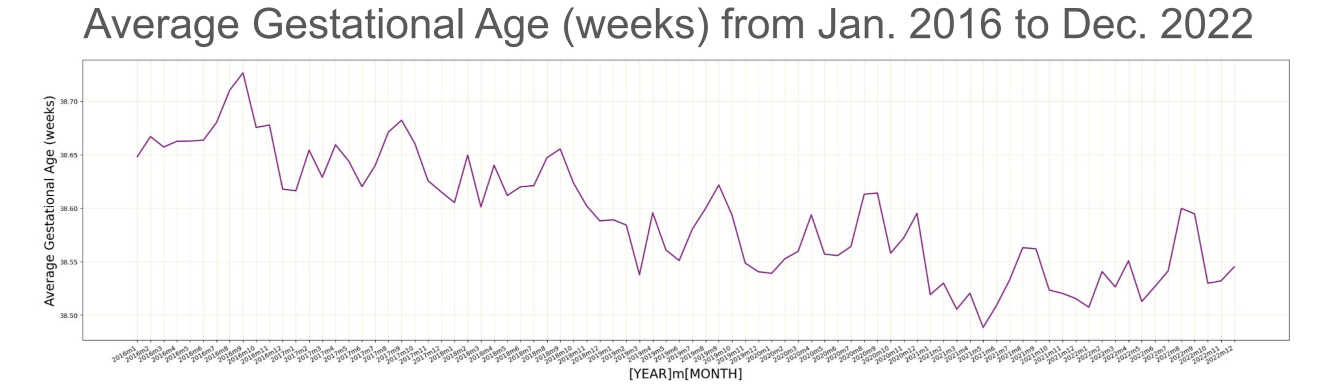
Data

The Global Summary of the Month (GSOM), Version 1 dataset is used as it collects 40+ weather metrics from stations worldwide since 1763. The station-level data is aggregated to the county-level by averaging the five closest stations that are less than 100 km from the county's centroid for 2016 and the years between 2019-2022. This process is performed separately for the snowfall related variables and the temperature and precipitation related variables to gain more observations. The Natality Dataset collected through the Vital Statistics Cooperative Program is also used as it records information about live births in the United States from 2016-2022 for counties with at least 100,000 people.

Variation of Weather Control Variables (Jan. 2016 - Dec. 2022)



Average Gestational Age (weeks) from Jan. 2016 to Dec. 2022



Theoretical Framework

The model is built using the Becker-Grossman style 1-period model of health production. Z_1 represents private goods consumed that do not affect the mother's health while Z_2 focuses on private goods consumed for the mother's health. Solving the utility maximization problem shows that the mother's health affects the birth outcome, which highlights the need to consider it when studying the effect of climate on the infant's health. Additionally, the mother's health is a function of climate. Extreme weather events or outcomes of climate change like heat waves, wildfires, drought, hurricanes, and floods affect the biological health and physiological health through oxidative stress, malnutrition, inflammatory responses, infection, and endocrine disruptors. They also impact the physical environment through air pollution, pathogen spread, food scarcity, and water contamination, which affects the biological and physiological health. They also result in food and water insecurity, conflicts because of resource scarcity, and displacement, which leads to stress and mental health issues. These health reactions have a direct effect on gestational complications and preterm birth (Ha, 2022).

$$U = U(M, B, Z)$$
$$M = H_1(C, Z_2)$$
$$B = H_2(M)$$
$$I - Z_1 - pZ_2 = 0$$

M: Mother's Health; **B**: Child's Health; **C**: Climate; **Z**: Jointly Aggregated Consumption of Goods & Services; **I**: Income; Price of $\mathbf{Z_1}$ relative to $\mathbf{Z_2}$ is 1

Coefficient of Interest:

$$\frac{dB}{dC} = \frac{dH_2(M)}{dC} = \frac{dH_2(H_1(C,Z_2))}{dC} = \frac{\partial H_2}{\partial H_1} * \frac{\partial H_1}{\partial C}$$

Empirical Model

$$X_{c,t,m,v} = \sum_{i=1}^{R} \lambda_{c,i} f_{t,i} g_{m,i} \mu_{v,i} + \epsilon_{c,t,m,v}$$

c: county; **t**: year; **m**: month; **v**: 10 weather control variables

$\lambda_{c,i}$: county loadings (exposure of counties to common climate factors)

$f_{t,i}$: annual factor (latent vector driving long-term changes in climate)

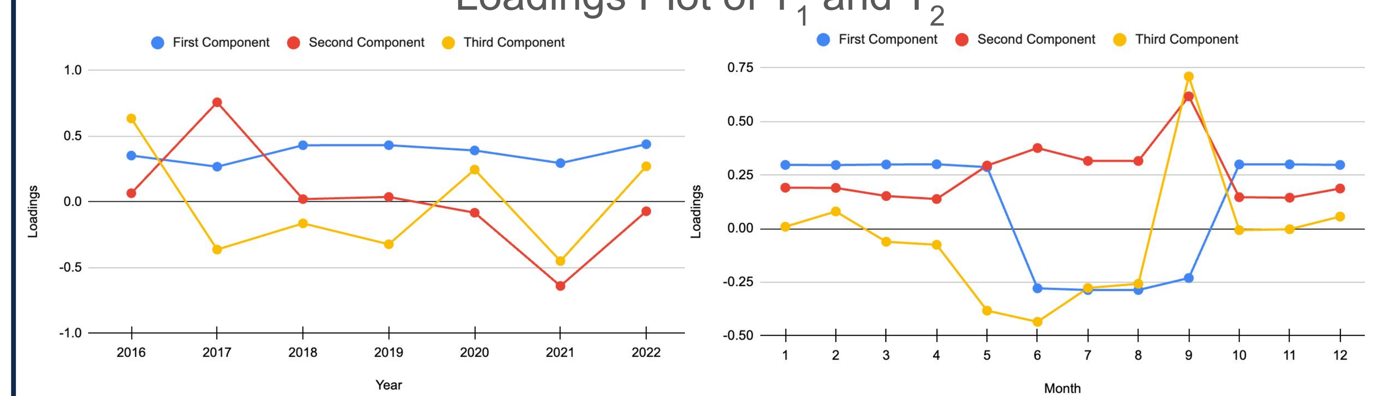
$g_{m,i}$: monthly factor (drives seasonal changes in weather)

$\mu_{v,i}$: weather loadings (exposure of each weather variable to factors)

A multi-way fixed effects model is used to estimate the causal effects of climate on gestational age using a panel dataset by including state fixed effects, year fixed effects, and month fixed effects. There are 10 control weather variables ranging from average monthly temperature to total monthly snowfall and 4 pregnancy control variables including average age of mother. Tensor PCA is performed on the 10 weather control variables by first unfolding the 4-dimensional tensor along each of its dimensions to obtain 4 matrices (year (T_1), month (T_2), weather variables (T_3) and counties (T_4)). For each of the 4 matrices, PCA is performed to get the R eigenvectors (corresponding to the R largest eigenvalues) of the sample covariance matrix. These represent the factors or loadings for the dimension it was unfolded along. The value of R (3) is determined using the elbow of the 4 scree plots corresponding to the 4 matrices.

Results

Loadings Plot of T_1 and T_2

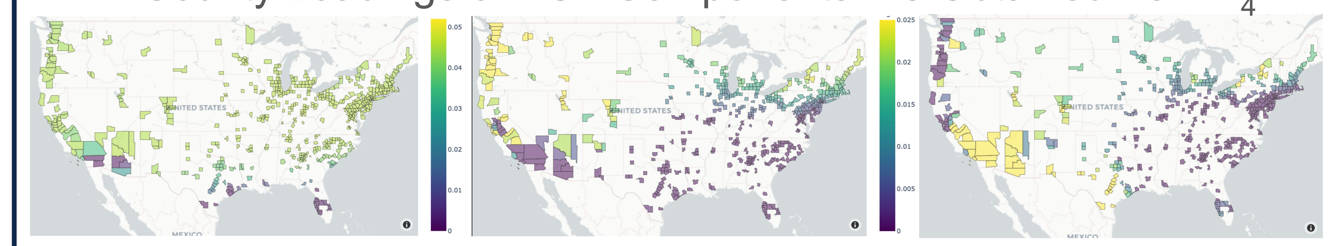


Loadings of 10 Control Weather Variables from T_3

	First Eigenvector	Second Eigenvector	Third Eigenvector
Average Monthly Temperature (°C)	0.2485	0.4100	-0.0093
Monthly Mean Maximum Temperature (°C)	0.3391	0.2842	-0.0554
Monthly Mean Minimum Temperature (°C)	0.0069	0.5169	0.0780
Total Monthly Cooling Degree Days (°F)	0.1041	0.4842	0.5564
Total Monthly Heating Degree Days (°F)	0.3719	-0.2045	-0.0720
Days with Minimum Temperature ≤ 32°F (Days)	0.3675	-0.2165	-0.0224
Total Precipitation in the Month (mm)	0.3987	0.0711	-0.3359
Days with Precipitation ≥ 1 inch (Days)	0.3819	0.1401	-0.4291
Total Monthly Snowfall (mm)	0.3379	-0.2635	0.4805
Days with ≥ 1 inch Snowfall (Days)	0.3464	-0.2581	0.3842

The estimated multi-way fixed effects model uses tensor PCA for the 10 control weather variables. The number of the days in a month where the maximum temperature is at least 90°F and number of days in a month where the precipitation is at least 0.1 inch are the key variables of interest, and they have a statistically significant causal effect on gestational age. The coefficients are approximately similar to the baseline multi-way fixed effects models but the use of tensor PCA instead of PCA leads to a drastic decrease in robust standard errors. With tensor PCA, only 1,662 parameters are estimated instead of 132,300 leading to more precise estimates and reduced risk of overfitting the data.

County Loadings of PCA Components 1-3 Obtained from T_4



Multi-Way Fixed Effects Model with Tensor PCA

Variables	(1) Average Gestational Age (weeks)	(2) Average Gestational Age (weeks)
Days with Maximum Temperature ≥ 90°F (Days)	-0.0018***	-
	(0.0001)	
Days with Precipitation ≥ 0.1 inch (Days)	-	0.0014***
		(0.0001)
Tensor	-0.4447**	-0.7270***
	(0.2134)	(0.2130)
Average Age of Mother (Years)	0.0073***	0.0074***
	(0.0005)	(0.0005)
Average Pre-pregnancy BMI	-0.0284***	-0.0287***
	(0.0006)	(0.0006)
Average Number of Prenatal Visits	0.0293***	0.0296***
	(0.0004)	(0.0004)
Average Interval Since Last Live Birth (Months)	-0.0078***	-0.0079***
	(0.0001)	(0.0001)
Constant	39.2218***	39.2092***
	(0.0278)	(0.0278)
Observations	441,000	441,000
R-squared	0.3271	0.3267
State Fixed Effects	Yes	Yes
Year Fixed Effects	Yes	Yes
Month Fixed Effects	Yes	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Conclusion

After controlling for pregnancy-related variables and 10 weather variables using tensor PCA, it was found that a one day increase in the number of days in a month with the maximum temperature of at least 90°F causes a decrease in average gestational age by 0.0018 weeks. This is due to the changed heat loss capacity, increase in hypothalamic-pituitary-adrenal axis' hormones, and dehydration. On the other hand, increasing the number of days in a month that experience precipitation of at least 0.1 inch by a day causes an increase in gestational age by 0.0014 weeks. This is because water scarcity and poor sanitation affects the absorption of nutrients for the mother and may also cause anemia. These results are the main empirical contribution to existing research, as it controls for several weather variables when estimating the causal effects. Moreover, because many weather variables are considered in the models, a comprehensive weather dataset is used which differs from existing studies. Also, different dimensionality reduction methods are analyzed to incorporate the 10 weather control variables. Tensor PCA provides estimates that are more precise and effectively controls for 10 weather variables with only one tensor, thereby overcoming the omitted variable bias present in several existing studies.