Final project

2023-10-29

Part 2

Study period: Feb 2020-Oct 2020

Mobility: The data were compared for COVID phase (ie, March 14 - October 15 2020) with pre-COVID phase (ie, January 3-February 6, 2020).

The mathematical model (multilevel form)

$$\textit{For time } t \textit{ in site } i$$

$$PM2.5_{ti} = \pi_{0i} + \pi_{1i}date_{ti} + \pi_{2i}month_{ti} + \pi_{3i}workplace_{ti} + \epsilon_{ti}$$

$$\epsilon_{ti} \sim N(0, \sigma^2)$$

$$\begin{aligned}
\pi_{0i} &= \gamma_{00} + \gamma_{01}HDI_i + \gamma_{02}Population_i + u_{0i} \\
\pi_{1i} &= \gamma_{10} + u_{1i} \\
\pi_{2i} &= \gamma_{20} + \gamma_{21}hemisphere_i \\
\pi_{3i} &= \gamma_{30} + \gamma_{31}HDI_i
\end{aligned}$$

$$\begin{bmatrix} u_{0i} \\ u_{1i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \\ \tau_{10} & \tau_{11} \end{bmatrix} \right)$$

At level 1 (time-varying) covariates: time, measured in days
$mobility_{ti}$: mobility data for each day
$month_{ti}$: month of of the day
$workplace_{ti}$: workplace percentage change for each day

At level 2 (site specific invariant covariates):
HDI_i: HDI for each site
$population_i$: population density for each site

- $time_{ti}$: time is the days at measurement t for country i
- $mobility_{ti}$: % change in mobility compared to the median value of the day of the week during baseline days before COVID (5 weeks, from January 3 to February 6, 2020), at measurement t for country i
- π_{0i}: a country's baseline PM2.5 level, i.e., # PM2.5 level at $time_{ti} = 0$ and $mobility_{ti} = 0$
- π_{1i}: an country's PM2.5 change rate per unit of time, i.e., expected # PM2.5 change per day
- π_{2i}: an country's PM2.5 change rate per unit of mobility, i.e., expected # PM2.5 change per unit change in mobility
- γ_{00}: the "grand intercept" - mean baseline PM2.5 for all countries (when percent change in mobility and time since February 6, 2020 are 0)
- γ_{01}: the slope for GDP - mean PM2.5 change rate for each unit increase in GDP for all countries

- γ_{02}: the slope for population density - mean PM2.5 change rate for each unit increase in population density for all countries
- γ_{10}: the "grand slope" - mean PM2.5 change rate per day for all countries
- τ_{00}: extent of country variation around the mean baseline PM2.5 level (variance of π_{0i})
- τ_{11}: extent of country variation around the mean PM2.5 change rate/curve (variance of π_{1i})
- τ_{22}: extent of country variation around the mean mobility vs. PM2.5 change rate/curve (variance of π_{2i})
- $\rho = \frac{\tau_{10}}{\sqrt{\tau_{00}\tau_{11}}}$: correlation between initial status (intercepts u_{0i}) and rate of change (slopes u_{1i})
- ϵ_{ti}: residual for measurement of PM2.5 at time t for country i - how far a particular measurement is different from the model's prediction
- σ^2: variance of all of the residuals (how much of the data is left unexplained after our model)

Mathematical model (reduced form or lmer command)

Lmer Model Formula

model <- lmer(PM2.5 ~ Change_in_Mobility + Time + GDP + Population_Density + (1 + Time | Country), data = dat)

We included a random intercept for country because the baseline PM2.5 level is different for each country. We include a random slope for mobility because we think that the relationship between mobility change and PM2.5 trends will be different for different countries. We also included a random slope for time due to the following: 1. Different countries may experience changes in PM2.5 levels differently over time 2. Including time as a random slope can help capture seasonality by allowing countries to have their own unique seasonal patterns. 3. Some countries may be experiencing increasing PM2.5 level over the long-term, while others maybe experiencing decreasing PM2.5 trends. Allowing each country to have its own trend in PM2.5 levels over time is a way to account for the heterogeneity between countries in terms of how they respond to time. ***Marshae: Need to reference GPT here, and this part can be shortened***

Short description of how you may fit different versions of your model (e.g., adding interaction terms) to answer extensions to your primary research questions

Interaction term: mobility*GDP

Justification of using MLM or cluster-robust approaches (depending on what you are doing). Is MLM needed for your project? Or do you just like to use it?

In this study, we intend to look at how mobility (as a proxy of how strict COVID lock down policy is) might influence PM2.5 levels over time. Our study is a time-series study with repeated measures of the PM2.5 and mobility change from Feb to Oct 2020. Therefore, a growth model is needed for our project. The measurements of mobility and PM2.5 are clustered within countries. Therefor, MLM is also needed for our project.