Causal inference on temperature-mortality relationship using time-series data: comparing the distributed lag nonlinear model and Rubin causal model

Jaemin Oh[1] and Yeonseung Chung[2]

[1,2]Department of Mathematical Sciences, KAIST, Daejeon, South Korea

October 31, 2022

Contents

Abstract

Temperature-mortality relationship has been analyzed by using regional time series data under the DLNM framework. There has been many concerns about causal interpretation on the temperature-mortality relationship, because of unmeasured confounders, model misspecification, and mixing of the design stage and the analysis stage. In this paper, we used the potential outcome framework to deal with the last two issues, and obtained the consistent result compared to the previous studies. This work shines a light on the possibility of causal interpretation on temperature-mortality relationship analyzed so far.

1 Introduction

Due to global warming and climate change, analyzing the effect of the ambient temperature on human health is an important research topic [5, 10]. Usually, the relationship has been analyzed by using regional time series data inspite of obstacles such as temporal trend of death counts and temperature, and the existence of delayed effect. These difficulties have been addressed by the distributed lag nonlinear model (DLNM) framework [3] producing exposure-response surface $\mu : (w, l) \mapsto \mu(w, l) \in \mathbb{R}$ where w is the ambient temperature and l is a time lag. In general, regression methods containing the DLNM framework have some advantages: identification of lagged effects and low dimensional summary of estimated nonlinear exposure-response curve. These advantages make regression approach popular in environmental epidemiology, especially for the topic of temperature-mortality relationship.

However, there has been a concern about causal interpretation on the results obtained by regression analysis. In recent debate on air pollution study [6] that uses similar tools to analyze time-series data, two aspects were pointed out: mixing of design stage & analysis stage, and model selection problem. Rubin said, in observational study, design stage that adjusts confounding bias and analyis stage that relates treatments and outcomes should be separated to approximate the gold standard of causal inference, the randomized experiment [15]. However, regression methods, e.g., the DLNM framework mixes two stages by fitting additional spline for each year to adjust for temporal confounding. Moreover, the DLNM framework is susceptible to model selection problem [4] since we don't know the exact placement of knots and the exact degrees of freedom. Therefore, we have to solve these issues first to make causal interpretation possible, together with collecting more data to remove the existence of unmeasured confounders.

As a solution, we suggests to use Rubin causal model (RCM) [7] known as the potential

outcome framework. It separates the design stage and the analysis stage, and does not need to select any parametric model for outcome generating process. RCM was first introduced to analyze the data of randomized experiments [14], but now widely used in observational studies [17], and even in time series data [1]. In this paper, we used RCM to estimate the log of relative risk (logRR) curve of the ambient temperature and compared the result to the DLNM framework.

2 Dataset

The data is composed of daily mean temperature in the first decimal place, and all-cause mortality count during the period from 1997 to 2018 across 36 regions in South Korea. Figure 1 presents daily mean temperature and daily all-cause deaths in Seoul between 2010 and 2018. The daily mean temperature shows apparent seasonality and the peaks are increasing due to global warming. The number of deaths shows annual seasonality with an increase in cold seasons and a decrease in warm seasons. It shows long-term increasing trend also.

3 Method

For $i = 1, \ldots, N(= 36)$ and $t = 1, \ldots, T(= 8054)$, the state of i-th region at time t can be described by $(Y_{i,t}, W_{i,t}, C_{i,t})$ where Y is the all-cause death counts, W is the mean temperature, and C is the vector of the year, the month of year, the week of year, and the day of year. Hereafter, we consider a fixed region and drop the subscript i to simplify the notation.

3.1 Distributed Lag Nonlinear Model

In the DLNM framework, the temperature-mortality association is described by the equations

$$Y_t \sim quasi - Possion(\lambda_t),$$
$$\log(\lambda_t) = \alpha + cb(W_t, \ldots, W_{t-L}; \beta) + \eta(C_t; \gamma).$$

where λ_t is the mean of the Poisson distribution with overdispersion (namely, Quasi-Possion [16]), cb is the cross basis function with specified lag L, and η is spline basis. By maximizing quasi-likelihood function, we get estimates $\hat{\alpha}, \hat{\beta}$ and $\hat{\gamma}$. Here, we emphasize that including η in the model is equivalent to eliminating temporal trend in all-cause deaths, and fitting cb function is estimating temperature effect, thus the design stage & the analysis stage are mixed.

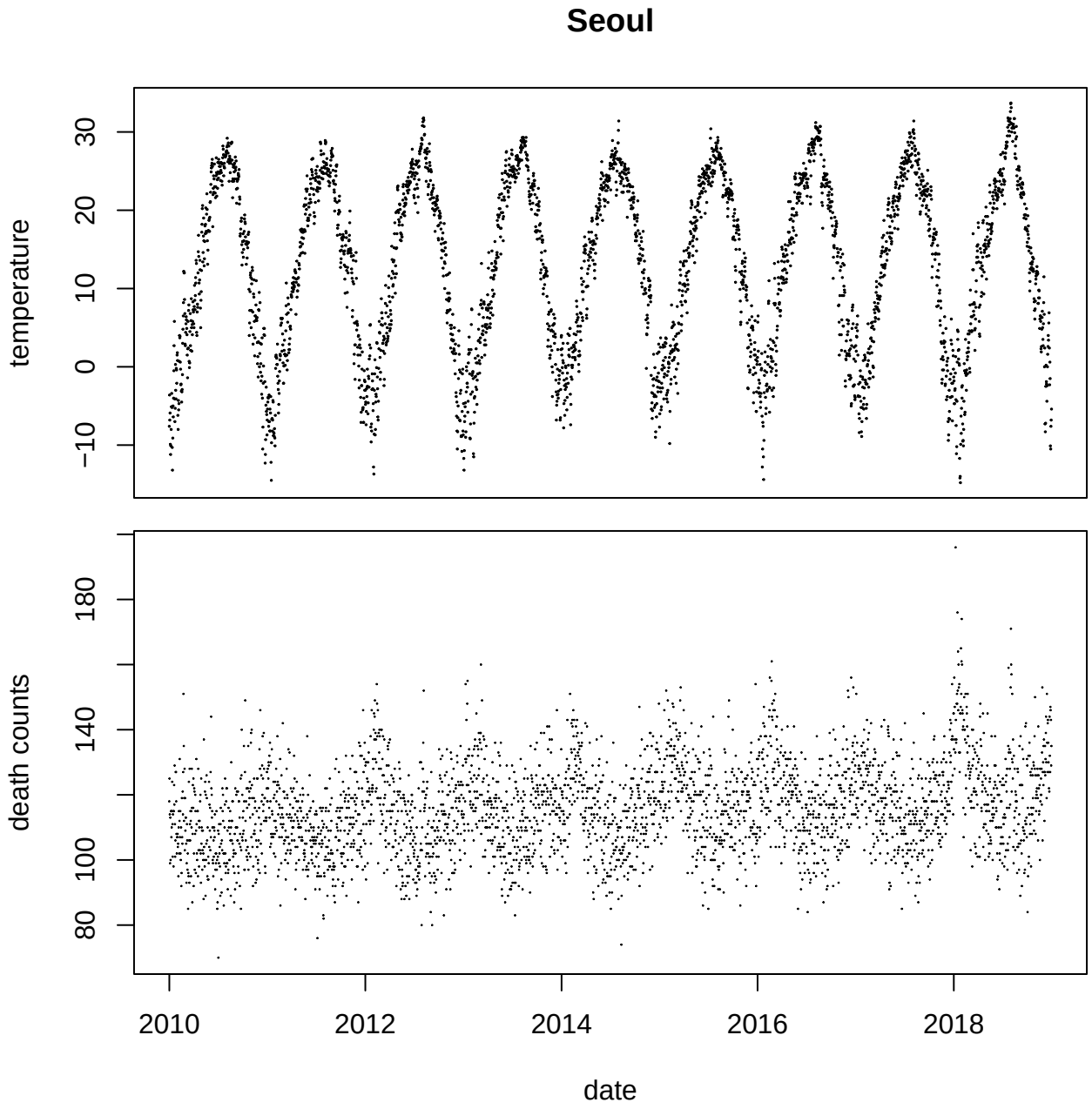

Figure 1: Daily times series of mean temperature (in Celcius) and all-cause death in Seoul from 2010 to 2018

3.2 Rubin Causal Model

Now we introduce the notation of potential outcomes. $Y_t(w)$ refers to the outcome variable at time t that would have been observed under the treatment value $W_t = w$. $Y_t(w')$ is the outcome variable that would have been obeserved by the counterfactual imagination that $W_t = w'$ had been observed instead of $W_t = w$. Observed outcome Y_t is equal to the potential outcome under observed treatment value, $Y_t(W_t)$. This is called consistency assumption 1. See the reference [12] for more detailed explanation about the definition.

We might be interested in the individual risk ratio

$$\frac{Y_t(w)}{Y_t(w')}$$

if we know true values of $Y_t(w)$ and $Y_t(w')$. However, we never know the true potential outcomes of unmeasured treatment, due to its counterfactual nature. This is called the fundamental problem of causal inference [7]. Instead, we concentrate on the average risk ratio

$$\frac{\mathbb{E}\left[Y_t(w)\right]}{\mathbb{E}\left[Y_t(w')\right]}.$$

There has been many studies to estimate $\mathbb{E}[Y_t(w)]$. In marginally randomized experiment, $\mu(w) = \mathbb{E}[Y(w)]$ can be estimated from observed data [14]. In observational studies, one can estimate causal estimand $\mu(w)$ by preprocessing the data to approximate randomization e.g., inverse probability weighting, standardization, matching [13]. Note that we dropped the subscript t to indicate more general situation than time-series setting. Throughout those techniques, the fundamental assumptions that make it possible to estimate the causal estimand are below:

Assumption 1 (Consistency)

Potential outcome for observed treatment is equal to the observed outcome. That is, $Y_t(W_t) = Y_t$.

Assumption 2 (Positivity)

Discrete treatment: For all w *and* C_t, $p(w|C_t) = Pr\left(W_t = w|C_t\right) \in (0, 1)$.

Continuous treatment: For all w *and* C_t, $p(w|C_t) > 0$ *where* $p(w|C_t)$ *is a conditional density.*

Assumption 3 (Weak Unconfoundedness)

For all w, $Y_t(w) \perp W_t|C_t$.

Positivity assumption says all treatments are possible for each confounder. Weak unconfoundedness assumption, also known as "weak ignorability" or "selection on observables" in different context, says conditional on current confounders, assignment mechanism is random to potential outcomes. Under these three assumptions, the causal estimand can be calculated as

$$\begin{aligned}\mathbb{E}\left[Y_t \frac{1_{(W_t=w)}}{p(w|C_t)}\right] &= \mathbb{E}\left[\mathbb{E}\left(Y_t(w)\frac{1_{(W_t=w)}}{p(w|C_t)}|C_t\right)\right] \\ &= \mathbb{E}\left[Y_t(w)\frac{\mathbb{E}\left(1_{(W_t=w)}|C_t\right)}{p(w|C_t)}\right] \\ &= E\left[Y_t(w)\right] = \mu(w),\end{aligned}$$

where p is a mass or density function for discrete or continuous treatment respectively. Note that the first equality comes from the interated expectation formula and consistency assumption, the second equality comes from weak unconfoundedness assumption, and the third equality is due to the definition of $p(w|C_t)$. By positivity assumption, we can divide by $p(w|C_t)$. This is called "inverse probability weighting" (IPW). Thus, a natural estimator of the causal estimand is

$$\hat{\mu}(w) = \frac{1}{T}\sum_{t=1}^{T} Y_t \frac{1_{(W_t=w)}}{p(w|C_t)}.$$

Still, we need to estimate $p(w|C_t)$ since it is unknown to us in general. When the treatment is binary, $p(w|C_t)$ is called propensity score, and it is used to adjust for confounding bias [13]. Propensity score can be extended to "generalized propensity score" (GPS) for categorical or continuous treatment [9]. For binary treatment, one can estimate propensity score by fitting logit model to data. For categorical or continuous treatment, GPS can be estimated by fitting ordered probit model or boosting.

PS and GPS have two nice properties [8, 13]. The first one is balancing property, which means that conditional on the same PS (or GPS), treatment and covariates are independent. The second one is PS-unconfoundedness, which means that conditional independence of potential outcome and treatment given PS. PS-unconfoundedness is implied by balancing property and unconfoundedness. These properties are the basis of propensity score based matching methods. But in this paper, we used inverse probability weighting so these properties are not necessary to be explained more.

The main reason why we use inverse probability weighting by GPS is to achieve covariate balance. In randomized experiment, distribution of covariates are similar across each treated group. However, we are now dealing with observational data which is not randomized. So,

we generate a pseudo-population by imposing appropriate weights to each observation and look forward to the pseudo-population achieve covariate balance. One criteria for covariate balance is the absolute correlation (AC) [19]. If treatment and covariates are independent, then their correlation must be zero. So, small value of AC can be an evidence of covariate balance. Usually, AC with < 0.1 is considered as acceptible.

4 Application

We set the reference temperature by 20°C.

4.1 Distributed Lag Nonlinear Model

We fitted the DLNM to our data to obtain region specific effect estimates. For temperature dimension, we used quadratic B-spline, and placed knots at 10th, 75th, 90th quantiles. For lag dimension, we considered 21 lags, used natural B-spline, and placed 3 knots at equally spaced on log scale. For temporal trend adjustment, we fitted additional natural B-spline with 8 degrees of freedom for each year.

With multiple effect estimates from various reigions, we pooled those estimates by multivariate meta-analysis. We modeled effect estimates $\hat{\beta}_i$ from i-th region as a mixed-effect model

$$\hat{\beta}_i \sim N_m(\beta, S_i + V)$$

where N_m denotes m-dimensional multivariate normal distribution, S_i and V are with-in and between study error covariances, respectively, and β is the true aggregated effect. We used R package 'mixmeta' to estimate β and its confidence interval. See the upper left panel of figure 2.

4.2 Rubin Causal Model

We rounded daily mean temperature to integer value. The first stage is design stage to adjust for time confounding. We adjusted confounding bias by stabilized inverse probability weighting [18] that assign weights

$$q_t = \min\left\{\frac{\hat{p}(W_t)}{\hat{p}(W_t|C_t)}, 10\right\}$$

to each obsevation. Here, we don't know true probability densities, so we should use estimated values.

To estimate $p(W_t|C_t)$, we assumed that the conditional distribution of treatment given covariates is a normal distribution, i.e.

$$W_t|C_t \sim N(m(C_t), \sigma(C_t)^2)$$

where $m(C_t)$, and $\sigma(C_t)$ are some functions of C_t. Since they are unknown in general, we should estimate them. The estimated mean $\hat{m}(C_t)$ was obtained by regressing W_t on C_t with boosting, and the estimated standard deviation $\hat{\sigma}(C)$ was calculated by boosting residuals [8,19]. Hyperparameters such as depth of tree(3), shrinkage(0.1), and the number of trees(20) were determined heuristically to minimize the absolute correlation. As an estimate of $p(W_t)$, we used relative frequency as $\hat{p}(W_t)$. One may use normal density with sample mean and sample variance too, but weighting method with relative frequency achieved lower absolute correlation in this application. See the second row and third row of Table 1. We trimmed weights bigger than 10 by 10, because some untrimmed weights were too large so effect estimate was heavily dependent to those observations.

We calculated absolute correlation (AC) [19] to see whether the covariate balance is achieved (here, AC < 0.1). Let c_t be a component of C_t, then absolute correlation with weight q_t is the absolute value of Pearson correlation coefficient between c_t and W_t, regarding each observation as q_t observations with the same values. See the table 1.

	year	month	week of the year	day of the year
1	0.01405874	0.25223312	0.25127380	0.25115386
2	0.03168787	0.08356729	0.08222320	0.08186817
3	0.03033376	0.12613061	0.12417815	0.12356290

Table 1: Absolute correlation (AC) before/after adjustment by IPW. The first row is AC before adjustment; the second row is AC after adjustment, with relative frequency of temperature as marginal probability; the third row is AC after adjustment, with normal assumption on marginal distribution of temperature.

The second stage is the analysis stage that relates treatments and outcomes with weights. We estimated the causal effect by Horvitz-Thompson estimator with trimmed stabilized weights,

$$\hat{\mu}(w) = \frac{\sum_{t=1}^{T} q_t Y_t 1_{(W_t=w)}}{\sum_{t=1}^{T} q_t 1_{(W_t=w)}}.$$

To be consistent with previous studies and standardize different population size across regions, we calculated logRR curve by $\log \hat{\mu}(w) - \log \hat{\mu}(20)$ instead of risk difference. However,

there is no known result about the distribution of estimator of risk ratio. So we measured the uncertainty of logRR curve by Moving Block Bootstrapping (MBB) [11] with 2000 bootstrap samples, 20 blocks, and each block is length of 400. In the bootstrap procedure, we did not fit gps model repeatedely for each bootstrap sample since we need to re-sample the pseudo-population itself that acheives covariate balance.

To pool the estimates, we assumed $X_i = (Y_{i,t}, W_{i,t}, C_{i,t})_{t=1}^{T}$ for $i = 1, \ldots, N$ are independent and potential outcomes of one region is independent of other regions' potential outcomes and treatments.

We estimated $\hat{\mu}_i(w)$ for each region i previously. To obtain aggregated estimate, we assumed

$$\hat{\mu}_i(w) = \mu(w) + \epsilon_i + \tau,$$

where $\epsilon_i \sim N(0, s_i)$ is within study error and $\tau \sim N(0, v)$ is between study error. s_i is estimated by bootstrap, so estimation of v remains. We estimated v and pooled estimates by taking weighted average of region specific effect estimates where weight is inversely proportional to the variance $s_i + v$ of estimator,

$$\hat{\mu}(w) = \sum_{i=1}^{N} \frac{\hat{\mu}_i(w)}{s_i + \hat{v}} / \sum_{i=1}^{N} \frac{1}{s_i + \hat{v}}.$$

Precision of pooled estimator is sum of precisions of region specific estimator,

$$\frac{1}{\hat{\sigma}^2} = \sum_{i=1}^{N} \frac{1}{s_i + \hat{v}}.$$

Confidence interval is obtained by

$$\hat{\mu}(w) \pm 1.96\hat{\sigma}.$$

See the upper right panel of figure 2.

4.3 Result

In figure 2, the upper left panel is a logRR curve obtained under the DLNM framework; the upper right panel is obtained by applying potential outcome framework; the lower left panel is smoothed version of upper right pannel (kernel: Gaussian, bandwidth: 6); the lower right pannel is a logRR curve without adjusting temporal trend under the DLNM framework.

The estimated logRR curve of the lower right panel has exaggerated values at extreme temperatures, compared to the upper left panel. Since the model of the lower right panel does

not consider temporal trend, the difference between two panels comes from autocorrelation of outcome variable.

The logRR curve of the upper right panel is spiky, because we estimated it by model free method, so it heavily depends on the observations. For the most cold temperature, we can see that the confidence interval is narrow compared to the most hot temperature. This is because there is only one such observation, so uncertainty captured by bootstrap is due to the variation of effect estimate at the reference temperature.

In the lower left pannel, we applied kernel smoothing to our estimate to remove spikes. The smoothed curve and the curve of the upper left panel have similar values compared to the curve of the lower right panel. From this point of view, we may say potential outcome framework can adjust temporal confounding bias in some extent. Moreover, we don't know what is the true logRR curve, but we may insist that the logRR curve obtained from potential outcome framework is more general in the sense that it becomes similar to the curve of DLNM after kernel smoothing.

5 Discussion

In the extent of my knowledge, there has been no study analyzing the short term relationship between the ambient temperature and the all-cause mortality using the potential outcome framework. In this work, the causal link between the ambient temperature and mortality was investigated under the potential outcome framework. Consistency between new approach and existing regression method reinforced usefulness of regression method, and added an evidence of causal relationship found by previous studies. However, there are some unsolved limitations.

In the potential outcome framework, we could not do several analyses that the DLNM framework can do: identification of the lagged effect, distribution based confidence interval, and meta-regression.

The DLNM framework is able to produce exposure-response surface with treatment dimension and lag dimension. Therefore, the lagged effect of any treatment level can be identified after obtaining the surface. In contrast, we estimated the effect of the treatment by only using the current outcome variable. This can produce exposure-response curve, not the surface since we did not use any information about lagged treatment or lagged outcome. Thus our approach does not have ability to measure the lagged effect of the treatment. Actually, there

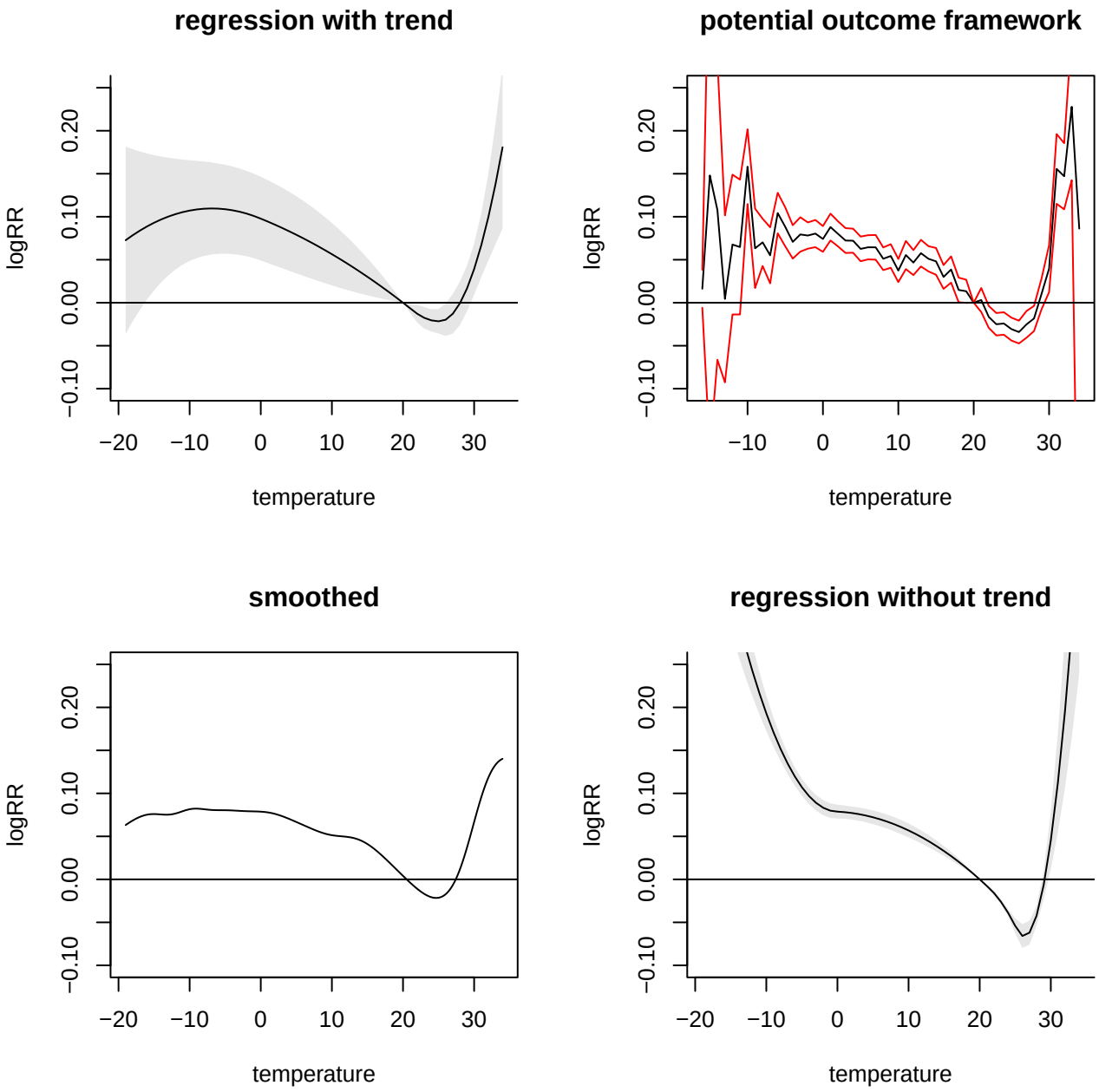

Figure 2: Estimated overall effect. For extreme hot or cold temperature, estimated effects have quite large uncertainties.

is a method to measure the lagged effect in a single time series setting [2]. However, this paper is limited to binary (sequential) treatment setting, so the curse of dimension makes it hard to be directly applicable to continuous treatment case. Someone may say to us that the calculated logRR curve does not represent overall effect but represents instant effect since we did not account for lagged effect. However, under the assumption that treatment history would have been almost the same during short period, due to their high autocorrelation, our effect estimate represents overall effect.

The DLNM framework provides the confidence interval based on asymptotic normality followed by maximizing quasi-likelihood function. However, in our case, we quantified the uncertainty of estimated logRR curve by moving block bootstrap to include information about temporal correlation. This method has a crucial disadvantage. Consider the situation that we have only one observation for some treatment level w. Then its uncertainty based on MBB must be zero. It contradicts to our intuition that larger sample size gives lower uncertainty. Although, there was no alternative way to obtain standard error of logRR curve. The difficulty of getting the exact or asymptotic distribution of our estimator comes from the definition of RR, which is the ratio between two effects, not the difference. Moreover, the standard error of pooled logRR curve seems not similar to the one from the DLNM framework. This is because, the precision of overall effect was obtained by summing up all precisions from each region, but some regions did not have observations of extremal temperature. This leads us to relatively low precision of overall effect estimates in extreme temperature.

It is well known that the temperature-mortality relationship is heterogeneous across regions(citation). The heterogeneity has been explained by meta-regression that has spline coefficients obtained from the DLNM framework as the response variable, and regional level variables such as latitude or climate zone as meta-predictors. Loosely speaking, the meta-regression with coefficients is analyzing the heterogeneity of curve itself across regions. However, in our suggestion, we just pooled the effect estimates from each region by naively taking weighted average for each treatment value. It did not use the information near a given treatment value, since the covariance of estimated effects was unavailable.

In addition to the weekness compared to the regression method stated above, our method has several disadvantages of itself to be addressed: incorrect treatment assignment mechanism, high variance, and violation of assumptions.

IPW is a method to generate a pseudo-population with marginally randomized treatment from the data of conditionally randomized experiment. We insist that the daily mean temper-

ature is conditionally randomized. Because from the viewpoint of the Earth, the "assignment mechanism" of daily mean temperature is heavily dependent on the meridinal altitude. Moreover, the meridinal altitude is able to be predicted almost perfectly by the date. So, we can say that conditional on the date, daily mean temperature is randomly determined where the randomness comes from cloud, rain, air mass, typhoon, CO2 emission, global warming, etc. It would be very good if we can include those factors into our GPS model to adjust for confounding bias of such meteorological variables, but it is impossible because of practical issue. Rather, by exploiting the fact that there are so many factors that may have influence on daily mean temperature, we gave normal assumption on the distribution of daily mean temperature based on the central limit theorem.

We estimated logRR curve without any modeling assumption after confounding adjustment. Also, the estimator does not borrow information near given treatment value. This makes the estimate soley rely on observed values. However, we have few observations for extreme temperature. So few observations play an important role in effect estimating procedure, and it means our approach may have high variance and low bias. It is the same for previous studies that there are only few extreme temperature observations, but they assumed parametric model to the outcome generating process so effect estimates at extreme temperature is a result of borrowing information near the temperature point, and it makes extreme cases play somewhat shrinked role compared to our approach, which means lower variance and higher bias.

There is a possibility of violation on two key assumptions in the potential outcome framework. The first one is unconfoundedness that potential outcome and treatment are independent given measured confounders. This assumption may be violated when there is an unmeasured confounder, since unmeasured confounder can change the distribution of potential outcome. In our case, we cannot measure everything related to temperature-mortality relationship so it is plausible to think unconfoundedness assumption is violated. The second one is positivity (overlap) assumption that any treatment has positive probability of being assigned for each confounder. In our case, the confounder is time and there is always some chance of extreme temperature because of catastrophic events in theory. However those events rarely happen in reality, so stochastic positivity violation [20] can happen. The reason is we don't have enough sample size. To address this issue, we made normal assumption on gps, and trimmed very small probability to 0.1 to ensure stability.

References

[1] Joshua D. Angrist, Òscar Jordà, and Guido M. Kuersteiner. Semiparametric estimates of monetary policy effects: String theory revisited. *Journal of Business & Economic Statistics*, 36(3):371–387, 2018.

[2] Iavor Bojinov and Neil Shephard. Time series experiments and causal estimands: Exact randomization tests and trading. *Journal of the American Statistical Association*, 114(528):1665–1682, 2019.

[3] A. Gasparrini, B. Armstrong, and M. G. Kenward. Distributed lag non-linear models. *Statistics in Medicine*, 29(21):2224–2234, 2010.

[4] Antonio Gasparrini. Modelling Lagged Associations in Environmental Time Series Data: A Simulation Study. *Epidemiology*, 27(6), 2016.

[5] Antonio Gasparrini, Yuming Guo, Francesco Sera, Ana Maria Vicedo-Cabrera, Veronika Huber, Shilu Tong, Micheline de Sousa Zanotti Stagliorio Coelho, Paulo Hilario Nascimento Saldiva, Eric Lavigne, Patricia Matus Correa, Nicolas Valdes Ortega, Haidong Kan, Samuel Osorio, Jan Kyselý, Aleš Urban, Jouni J K Jaakkola, Niilo R I Ryti, Mathilde Pascal, Patrick G Goodman, Ariana Zeka, Paola Michelozzi, Matteo Scortichini, Masahiro Hashizume, Yasushi Honda, Magali Hurtado-Diaz, Julio Cesar Cruz, Xerxes Seposo, Ho Kim, Aurelio Tobias, Carmen Iñiguez, Bertil Forsberg, Daniel Oudin Åström, Martina S Ragettli, Yue Leon Guo, Chang fu Wu, Antonella Zanobetti, Joel Schwartz, Michelle L Bell, Tran Ngoc Dang, Dung Do Van, Clare Heaviside, Sotiris Vardoulakis, Shakoor Hajat, Andy Haines, and Ben Armstrong. Projections of temperature-related excess mortality under climate change scenarios. *The Lancet Planetary Health*, 1(9):e360–e367, 2017.

[6] Gretchen T. Goldman and Francesca Dominici. Don't abandon evidence and process on air pollution policy. *Science*, 363(6434):1398–1400, 2019.

[7] Paul W. Holland. Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960, 1986.

[8] Guido Imbens and Keisuke Hirano. The propensity score with continuous treatments. 2004.

[9] Guido W. Imbens. The role of the propensity score in estimating dose-response functions. *Biometrika*, 87(3):706–710, 2000.

[10] Yoonhee Kim, Ho Kim, Antonio Gasparrini, Ben Armstrong, Yasushi Honda, Yeonseung Chung, Chris Fook Sheng Ng, Aurelio Tobias, Carmen Íñiguez, Eric Lavigne, Francesco Sera, Ana M. Vicedo-Cabrera, Martina S. Ragettli, Noah Scovronick, Fiorella Acquaotta, Bing-Yu Chen, Yue-Liang Leon Guo, Xerxes Seposo, Tran Ngoc Dang, Micheline de Sousa Zanotti Stagliorio Coelho, Paulo Hilario Nascimento Saldiva, Anna Kosheleva, Antonella Zanobetti, Joel Schwartz, Michelle L. Bell, and Masahiro Hashizume. Suicide and ambient temperature: A multi-country multi-city study. *Environmental Health Perspectives*, 127(11):117007, 2019.

[11] Hans R. Kunsch. The Jackknife and the Bootstrap for General Stationary Observations. *The Annals of Statistics*, 17(3):1217 – 1241, 1989.

[12] Hernán MA and Robins JM. *Causal Inference: What If*. Chapman & Hall/CRC, 2020.

[13] PAUL R. ROSENBAUM and DONALD B. RUBIN. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 04 1983.

[14] Donald B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66:688–701, 1974.

[15] Donald B. Rubin. The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Statistics in Medicine*, 26(1):20–36, 2007.

[16] R. W. M. WEDDERBURN. Quasi-likelihood functions, generalized linear models, and the Gauss—Newton method. *Biometrika*, 61(3):439–447, 12 1974.

[17] X. Wu, D. Braun, J. Schwartz, M. A. Kioumourtzoglou, and F. Dominici. Evaluating the impact of long-term exposure to fine particulate matter on mortality among the elderly. *Science Advances*, 6(29):eaba5692, 2020.

[18] Stanley Xu, Colleen Ross, Marsha A. Raebel, Susan Shetterly, Christopher Blanchette, and David Smith. Use of stabilized inverse propensity scores as weights to directly estimate relative risk and its confidence intervals. *Value in Health*, 13(2):273–277, 2010.

[19] Ghosh D. Zhu Y, Coffman DL. A boosting algorithm for estimating generalized propensity scores with continuous treatments. *Journal of Causal Inference*, 3(1):25–40, 2015.

[20] Paul N Zivich, Stephen R Cole, and Daniel Westreich. Positivity: Identifiability and estimability. 2022.