mathematics

MDPI

Article

Semiparametric Permutation-Based Change Point Detection with an Application on Chicago Cardiovascular Mortality Data

Hamdy F. F. Mahmoud [1,2]

[1] Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA; ehamdy@vt.edu

[2] Department of Statistics, Mathematics and Insurance, Faculty of Commerce, Assiut University, Assiut 71515, Egypt

Abstract: Climate change has several negative effects on health, including cardiovascular disease. Many studies have considered the effect of temperature on cardiovascular disease and found that there is an association between extreme levels of temperature, cold and hot, and cardiovascular disease. However, the number of articles that have studied the change point or the threshold in temperature is very limited. To the best of our knowledge, there have been no studies focusing on detecting and testing the significance of the change point in the temperature–cardiovascular relationship. Identifying the change point in cities may help to design better adaptive strategies in view of predicted weather changes in the future. Knowing the change points of temperature may prevent further mortality associated with the weather changes. Therefore, in this paper, we propose a unified approach that simultaneously estimates the semiparametric relationship and detects the significant point. A semiparametric generalized change point single index model is introduced as our unified approach by adjusting for several weather variables. A permutation-based testing procedure to detect the change point is introduced as well. A simulation study is conducted to evaluate the proposed algorithm. The advantage of our proposed approach is demonstrated using the cardiovascular mortality data of the city of Chicago, USA.

Keywords: cardiovascular disease; threshold detection; semiparametric regression; single index model

Citation: Mahmoud, H.F.F. Semiparametric Permutation-Based Change Point Detection with an Application on Chicago Cardiovascular Mortality Data. *Mathematics* **2022**, *10*, 857. https://doi.org/10.3390/math10060857

Academic Editor: Christophe Chesneau

Received: 5 February 2022
Accepted: 7 March 2022
Published: 8 March 2022

1. Introduction

Climate change has several negative effects on health. Extreme temperature events are expected to continue to increase as a result of climate change. The increase in cardiovascular deaths is associated with the increase in exposure to climate change-related heat [1]. Cardiovascular disease (CVD), which is globally considered as the leading cause of death, includes heart failure, heart disease, stroke, heart attack, heart valve problems, and arrhythmia. In 2019, the World Health Organization (WHO) estimated that 17.9 million people died from cardiovascular diseases, which represents 32% of all global deaths. Out of the 17.9 million people, 85% died due to stroke and heart attack (World Health Organization, 2021). In the United States, the American Heart Association (AHA) stated that cardiovascular disease is the leading cause of death for both men and women and reported that approximately 82.6 million Americans currently have at least one form of cardiovascular disease [2]. This disease produces immense health and economic burdens in the United States and globally as well. Each day, approximately 2150 Americans die due to cardiovascular diseases, accounting for a person every 40 seconds, which represents 1 in every 3 deaths in the United States. The number of people who die due to stroke and heart disease, in the Unites States, continues to increase more than any other cause of death, despite, and likely even due to, the impact of the COVID-19 pandemic during the last year. The number of deaths due to heart disease increased during the pandemic in 2020 [3]. Of those who have a heart attack or stroke, many may experience a diminished quality of life due to discomfort and disability. Among stroke survivors, 15–30% suffer some form of permanent disability. Cardiovascular disease is the leading cause of death in Chicago, as it is in the United States. In terms of health care expenditures and related expenses, in 2010, stroke and heart disease accounted for more than USD 500 billion, and it is considered one of the most costly and widespread health problems that face the Unites States and the city of Chicago. With approximately 5500 Chicagoans being killed each year because of cardiovascular diseases, heart disease is considered the leading cause of death.

Temperature extremes, both cold and hot, have impacts on mortality and the morbidities of cardiovascular disease. Many articles have studied the effect of temperature on cardiovascular mortality ([1,3–11]) and morbidity, hospital admissions due to cardiovascular disease ([12–17]). In some cities, it is found that the cold effect was more harmful than heat [5,6,8,9]). In other cities, it is found that the cold only has an impact on cardiovascular disease and heat has no impact [13,14]. That is most likely due to the increasing use of air conditioning [5]. A few studies have focused on the association between heat and cardiovascular disease only [1,9,12]. A few articles adjusted the cardiovascular-temperature relationship with weather variables [12,13,15,17]. Ref. [12] found that there is an association between the disease and humidity and [13] found an association with pressure. The weather variables (temperature, humidity, pressure) are correlated. So, in our proposed model, in this article, we included these variables as a block (single index) to study its effect on cardiovascular disease. A unified approach is introduced to estimate the proposed model and test the significance of the change point simultaneously. A very limited number of studies were interested in finding the change point of temperature [8,12]. Ref. [8] considered the change point as the specific temperature associated with the lowest mortality and [8] as the temperature associated with the lowest deviance. None of the studies mainly focused on detecting the change point or on testing the significance of that change point to find the optimal temperature or the temperature range where, if out of this range, people's health is at risk due to cardiovascular disease. In addition, the existence of a significant change point means that both heat and cold have an impact on cardiovascular disease.

The distributed lag nonlinear model [18], the Poisson regression model, and the generalized additive model, were commonly used to estimate the association between temperature and cardiovascular disease. This paper will introduce a semiparametric change point single index model to model the relationship and a permutation-based approach to test the significance of the detected change point.

This paper aims to (1) propose a model that includes weather variables as a block to study its effect on the cardiovascular disease; (2) detect the change point in temperature and test its significance using a permutation-based procedure simultaneously with estimating the proposed model; and (3) study the cardiovascular-temperature relationship adjusted with other covariates for the city of Chicago, USA. The remainder of this paper is organized as follows. In Section 2, a semiparametric change point single index model will be introduced and a unified approach for model estimation and change-point testing approach based on a permutation test will be displayed. A simulation study to evaluate the estimation procedure and the permutation test is presented in Section 3. Section 4 considers applying our unified approach to data of Chicago. Section 5 provides a discussion and the conclusion.

2. Semiparametric Change Point Single Index Model and a Permutation Test

2.1. The Generalized Semiparametric Change Point Single Index Model (GSCP-SIM)

Let $\mathbf{y}$ be the response variable and assume that we have n observations and p covariates. Without loss of generality, we assume that there is a change point in variable $\mathbf{x_1}$. The proposed model with one change point, θ, takes the form

$$y_i \sim P_d(y_i|\mu_i), i = 1, 2, \ldots, n; \mu_i = g(f_1[\beta_1 x_{1i} + \beta_{11}(x_{1i} - \theta)_+ + \ldots + \beta_p x_{pi}] + f_2[z_{1i}] + \ldots + f_q[z_{qi}]), \quad (1)$$

where P_d stands for the probability distribution of the response variables $\mathbf{y}$, g is a link function of the generalized model, θ is the change point associated with the variable x_1,

$(f_1, \ldots, f_q)$ are the unknown functions that need to be estimated, $(\beta_1, \ldots, \beta_p)$ are the single index coefficient parameters, and $(x_{1i} - \theta)_+ = max(0, x_{1i} - \theta)$, $i = 1, 2, \ldots, n$. To solve the identifiability problem of the single index function, f_1, $||\beta|| = 1$ is used. For our motivating data, g is the log link function where the response variable, the number of Cardiovascular deaths, follows the Poisson distribution.

This model has several advantages: (1) it does not have the curse of the dimensionality problem compared to the generalized additive model, which is commonly used, (2) it enables us to detect the significant change point in the temperature–cardiovascular relationship, and detecting the change point, using this model, does not get affected by smoothing the unknown function, f_1, (3) it enables us to study the effect of weather variables as a block on cardiovascular disease. In the matrix form, Equation (1) can be written as

$$\mathbf{y} \sim P_d(\mathbf{y}|\boldsymbol{\mu}); \boldsymbol{\mu} = g(f_1[\boldsymbol{\beta}X] + f_2[\mathbf{z}_1] + \ldots + f_q[\mathbf{z}_q]), \tag{2}$$

where $X = \{\mathbf{x}_1, [\mathbf{x}_1 - \theta]_+, \ldots, \mathbf{x}_p\}$ is a $n \times (p+1)$ matrix of regressors values, and $\boldsymbol{\beta}$ is a $(p+1) \times 1$ vector of parameters. We define $[\mathbf{x}_1 - \theta]_+ = \max[0, \mathbf{x}_1 - \theta]$, where θ is the change point associated with $\mathbf{x}_1$ and $\mathbf{y}$ is the vector of observations that follow the Poisson distribution.

2.2. Estimation and Change Point Detection Unified Approach

In this subsection, a unified approach for simultaneously estimating the proposed model parameters of the single index function coefficients, the unknown functions, and detecting and testing the change point, is presented. Assume the proposed model under the alternative hypothesis, H_a: there is a change point associated with the explanatory variable, x_1, which takes the form

$$y_i \sim P_d(y_i|\mu_i), i = 1, 2, \ldots, n; \mu_i = g(f_1[\beta_1 x_{1i} + \beta_{11}(x_{1i} - \theta)_+ \ldots + \beta_p x_{pi}] + f_2[z_{1i}] + \ldots + f_q[z_{qi}]) \tag{3}$$

Calculate the residuals

$$\epsilon_i^{(a)} = y_i - \mu_i^{(a)}, i = 1, 2, \ldots, n. \tag{4}$$

Assume the proposed model under the null hypothesis, H_0: there is no change point associated with the explanatory variable, x_1, which takes the form

$$y_i \sim P_d(y_i|\mu_i), i = 1, 2, \ldots, n; \mu_i = g(f_1[\beta_1 x_{1i} + \ldots + \beta_p x_{pi}] + f_2[z_{1i}] + \ldots + f_q[z_{qi}]) \tag{5}$$

Calculate the residuals

$$\epsilon_i^{(0)} = y_i - \mu_i^{(0)}, i = 1, 2, \ldots, n, \tag{6}$$

where $\epsilon_i^{(a)}$ and $\mu_i^{(a)}$ are the error and mean value under the alternative hypothesis and $\epsilon_i^{(0)}$ and $\mu_i^{(0)}$ are the error and mean value under the null hypothesis. Under the null hypothesis, the permutation test assumptions are $E(\epsilon_i^{(0)}) = 0$, $\text{var}(\epsilon_i^{(0)}) = \sigma^2$ for all i, and $\text{cov}(\epsilon_i^{(0)}, \epsilon_j^{(0)}) = 0$ for all $i \neq j$.

The following algorithm steps are for simultaneously estimating the GSCP-SIM parameters $(\beta_1, \ldots, \beta_p)$, estimating the unknown functions $(f_1, \ldots, f_q)$, and detecting and testing the change point, θ, which is associated with the explanatory variable x_1.

Step 0: use a grid search to find a candidate change point $\theta^{(0)}$, create the design matrix X;

Step 1: under the null hypothesis, estimate the coefficient parameters, $\boldsymbol{\beta}$, of the single index function, f_1, using a general-purpose optimizer to reduce the p-dimension to one dimension, the *nlm* function in R is used in this paper. Then, fit the unknown functions $(f_1, \ldots, f_q)$ using the *gam* function in R. Compute the mean response, $\hat{\mu}_i^{(0)}$, and residuals, $\hat{\epsilon}_i^{(0)} = y_i - \hat{\mu}_i^{(0)}$;

Step 2: under the alternative hypothesis, estimate the coefficient parameters, $\boldsymbol{\beta}$, of the single index function, f_1, using a general-purpose optimizer to reduce the p-dimension to one dimension, the *nlm* function in R is used in this paper. Then, fit the unknown functions $(f_1, \ldots, f_q)$ using the *gam* function in R. Compute the mean response, $\hat{\mu}_i^{(a)}$, and residuals, $\hat{\epsilon}_i^{(a)} = y_i - \hat{\mu}_i^{(a)}$;

Step 3: calculate the test statistic,

$$T_{\mathbf{y}} = \frac{[\hat{\boldsymbol{\epsilon}}_{\mathbf{y}}^{(0)}]^T [\hat{\boldsymbol{\epsilon}}_{\mathbf{y}}^{(0)}]}{[\hat{\boldsymbol{\epsilon}}_{\mathbf{y}}^{(a)}]^T [\hat{\boldsymbol{\epsilon}}_{\mathbf{y}}^{(a)}]}, \tag{7}$$

where $\hat{\boldsymbol{\epsilon}}_{\mathbf{y}}^{(a)}$ and $\hat{\boldsymbol{\epsilon}}_{\mathbf{y}}^{(0)}$ are the residuals under the alternative and null hypotheses of the original data set $\mathbf{y}$, respectively;

Step 4: obtain $\hat{\boldsymbol{\epsilon}}_m^{(0)}$ by permuting the residuals, $\hat{\boldsymbol{\epsilon}}_{\mathbf{y}}^{(0)}$, add the model means under the null hypothesis to the permuted residuals and obtain $\mathbf{y}_{(m)} = \hat{\boldsymbol{\mu}}^{(0)} + \hat{\boldsymbol{\epsilon}}_m^{(0)}$, where $\mathbf{y}_{(m)}$ is $n \times 1$ vector of permuted responses, $\hat{\boldsymbol{\epsilon}}_m^{(0)}$ is the permuted $n \times 1$ vector of residuals, and m stands for the mth permutation;

Step 5: estimate the model under the alternative and null hypotheses, for the permuted data set, $\mathbf{y}_{(m)}$, as described in steps 1–2 and calculate the following test statistic:

$$T_{\mathbf{y}_{(m)}} = \frac{[\hat{\boldsymbol{\epsilon}}_{\mathbf{y}_{(m)}}^{(0)}]^T [\hat{\boldsymbol{\epsilon}}_{\mathbf{y}_{(m)}}^{(0)}]}{[\hat{\boldsymbol{\epsilon}}_{\mathbf{y}_{(m)}}^{(a)}]^T [\hat{\boldsymbol{\epsilon}}_{\mathbf{y}_{(m)}}^{(a)}]}, \tag{8}$$

where, $\hat{\boldsymbol{\epsilon}}_{\mathbf{y}_{(m)}}^{(0)}$ and $\hat{\boldsymbol{\epsilon}}_{\mathbf{y}_{(m)}}^{(a)}$ are the residuals under the null and under the alternative hypotheses, respectively, for the mth permuted data set;

Step 6: repeat steps 4–5 for a big number, M, and obtain $T_{\mathbf{y}_{(m)}}, m = 1, 2, \ldots, M$, and finally compute the empirical p-value as follows

$$p - value = \frac{\sum I_{[T_{\mathbf{y}(m)} \geq T_{\mathbf{y}}]}}{M}, \ m = 1, 2, \ldots, M, \tag{9}$$

which measures how extreme the $T_{\mathbf{y}}$ value is;

Step 7: if the p-value is less than 0.05, we reject the null hypothesis and declare θ a significant change point, otherwise, we select another candidate change point and repeat Step 1–Step 6.

3. Simulation Study

This section is to conduct a simulation study to evaluate the introduced unified approach for estimating the proposed model, GSCP-SIM, and detecting and testing the change point using the permutation-based test. To evaluate the introduced permutation test in detecting the significant change point before applying it to the real data set, two evaluating criteria were used: (1) the probability of Type I error (the probability of rejecting the null hypothesis when it is in fact true), and (2) the power (the probability of rejecting the null hypothesis when it is in fact false).

To estimate the probability of Type I error, 1000 data sets were generated from the introduced model with no change point (H_0 is true). For each generated data set, the introduced permutation test described in Section 2.2, was run and the p-value is calculated. If p-value < 0.05, the null hypothesis is rejected, otherwise, the null hypothesis is failed to reject. To estimate the power, 1000 data sets were generated from the introduced model with a change point associated with a covariate variable (H_0 is false). For each generated data set, the introduced permutation test was run and the p-value was calculated. If p-value < 0.05, the null hypothesis is rejected, otherwise, the null hypothesis is failed to reject. In addition, to evaluate the algorithm in estimating the change point associated with the covariate variable, the change point was estimated simultaneously when calculating the power. For each generated data set out of 1000 data sets generated, under H_1 is correct, the detected change point was estimated, and then the mean of the estimates and the standard error of the mean were calculated. The following are the simulation settings of the data that were generated under H_0 and H_1:

Let $\mathbf{x_1} \sim Unif(0,1)$, $\mathbf{x_2} \sim Unif(0,1)$, $\mathbf{x_3} \sim Unif(0,1)$ and $\mathbf{z} \sim Unif(0,\pi)$. Without loss of generality, suppose we are interested in a change point in the variable $\mathbf{x_2}$ at the value 0.5, $\theta = 0.5$. The three variables $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ constitute the single index, $X\boldsymbol{\beta}$, with the coefficient vector of parameters $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_{21}, \beta_3) = (1, 1, -1, 1)/\sqrt(4)$. The sample size of each generated data is 200, $n = 200$. For fixing the identifiability problem of the single index function, $||\boldsymbol{\beta}|| = 1$ was used. Under this setting, $X\boldsymbol{\beta} = \mathbf{x}_1 + \mathbf{x}_2 - (\mathbf{x}_2 - 0.5)_+ + \mathbf{x}_3$, where $(\mathbf{x}_2 - 0.5)_+ = max(0, \mathbf{x}_2 - 0.5)$.

The GSCP-SIM under the alternative hypothesis, H_a: there is a change point associated with the explanatory variable x_2, which takes the form

$$\mathbf{y} \sim P_d(\mathbf{y}|\boldsymbol{\mu}); log(\boldsymbol{\mu}) = f_1(\boldsymbol{\beta}X/1.4)/4 + f_2(\mathbf{z}), \tag{10}$$

where $X = \{\mathbf{x}_1, [\mathbf{x}_1 - 0.5]_+, \mathbf{x}_2, \mathbf{x}_3\}$ is a 200×4 matrix of regressors values, $\boldsymbol{\beta}$ is a 4×1 vector of parameters, and $\mathbf{y}$ is the vector of observations that follows the Poisson distribution. Let $w = [\boldsymbol{\beta}X/1.4]$. Then $f_1(w)$ takes the form

$$f_1(w) = 0.2w^{11}(10(1-w))^6 + 10(10w)^3(1-w)^{10}, \tag{11}$$

and $f_2(z) = sin(z)$. Figure 1 shows the smoothed functions, $f_1(w)$ and $f_2(z)$, for one of the simulated data sets.

The proposed model, GSCP-SIM, under the null hypothesis, H_0: there is no change point associated with the explanatory variable x_2, takes the form

$$\mathbf{y} \sim P_d(\mathbf{y}|\boldsymbol{\mu}); log(\boldsymbol{\mu}) = f_1(\boldsymbol{\beta}X/1.4)/4 + f_2(\mathbf{z}), \tag{12}$$

where $X = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ is a 200×3 matrix of regressors values, and $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3) = (1, 1, 1)/\sqrt(3)$ is a 3×1 vector of parameters.

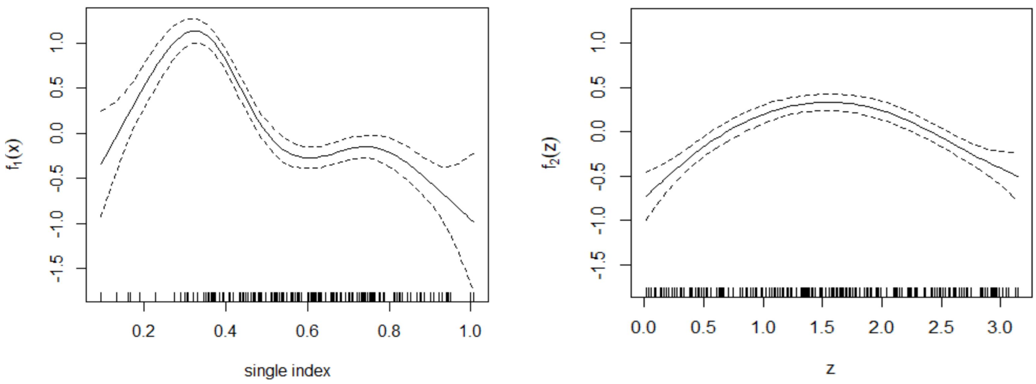

Figure 1. The p-spline smooth single index function, $f_1(w)$ and its 95% confidence interval (**left**), and the p-spline smooth function of the covariate variable, $f_2(z)$ (**right**) for one of the simulated data sets with a change point at 0.5, $\theta = 0.5$.

Form the simulation study for estimating Type I error, it was found that out of the generated 1000 data sets the null hypothesis was rejected 51 times. As a result, the estimated probability of Type I error for the introduce permutation test is 0.051 which is close the nominal value 0.05. From the simulation study of estimating the power of the test, it is found that out of the generated 1000 data sets the null hypothesis was rejected 989 times. As a result, the estimated power of the test is 0.989, which is close to 1. According to the two criteria, the introduced permutation-based test performs well in detecting the significant change point. For estimating the change point value, it was found that the mean is 0.502, which is close to the true value, $\theta = 0.5$, and the standard error is 0.0021, which is considered small. In sum, according to the simulation studies, it was concluded that the permutation test works well in detecting the change point and the estimation algorithm works well in estimating the change point value.

4. Real Data Application

In the National Morbidity, Mortality, and Air Pollution Study (NMMAPS), data were collected daily over the period from 1 January 1987 to 31 December 2000 from 108 U.S. to study the impact of air pollution on health in the United States [19]. The data have many variables, such as temperature, cardiovascular deaths, nitrogen dioxide, relative humidity, ozone, PM10, and many other variables. Each pollutant variable has been studied individually in the past with no significant effect detected on the CVD death rate [20,21]. Daily mortality counts were obtained from the National Center for Health Statistics and pollution data were obtained from the U.S. national air monitoring network provided by the US Environmental Protection Agency Aerometric Information Retrieval System database. The weather data came from the National Climatic Data Center. The data were made available in June 2004 and were removed from public availability in 2011 due to privacy concerns. The Chicago data is still available through the R package "dlnm" that was used in this paper. The response variables is CVD and relative humidity (%), pressure (sea level pressure, Hg) and temperature (°C) are the covariates. Pressure data are not available in the "dlnm" R package, so it is collected from the www.wunderground.com website. The main goals are to estimate the unknown cardiovascular–weather variables relationship and simultaneously detect and test the change point using the proposed model that was described in Section 2.1 and the permutation-based procedure that was described in Section 2.2.

4.1. Explanatory Data Analysis

Table 1 shows the numerical summary of the cardiovascular deaths and the weather variables by season. In total, the number of cardiovascular deaths in Chicago data was 260,314 collected during the study period from 1 January 1987 to 31 December 2000. The three-dimension plots for the cardiovascular deaths and the weather variables are displayed in Figure 2. From Figure 2, one can see that CVD deaths reaches the highest level at low levels of temperature and relative humidity, and high levels of relative humidity and pressure. Those figures are in three dimensions. We need to simultaneously study the effect of the weather variables as a block on the CVD deaths because they are correlated, and see whether there is a significant change point in the temperature–CVD relationship as well. Figure 3 shows the smoothed function of temperature and CVD and its 95% confidence interval, and the boxplots of CVD deaths over the 12 months of the year. Figure 3(left) reveals a potential change point in cardiovascular deaths associated with a temperature level within the interval (20 °C and 30 °C). Figure 3(right) shows that the CVD deaths level is high at the beginning of the year and at the end, and has a low level during the middle of the year with some outliers.

This motivates us to develop a single weather variables index that combines weather variables in a way that best describes the severity of the overall weather variables in terms of the CVD death rate and detect the change point in the temperature–cardiovascular relationship. To achieve these goals, we use the proposed model, GSCP-SIM, and the introduced permutation procedure.

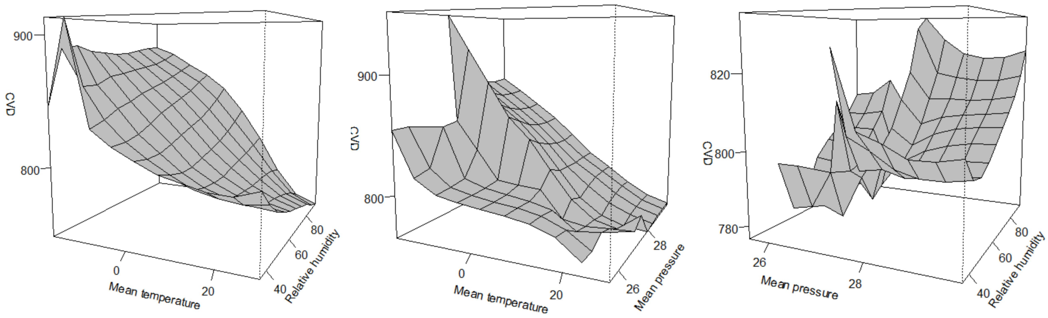

Figure 2. Three-dimensions scatterplot of mean temperature, relative humidity and CVD (**left**), mean temperature, mean pressure and CVD (**middle**), and mean pressure, relative humidity and CVD (**right**).

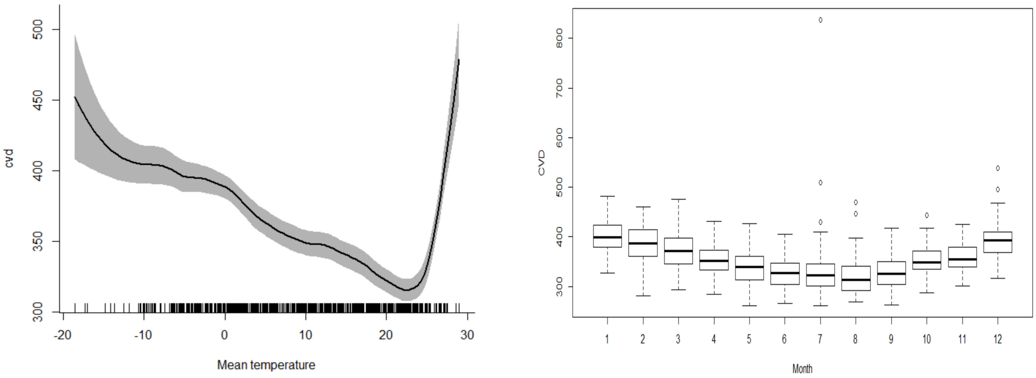

Figure 3. Mean temperature and CVD p-spline smoothed function of the unknown nonlinear relationship and its 95% confidence interval (**left**), and the boxplots of CVD deaths over the 12 months of the year (**right**).

Table 1. Numerical summary of weather variables and cardiovascular deaths by season in Chicago.

Variable		Mean ± SE	Minimum	Median	Maximum
Cardiovascular deaths	Winter	394.2 ± 4.96	282	393	538
	Spring	356.0 ± 4.49	262	354	476
	Summer	328.2 ± 6.22	262	321	837
	Fall	346.7 ± 4.11	263	342	444
Temperature	Winter	−2.58 ± 0.52	−18.53	−1.71	9.84
	Spring	9.84 ± 0.45	−6.67	9.76	25.00
	Summer	22.27 ± 0.34	13.88	22.22	29.00
	Fall	10.58 ± 0.44	−3.81	10.59	24.09
Humidity		69.65 ± 0.36	35.55	70.23	92.45
Pressure		29.27 ± 0.01	25.55	29.32	29.77

4.2. The Model and Change Point Detection

Let x_1 be humidity, x_2 be pressure, x_3 be temperature, z be time, and $(x_3 - \theta)_+ = max(0, x_3 - \theta)$. Then the proposed model (GSCP-SIM) for the Chicago data with a change point in temperature, θ, takes the form

$$\mathbf{y} \sim Poisson(\mathbf{y}|\boldsymbol{\mu}); log(\boldsymbol{\mu}) = f_1(a) + f_2(\mathbf{z}), \tag{13}$$

where the $a = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x3 + \beta_{31}(x_3 - \theta)_+$ is the weather index with the change point in temperature, f_1 is the semiparametric function of weather variables, and f_2 is the nonparametric function of time.

Using the algorithm that was described in Section 2.2, the model parameters are estimated and the change point was detected and tested simultaneously. Figure 4 shows the estimated function of the weather variables and the time. The significant change point that was detected is equal to 23.1 °C with a 95% confidence interval equal to (22.8 °C and 23.4 °C). The standard error for calculating the 95% confidence interval was calculated using the bootstrapping approach. The change point, $\theta = 23.1$, was found to be significant using the permutation-based approach, which means that the cold and heat have a significant effect on cardiovascular mortality in Chicago, USA.

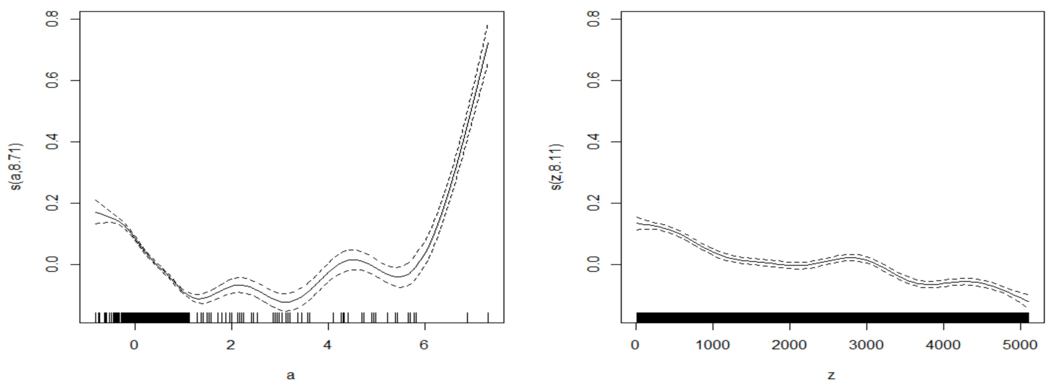

Figure 4. Smoothed single index function of weather variables and cardiovascular, $a = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x3 + \beta_{31}(x_3 - \theta)_+$ (**left**) and smoothed function of time (**right**).

5. Discussion and Conclusions

In this paper, a proposed model for estimating the relationship between weather variables and cardiovascular mortality with a change point in temperature is introduced: the generalized semiparametric change point single index model (GSCP-SIM). A unified approach is proposed for estimating the model parameters and unknown functions, as well as a permutation-based procedure for detecting and testing the change point simultaneously. A simulation study was conducted and it was found that the model estimation algorithm and the permutation test work well. The computed empirical Type I error was 5.1% and the power of the test is 0.989. Our approach was applied to Chicago cardiovascular mortality data and it was found that the change point in temperature is about 23.1 °C with a 95% confidence interval (22.8 °C, 23.4 °C). This means that both cold and heat have a significant impact on cardiovascular deaths in Chicago city. Our results may have important implications for the development of policies to reduce CVD deaths from extreme temperatures in Chicago. In other words, such knowledge might be useful to develop strategies to reduce the impact of extreme temperature episodes on human health due to climate changes.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is available through the R package “dlnm”.

Acknowledgments: The author would like to thank the editor and the anonymous referees for their valuable comments and suggestions to improve the paper. We acknowledge VT’s OASF support in the publishing of the article.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CVD	Cardiovascular disease
AHA	American Heart Association
GSCP-SIM	Generalized semiparametric change point single index model
Unif	Uniform distribution

References

1. Bloisa, D.J.; Kjellstromb, T.; Agewalld, S.; Ezekowitzf, J.A.; Armstrong, P.W.; Atar, D. The Effects of Climate Change on Cardiac Health. *Cardiology* **2018**, *131*, 209–217. [CrossRef] [PubMed]
2. Roger, V.L.; Go, A.S.; Lloyd-Jones, D.M.; Adams, R.J.; Berry, J.D.; Brown, T.M.; Carnethon, M.R.; Dai, S.; de Simone, G.; On behalf of the American Heart Association Statistics Committee Stroke Statistics Subcommittee; et al. Heart disease and stroke statistics—2011 update: A report from the American Heart Association. *Circulation* **2010**, *123*, e18–e209. [CrossRef] [PubMed]
3. Wadhera, R.K.; Figueroa, J.F.; Rodriguez, F.; Liu, M.; Tian, W.; Kazi, D.S.; Song, Y.; Yeh, R.W.; Maddox, K.E.J. Racial and Ethnic Disparities in Heart and Cerebrovascular Disease Deaths During the COVID-19 Pandemic in the United States. *Circulation* **2021**, *143*, 2346–2354. [CrossRef] [PubMed]
4. Mazidia, M.; Speakman, J.R. Predicted impact of increasing average ambient temperature over the coming century on mortality from cardiovascular disease and stroke in the USA. *Atherosclerosis* **2020**, *313*, 1–7. [CrossRef] [PubMed]
5. Barnett, A.G. Temperature and Cardiovascular Deaths in the US Elderly: Changes over Time. *Epidemiology* **2007**, *18*, 369–372. [CrossRef]
6. Wu, P.C.; Lin, C.Y.; Lung, S.C.; Guo, H.R.; Chou, C.H. Cardiovascular mortality during heat and cold events: Determinants of regional vulnerability in Taiwan. *Iran J. Public Health* **2011**, *49*, 1455–1466. [CrossRef]
7. Khanjani, N.; Bahrampour, A. Temperature and cardiovascular and respiratory mortality in desert climate. A case study of Kerman, Iran. *Iran. J. Environ. Health Sci. Eng.* **2013**, *10*, 11. [CrossRef]
8. Yang, J.; Yin, P.; Zhou, M.; Ou, C.Q.; Guo, Y.; Gasparrini, A.; Liu, Y.; Yue, Y.; Gu, S.; Sang, S.; et al. Cardiovascular mortality risk attributable to ambient temperature in China. *Heart* **2015**, *101*, 1966–1972. [CrossRef]
9. Huang, H.; Deng, X.; Yang, H.; Zhou, X.; Jia, Q. Spatio-Temporal Mechanism Underlying the Effect of Urban Heat Island on Cardiovascular Diseases. *Occup. Environ. Med.* **2020**, *68*, 525–530. [CrossRef]
10. Achebak, H.; Devolder, D.; Ballester, J. Trends in temperature-related age-specific and sex-specific mortality from cardiovascular diseases in Spain: A national time-series analysis. *Lancet Planet Health* **2019**, *3*, e297–e306. [CrossRef]
11. Péres, W.E.; Ribeiro, A.F.S.; Russo, A.; Nunes, B. The Association between Air Temperature and Mortality in Two Brazilian Health Regions. *Climate* **2020**, *8*, 32. [CrossRef]
12. Lin, S.; Luo, M.; Walker, R.J.; Liu, X.; Hwang, S.; Chinery, R. Extreme High Temperatures and Hospital Admissions for Respiratory and Cardiovascular Diseases. *Epidemiology* **2009**, *20*, 738–746. [CrossRef] [PubMed]
13. Giang, P.N.; Dung, D.; Giang, K.B.; Vinhc, H.V.; Rocklöv, J. The effect of temperature on cardiovascular disease hospital admissions among elderly people in Thai Nguyen Province, Vietnam. *Glob. Health Act.* **2014**, *7*, 23649. [CrossRef] [PubMed]
14. Ponjoan, A.; Parramon, D.; Blanch, J.; Alves-Cabratosa, L.; Martí-Lluch, R.; Comas-Cufí, M.; Parramon, D.; Garcia-Gil, M.M.; Ramos, R.; Petersen, I. Effects of extreme temperatures on cardiovascular emergency hospitalizations in a Mediterranean region: A self-controlled case series study. *Environ. Health* **2017**, *16*, 32. [CrossRef] [PubMed]
15. Tian, Y.; Liu, H.; Si, Y.; Cao, Y.; Song, J.; Li, M.; Wu, Y.; Wang, X.; Xiang, X.; Juan, J.; et al. Association between temperature variability and daily hospital admissions for cause- specific cardiovascular disease in urban China: A national time-series study. *PLoS Med.* **2019**, *16*, e1002738. [CrossRef]
16. Huang, J.; Wang, L.; Wang, S.; Lu, Y.; Zhang, W.; Wang, J. Spatial and temporal characteristics of temperature effects on cardiovascular disease in Southern China using the Empirical Mode Decomposition method. *Sci. Rep.* **2018**, *8*, 14775. [CrossRef]
17. Zhai, G.; Zhang, K.; Chai, G. Lag effect of ambient temperature on the cardiovascular disease hospital admission in Jiuquan, China. *Air Qual. Atmos. Health* **2021**, *14*, 181–189. [CrossRef]
18. Gasparrinia, A.; Armstronga, B.; Kenwardb, M.G. Distributed lag non-linear models. *Stat. Med.* **2010**, *29*, 2224–2234. [CrossRef]
19. Samet, J.M.; Dominici, F.; Zeger, S.L.; Schwartz, J.; Dockery, D.W. The national morbidity, mortality, and air pollution study. part I: Methods and methodologic issues. *Res. Rep. Health Eff. Inst.* **2000**, *5*, 5–14.
20. Cox, L.A.; Popken, D. Has reducing fine particulate matter and ozone caused reduced mortality rates in the United States? *Ann. Epidemiol.* **2015**, **25**, 162–173. [CrossRef]
21. Turner, M.C.; Jerrett, M.; Pope, C.A.; Krewski, D.; Gapstur, S.M.; Diver, W.R.; Beckerman, B.S.; Marshall, J.D.; Su, J.; Crouse, D.L.; et al. Long-Term Ozone Exposure and Mortality in a Large Prospective Study. *Am. J. Respir. Crit. Care Med.* **2016**, *193*, 1134–1142. [CrossRef] [PubMed]