

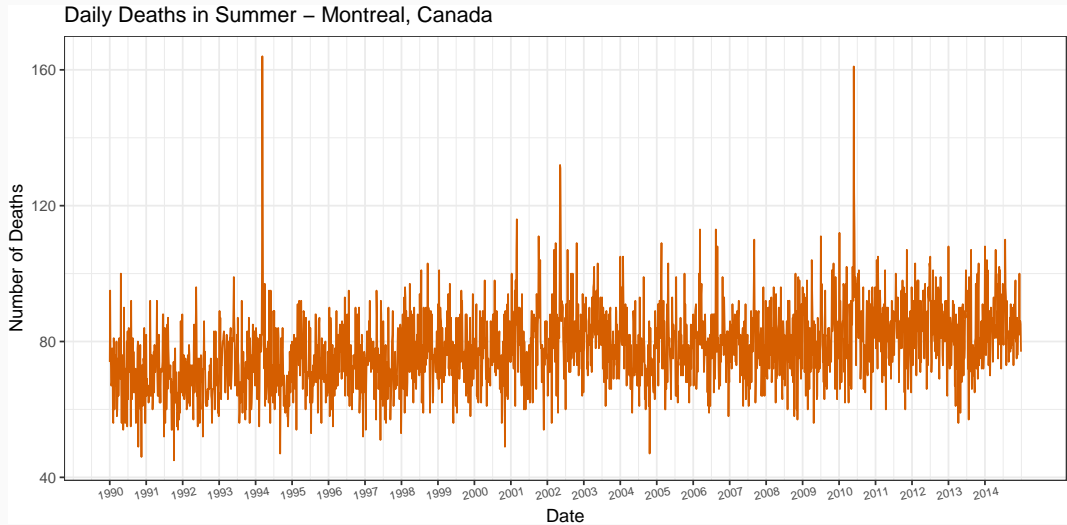
Sparse Multiple Index (SMI) Models for High-dimensional Nonparametric Forecasting

Nuwani Palihawadana

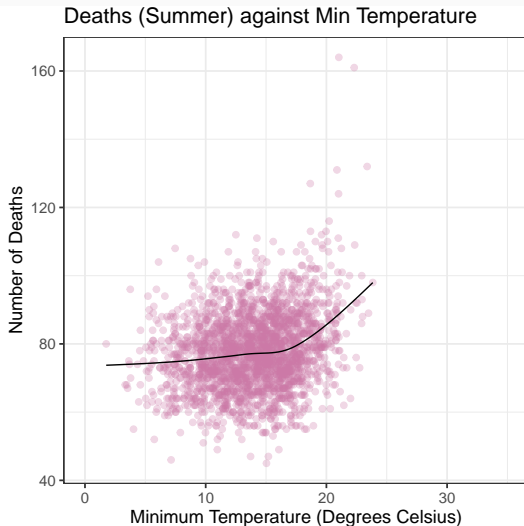
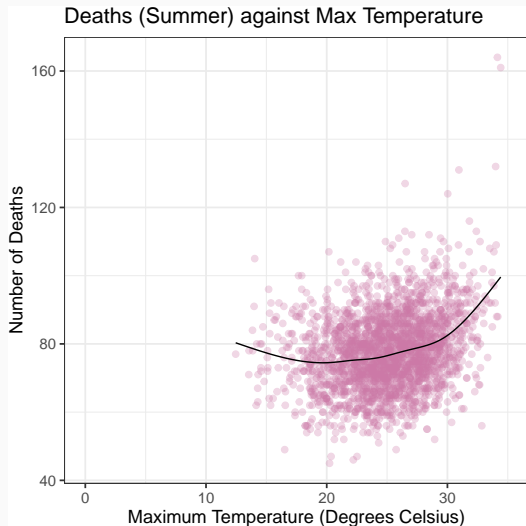
Joint work with : Rob Hyndman, Xiaoqian Wang

1 July 2024

Heat Exposure Related Daily Mortality



Heat Exposure Related Daily Mortality



■ *Nonlinear "Transfer Function" model*

$$y_t = f(\mathbf{x}_t, \mathbf{x}_{t-1}, \dots, \mathbf{x}_{t-p}, y_1, \dots, y_{t-k}) + \varepsilon_t$$

y_t – variable to forecast

\mathbf{x}_t – a vector of predictors

ε_t – random error

■ *Nonlinear "Transfer Function" model*

$$y_t = f(\mathbf{x}_t, \mathbf{x}_{t-1}, \dots, \mathbf{x}_{t-p}, y_1, \dots, y_{t-k}) + \varepsilon_t$$

y_t – variable to forecast

\mathbf{x}_t – a vector of predictors

ε_t – random error

- Impossible to estimate f for large p – **curse of dimensionality**

■ *Nonlinear "Transfer Function" model*

$$y_t = f(\mathbf{x}_t, \mathbf{x}_{t-1}, \dots, \mathbf{x}_{t-p}, y_1, \dots, y_{t-k}) + \varepsilon_t$$

y_t – variable to forecast

\mathbf{x}_t – a vector of predictors

ε_t – random error

- Impossible to estimate f for large p – **curse of dimensionality**
- Reasonable to impose additivity constraints

$$f(\mathbf{x}_t, \mathbf{x}_{t-1}, \dots, \mathbf{x}_{t-p}) = \sum_{i=0}^p f_i(\mathbf{x}_{t-i})$$

■ *Nonlinear "Transfer Function" model*

$$y_t = f(\mathbf{x}_t, \mathbf{x}_{t-1}, \dots, \mathbf{x}_{t-p}, y_1, \dots, y_{t-k}) + \varepsilon_t$$

y_t – variable to forecast

\mathbf{x}_t – a vector of predictors

ε_t – random error

- Impossible to estimate f for large p – **curse of dimensionality**
- Reasonable to impose additivity constraints

$$f(\mathbf{x}_t, \mathbf{x}_{t-1}, \dots, \mathbf{x}_{t-p}) = \sum_{i=0}^p f_i(\mathbf{x}_{t-i}) \leftarrow \text{Nonparametric Additive Model}$$

Issues:

- 1 Challenging to estimate in a high-dimensional setting
- 2 Subjectivity in predictor selection, and predictor grouping to model interactions

! Issues:

- 1 Challenging to estimate in a high-dimensional setting
- 2 Subjectivity in predictor selection, and predictor grouping to model interactions

i Index Models:

- Mitigate difficulty of estimating a nonparametric component for each predictor
- Improve flexibility

$$y_i = g(\alpha^T x_i) + \varepsilon_i$$

Semi-parametric model

$$y_i = \beta_0 + \sum_{j=1}^p g_j(\boldsymbol{\alpha}_j^T \mathbf{x}_{ij}) + \sum_{k=1}^d f_k(w_{ik}) + \boldsymbol{\theta}^T \mathbf{u}_i + \varepsilon_i, \quad i = 1, \dots, n,$$

- y_i – univariate response
- $\mathbf{x}_{ij} \in \mathbb{R}^{\ell_j}$, $j = 1, \dots, p$ – p subsets of predictors entering indices
- $\boldsymbol{\alpha}_j$ – ℓ_j -dimensional vectors of index coefficients
- g_j, f_k – smooth nonlinear functions
- Additional predictors :
 - ▶ w_{ik} – nonlinear
 - ▶ \mathbf{u}_i – linear

Semi-parametric model

$$y_i = \beta_0 + \sum_{j=1}^p g_j(\boldsymbol{\alpha}_j^T \mathbf{x}_{ij}) + \sum_{k=1}^d f_k(w_{ik}) + \boldsymbol{\theta}^T \mathbf{u}_i + \varepsilon_i, \quad i = 1, \dots, n,$$

- y_i – univariate response
- $\mathbf{x}_{ij} \in \mathbb{R}^{\ell_j}$, $j = 1, \dots, p$ – p subsets of predictors entering indices
- $\boldsymbol{\alpha}_j$ – ℓ_j -dimensional vectors of index coefficients
- g_j, f_k – smooth nonlinear functions
- Additional predictors :
 - ▶ w_{ik} – nonlinear
 - ▶ \mathbf{u}_i – linear

Allow elements equal to zero in $\boldsymbol{\alpha}_j$ – "Sparse"

Semi-parametric model

$$y_i = \beta_0 + \sum_{j=1}^p g_j(\boldsymbol{\alpha}_j^T \mathbf{x}_{ij}) + \sum_{k=1}^d f_k(w_{ik}) + \boldsymbol{\theta}^T \mathbf{u}_i + \varepsilon_i, \quad i = 1, \dots, n,$$

- y_i – univariate response
- $\mathbf{x}_{ij} \in \mathbb{R}^{\ell_j}$, $j = 1, \dots, p$ – p subsets of predictors entering indices
- $\boldsymbol{\alpha}_j$ – ℓ_j -dimensional vectors of index coefficients
- g_j, f_k – smooth nonlinear functions
- Additional predictors :
 - ▶ w_{ik} – nonlinear
 - ▶ \mathbf{u}_i – linear

Both "p" and the predictor grouping among indices are unknown.

Overlapping of predictors among indices is not allowed.

Optimisation Problem

Let q be the *total number of predictors* entering indices.

$$\begin{aligned} \min_{\beta_0, p, \alpha, g, f, \theta} \quad & \sum_{i=1}^n \left[y_i - \beta_0 - \sum_{j=1}^p g_j(\alpha_j^T \mathbf{x}_i) - \sum_{k=1}^d f_k(w_{ik}) - \boldsymbol{\theta}^T \mathbf{u}_i \right]^2 \\ & + \lambda_0 \sum_{j=1}^p \sum_{m=1}^q \mathbb{1}(\alpha_{jm} \neq 0) + \lambda_2 \sum_{j=1}^p \|\alpha_j\|_2^2 \\ \text{s.t.} \quad & \sum_{j=1}^p \mathbb{1}(\alpha_{jm} \neq 0) \in \{0, 1\} \quad \forall m \end{aligned}$$

Optimisation Problem

Let q be the *total number of predictors* entering indices.

$$\begin{aligned} \min_{\beta_0, p, \boldsymbol{\alpha}, \mathbf{g}, \mathbf{f}, \boldsymbol{\theta}} \quad & \sum_{i=1}^n \left[y_i - \beta_0 - \sum_{j=1}^p g_j(\boldsymbol{\alpha}_j^T \mathbf{x}_i) - \sum_{k=1}^d f_k(w_{ik}) - \boldsymbol{\theta}^T \mathbf{u}_i \right]^2 \\ & + \lambda_0 \sum_{j=1}^p \sum_{m=1}^q \mathbb{1}(\alpha_{jm} \neq 0) + \lambda_2 \sum_{j=1}^p \|\boldsymbol{\alpha}_j\|_2^2 \\ \text{s.t.} \quad & \sum_{j=1}^p \mathbb{1}(\alpha_{jm} \neq 0) \in \{0, 1\} \quad \forall m \end{aligned}$$

- $\lambda_0 > 0$ – controls the number of selected predictors

Optimisation Problem

Let q be the *total number of predictors* entering indices.

$$\begin{aligned} \min_{\beta_0, p, \boldsymbol{\alpha}, \mathbf{g}, \mathbf{f}, \boldsymbol{\theta}} \quad & \sum_{i=1}^n \left[y_i - \beta_0 - \sum_{j=1}^p g_j(\boldsymbol{\alpha}_j^T \mathbf{x}_i) - \sum_{k=1}^d f_k(w_{ik}) - \boldsymbol{\theta}^T \mathbf{u}_i \right]^2 \\ & + \lambda_0 \sum_{j=1}^p \sum_{m=1}^q \mathbb{1}(\alpha_{jm} \neq 0) + \lambda_2 \sum_{j=1}^p \|\boldsymbol{\alpha}_j\|_2^2 \\ \text{s.t.} \quad & \sum_{j=1}^p \mathbb{1}(\alpha_{jm} \neq 0) \in \{0, 1\} \quad \forall m \end{aligned}$$

- $\lambda_0 > 0$ – controls the number of selected predictors
- $\lambda_2 \geq 0$ – controls the strength of the additional shrinkage

MIQP Formulation

$$\min_{\beta_0, p, \alpha, g, f, \theta, z} \sum_{i=1}^n \left[y_i - \beta_0 - \sum_{j=1}^p g_j(\alpha_j^T \mathbf{x}_i) - \sum_{k=1}^d f_k(w_{ik}) - \theta^T \mathbf{u}_i \right]^2$$
$$+ \lambda_0 \sum_{j=1}^p \sum_{m=1}^q z_{jm} + \lambda_2 \sum_{j=1}^p \sum_{m=1}^q \alpha_{jm}^2$$

$$\text{s.t. } |\alpha_{jm}| \leq M z_{jm} \quad \forall j, \forall m,$$

$$\sum_{j=1}^p z_{jm} \leq 1 \quad \forall m,$$

$$z_{jm} \in \{0, 1\}$$

MIQP Formulation

$$\min_{\beta_0, p, \alpha, g, f, \theta, z} \sum_{i=1}^n \left[y_i - \beta_0 - \sum_{j=1}^p g_j(\alpha_j^T \mathbf{x}_i) - \sum_{k=1}^d f_k(w_{ik}) - \boldsymbol{\theta}^T \mathbf{u}_i \right]^2$$
$$+ \lambda_0 \sum_{j=1}^p \sum_{m=1}^q z_{jm} + \lambda_2 \sum_{j=1}^p \sum_{m=1}^q \alpha_{jm}^2$$

$$\text{s.t. } |\alpha_{jm}| \leq M z_{jm} \quad \forall j, \forall m,$$

$$\sum_{j=1}^p z_{jm} \leq 1 \quad \forall m,$$

$$z_{jm} \in \{0, 1\} \quad \leftarrow \quad z_{jm} = \mathbb{1}(\alpha_{jm} \neq 0)$$

MIQP Formulation

$$\begin{aligned} \min_{\beta_0, p, \alpha, g, f, \theta, z} \quad & \sum_{i=1}^n \left[y_i - \beta_0 - \sum_{j=1}^p g_j(\alpha_j^T \mathbf{x}_i) - \sum_{k=1}^d f_k(w_{ik}) - \boldsymbol{\theta}^T \mathbf{u}_i \right]^2 \\ & + \lambda_0 \sum_{j=1}^p \sum_{m=1}^q z_{jm} + \lambda_2 \sum_{j=1}^p \sum_{m=1}^q \alpha_{jm}^2 \\ \text{s.t.} \quad & |\alpha_{jm}| \leq M z_{jm} \quad \forall j, \forall m, \\ & \sum_{j=1}^p z_{jm} \leq 1 \quad \forall m, \\ & z_{jm} \in \{0, 1\} \quad \leftarrow \quad z_{jm} = \mathbb{1}(\alpha_{jm} \neq 0) \end{aligned}$$

- $M < \infty$: If α^* is an optimal solution, then $\max(\{|\alpha_{jm}^*|\}_{j \in [p], m \in [q]}) \leq M$

Estimation Algorithm

Step 1: Initialise index structure and index coefficients

Estimation Algorithm

Step 1: Initialise index structure and index coefficients

- **PPR:** Projection Pursuit Regression Based Initialisation
- **Additive:** Nonparametric Additive Model Based Initialisation
- **Linear:** Linear Regression Based Initialisation
- **Multiple:** Pick One From Multiple Initialisations

Estimation Algorithm

Step 1: Initialise index structure and index coefficients

Step 2: Estimate nonlinear functions

Estimation Algorithm

Step 1: Initialise index structure and index coefficients

Step 2: Estimate nonlinear functions

Step 3: Update index coefficients

Estimation Algorithm

Step 1 : Initialise index structure and index coefficients

Step 2 : Estimate nonlinear functions

Step 3 : Update index coefficients

Step 4 : Iterate steps 2 and 3 – until:

Estimation Algorithm

Step 1 : Initialise index structure and index coefficients

Step 2 : Estimate nonlinear functions

Step 3 : Update index coefficients

Step 4 : Iterate steps 2 and 3 – until:

- convergence
- loss increases for 3 consecutive iterations OR
- max iterations

Estimation Algorithm

Step 1: Initialise index structure and index coefficients

Step 2: Estimate nonlinear functions

Step 3: Update index coefficients

Step 4: Iterate steps 2 and 3 until stopping criteria are reached

Step 5: Add a new index with dropped predictors, and repeat step 4

Estimation Algorithm

Step 1: Initialise index structure and index coefficients

Step 2: Estimate nonlinear functions

Step 3: Update index coefficients

Step 4: Iterate steps 2 and 3 until stopping criteria are reached

Step 5: Add a new index with dropped predictors, and repeat step 4

Step 6: Increase p by 1 in each iteration of step 5 – until:

Estimation Algorithm

Step 1 : Initialise index structure and index coefficients

Step 2 : Estimate nonlinear functions

Step 3 : Update index coefficients

Step 4 : Iterate steps 2 and 3 until stopping criteria are reached

Step 5 : Add a new index with dropped predictors, and repeat step 4

Step 6 : Increase p by 1 in each iteration of step 5 – until:

- no.of indices reaches q
- loss increases after the increment model OR
- solution maintains same no.of indices as previous iteration, and $\text{abs}(\text{difference of index coefficients between two successive iterations}) \leq \text{tolerance}$

Forecasting Heat Exposure Related Daily Mortality

Variables

- **Response: Daily deaths in Summer**
– 1990 to 2014 – Montreal, Canada
- **Index Variables:**
 - ▶ Death lags
 - ▶ Max temperature lags
 - ▶ Min temperature lags
 - ▶ Vapor pressure lags
- **Nonlinear:** DOS (day of the season), Year

Forecasting Heat Exposure Related Daily Mortality

Variables

- **Response: Daily deaths in Summer**
– 1990 to 2014 – Montreal, Canada
- **Index Variables:**
 - ▶ Death lags
 - ▶ Max temperature lags
 - ▶ Min temperature lags
 - ▶ Vapor pressure lags
- **Nonlinear:** DOS (day of the season), Year

Data Split

- **Training Set:** 1990 to 2012
- **Validation Set:** 2013
- **Test Set:** 2014

Forecasting Heat Exposure Related Daily Mortality

Variables

- **Response: Daily deaths in Summer**
– 1990 to 2014 – Montreal, Canada
- **Index Variables:**
 - ▶ Death lags
 - ▶ Max temperature lags
 - ▶ Min temperature lags
 - ▶ Vapor pressure lags
- **Nonlinear:** DOS (day of the season), Year

Data Split

- **Training Set:** 1990 to 2012
- **Validation Set:** 2013
- **Test Set:** 2014

$$\text{Deaths} = \beta_0 + \sum_{j=1}^p g_j(X\alpha_j) + f_1(\text{DOS}) + f_2(\text{Year}) + \epsilon,$$

Results

Model	Predictors	Indices	Test Set 1		Test Set 2	
			MSE	MAE	MSE	MAE
SMI Model (5, 12) - PPR	61	7	85.233	7.140	97.353	7.772
SMI Model (1, 0) - Additive	61	59	96.398	7.481	112.199	8.156
SMI Model (6, 11) - Linear	61	2	100.231	7.719	120.542	8.598
Backward Elimination	40	—	136.204	9.319	140.867	9.385
Group-wise Additive Index Model	61	4	90.763	7.247	106.251	7.928
Projection Pursuit Regression	61	4	90.698	7.343	110.497	8.057

SMI Model (a, b) $\rightarrow \lambda_0 = a, \lambda_2 = b$

- **Test Set 1:** Three months (June, July and August 2014)
- **Test Set 2:** One month (June 2014)

i Key features:

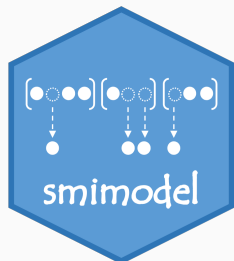
- Automatic selection of number of indices and predictor grouping
- Automatic predictor selection
- A wide spectrum: from single index models to additive models
- Flexibility to include separate nonlinear and linear predictors

Key features:

- Automatic selection of number of indices and predictor grouping
- Automatic predictor selection
- A wide spectrum: from single index models to additive models
- Flexibility to include separate nonlinear and linear predictors

Things to improve:

- Initialisation: we encourage trial-and-error
- Applicability: more applications are needed
- Computational time: increases with number of predictors and indices



■ **R package :**

github.com/nuwani-palihawadana/smimodel

■ **Paper :**

github.com/nuwani-palihawadana/smimodel_paper

Find me :



[in/nuwani-palihawadana](#)



[@nuwani-palihawadana](#)



nuwani.kodikarapalihawadana@monash.edu