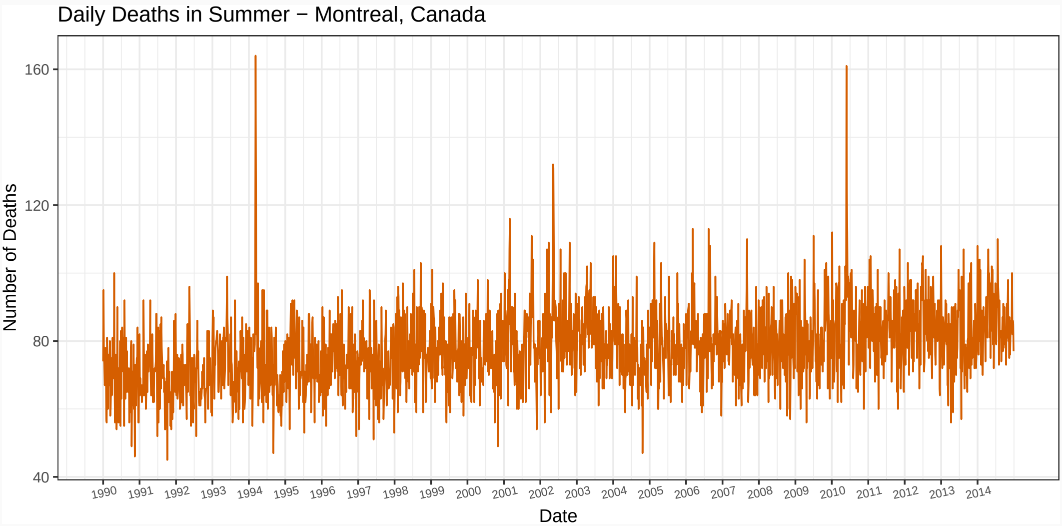
Sparse Multiple Index (SMI) Models for High-dimensional Nonparametric Forecasting

Nuwani Palihawadana

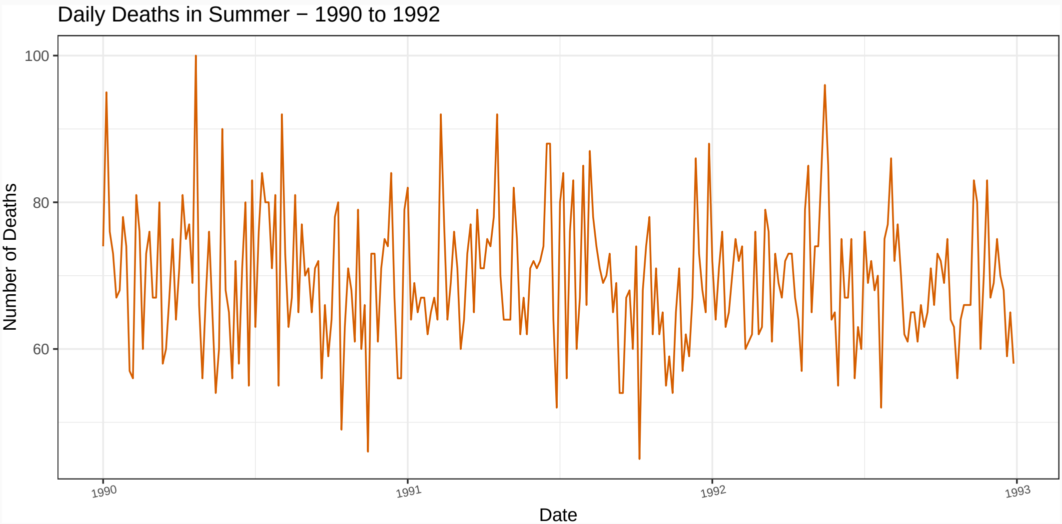
Joint work with : Rob Hyndman, Xiaoqian Wang

1 July 2024

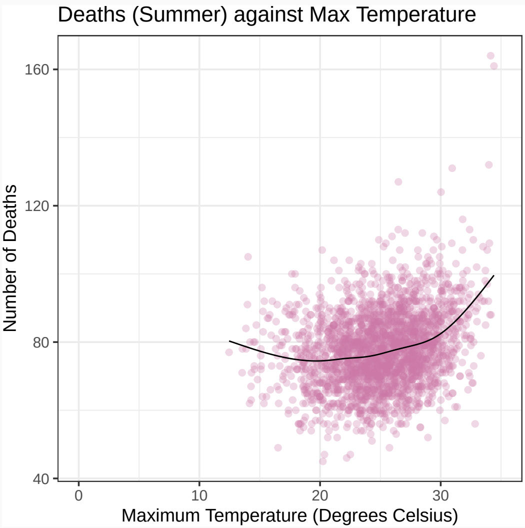
Heat Exposure Related Daily Mortality

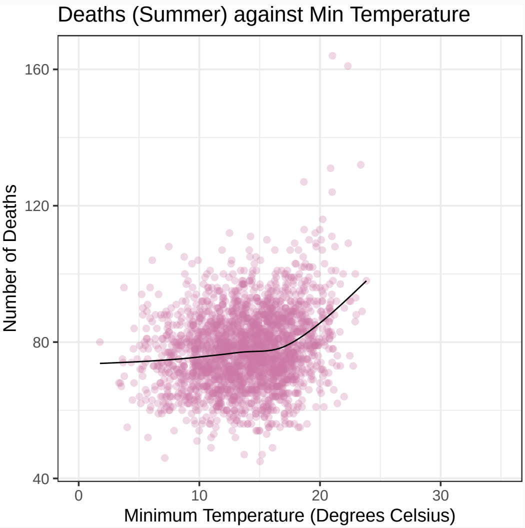


Heat Exposure Related Daily Mortality



Heat Exposure Related Daily Mortality





Background

- ***Nonlinear "Transfer Function" model***

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

y_t – variable to forecast
$\boldsymbol{x}_t$ – a vector of predictors
ε_t – random error

Background

- ***Nonlinear "Transfer Function" model***

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

y_t – variable to forecast
$\boldsymbol{x}_t$ – a vector of predictors
ε_t – random error

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

Background

- ***Nonlinear "Transfer Function" model***

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

y_t – variable to forecast
$\boldsymbol{x}_t$ – a vector of predictors
ε_t – random error

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

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

Background

- ***Nonlinear "Transfer Function" model***

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

y_t – variable to forecast
$\boldsymbol{x}_t$ – a vector of predictors
ε_t – random error

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

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

Background

Issues:

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

Background

Issues:

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

Index Models:

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

$$y_i = g\left(\boldsymbol{\alpha}^T \boldsymbol{x}_i\right) + \varepsilon_i$$

SMI Model

Semi-parametric model

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

- y_i – univariate response
- $\boldsymbol{x}_{ij} \in \mathbb{R}^{\ell_j}$, $j = 1, \dots, p$ – 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
  - $\boldsymbol{u}_i$ – linear

SMI Model

Semi-parametric model

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

- y_i – univariate response
- $\boldsymbol{x}_{ij} \in \mathbb{R}^{\ell_j}, j = 1, \dots, p$ – 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
  - $\boldsymbol{u}_i$ – linear

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

SMI Model

Semi-parametric model

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

- y_i – univariate response
- $\boldsymbol{x}_{ij} \in \mathbb{R}^{\ell_j}$, $j = 1, \dots, p$ – 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
  - $\boldsymbol{u}_i$ – 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.

$$\min_{\beta_0, p, \boldsymbol{\alpha}, \boldsymbol{g}, \boldsymbol{f}, \boldsymbol{\theta}} \quad \sum_{i=1}^{n} \left[y_i - \beta_0 - \sum_{j=1}^{p} g_j(\boldsymbol{\alpha}_j^T \boldsymbol{x}_i) - \sum_{k=1}^{d} f_k(w_{ik}) - \boldsymbol{\theta}^T \boldsymbol{u}_i \right]^2$$

$$+ \lambda_0 \sum_{j=1}^{p} \sum_{m=1}^{q} \mathbb{1}(\alpha_{jm} \neq 0) + \lambda_2 \sum_{j=1}^{p} \|\boldsymbol{\alpha}_j\|_2^2$$

$$\text{s.t.} \quad \sum_{j=1}^{p} \mathbb{1}(\alpha_{jm} \neq 0) \in \{0, 1\} \quad \forall m$$

Optimisation Problem

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

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

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

Optimisation Problem

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

$$\min_{\beta_0, p, \boldsymbol{\alpha}, \boldsymbol{g}, \boldsymbol{f}, \boldsymbol{\theta}} \sum_{i=1}^{n} \left[y_i - \beta_0 - \sum_{j=1}^{p} g_j(\boldsymbol{\alpha}_j^T \boldsymbol{x}_i) - \sum_{k=1}^{d} f_k(w_{ik}) - \boldsymbol{\theta}^T \boldsymbol{u}_i \right]^2$$

$$+ \lambda_0 \sum_{j=1}^{p} \sum_{m=1}^{q} \mathbb{1}(\alpha_{jm} \neq 0) + \lambda_2 \sum_{j=1}^{p} \|\boldsymbol{\alpha}_j\|_2^2$$

$$\text{s.t.} \quad \sum_{j=1}^{p} \mathbb{1}(\alpha_{jm} \neq 0) \in \{0, 1\} \quad \forall m$$

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

MIQP Formulation

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

MIQP Formulation

$$\min_{\beta_0, p, \boldsymbol{\alpha}, \boldsymbol{g}, \boldsymbol{f}, \boldsymbol{\theta}, \boldsymbol{z}} \quad \sum_{i=1}^{n} \left[y_i - \beta_0 - \sum_{j=1}^{p} g_j(\boldsymbol{\alpha}_j^T \boldsymbol{x}_i) - \sum_{k=1}^{d} f_k(w_{ik}) - \boldsymbol{\theta}^T \boldsymbol{u}_i \right]^2$$

$$+ \lambda_0 \sum_{j=1}^{p} \sum_{m=1}^{q} z_{jm} + \lambda_2 \sum_{j=1}^{p} \sum_{m=1}^{q} \alpha_{jm}^2$$

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

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

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

MIQP Formulation

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

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

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

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

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

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 abs(difference of index coefficients between two successive iterations) $<=$ 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

$$\textbf{Deaths} = \beta_0 + \sum_{j=1}^{p} g_j(\boldsymbol{X\alpha}_j) + f_1(\textbf{DOS}) + f_2(\textbf{Year}) + \varepsilon,$$

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) $\rightarrow$ $\boldsymbol{\lambda}_0 = \boldsymbol{a}, \boldsymbol{\lambda}_2 = \boldsymbol{b}$

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

Conclusion

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

Conclusion

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