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

Nuwani Palihawadana

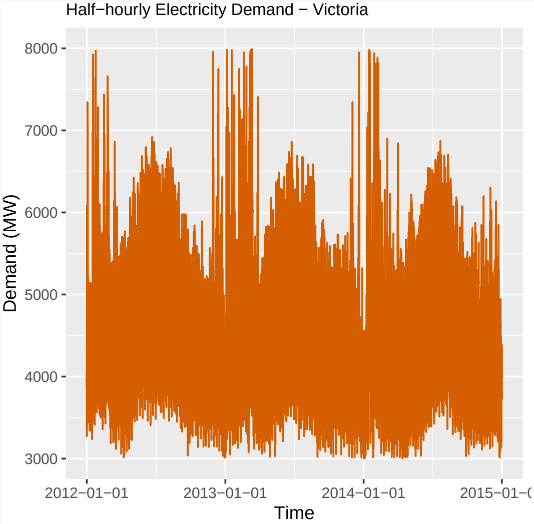
Joint work with : Rob Hyndman, Xiaoqian Wang

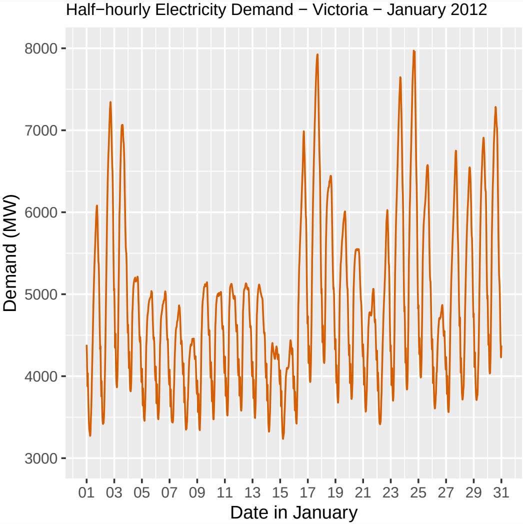
1 July 2024

Outline

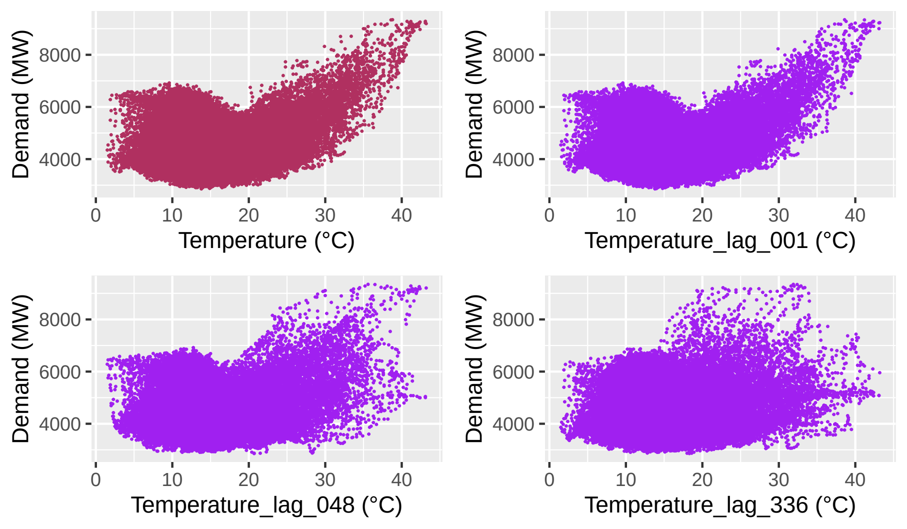
Outline

Electricity Demand Data





Electricity Demand Data



Outline

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

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

Outline

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

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
- Additional predictors :
  - w_{ik} – nonlinear
  - $\boldsymbol{u}_i$ – linear
- g_j, f_k : smooth nonlinear functions

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.

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

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

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

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

Optimisation Problem

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

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

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

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

- $\lambda_0 > 0$ – 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

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

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

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

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

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

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

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

- $M < \infty$: 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: Initialising the Index Structure and Index Coefficients

- Obtain a feasible initialisation :

Estimation Algorithm

Step 1: Initialising the Index Structure and Index Coefficients

- Obtain a feasible initialisation :

  1. **PPR:** Projection Pursuit Regression Based Initialisation

Estimation Algorithm

Step 1: Initialising the Index Structure and Index Coefficients

- Obtain a feasible initialisation :

  1. **PPR:** Projection Pursuit Regression Based Initialisation
  2. **Additive:** Nonparametric Additive Model Based Initialisation

Estimation Algorithm

Step 1: Initialising the Index Structure and Index Coefficients

- Obtain a feasible initialisation :

1. **PPR:** Projection Pursuit Regression Based Initialisation
2. **Additive:** Nonparametric Additive Model Based Initialisation
3. **Linear:** Linear Regression Based Initialisation

Estimation Algorithm

Step 1: Initialising the Index Structure and Index Coefficients

- Obtain a feasible initialisation :

  1. **PPR:** Projection Pursuit Regression Based Initialisation
  2. **Additive:** Nonparametric Additive Model Based Initialisation
  3. **Linear:** Linear Regression Based Initialisation
  4. **Multiple:** Picking One From Multiple Initialisations

- Scale each $\hat{\alpha}_j$ to have unit norm

Estimation Algorithm

Step 2: Estimating Nonlinear Functions

- Estimate a GAM :

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

where

- y_i – response
- $\hat{h}_{ij} = \hat{\boldsymbol{\alpha}}_j^T \boldsymbol{x}_i, j = 1, \ldots, p$ – estimated indices

Estimation Algorithm

Step 3: Updating Index Coefficients

$$\min_{\boldsymbol{\alpha}^{\text{new}},\boldsymbol{z}^{\text{new}}} (\boldsymbol{\alpha}^{\text{new}} - \boldsymbol{\alpha}^{\text{old}})^T \boldsymbol{V}^T \boldsymbol{V} (\boldsymbol{\alpha}^{\text{new}} - \boldsymbol{\alpha}^{\text{old}}) - 2(\boldsymbol{\alpha}^{\text{new}} - \boldsymbol{\alpha}^{\text{old}})^T \boldsymbol{V}^T \boldsymbol{r}$$

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

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

$$z_{jm}^{\text{new}} \in \{0, 1\},$$

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

- $\boldsymbol{V}$ – matrix of partial derivatives of RHS of SMI model equation, with respect to $\boldsymbol{\alpha}_j$
- $\boldsymbol{r}$ – current residual vector

Outline

Outline

Forecasting Heat Exposure Related Mortality

Forecasting Solar Intensity

Outline