Introduce.

a spline-based multilevel approach for effectively analyzing complex data.

developed is estimation procedure combining Expectation-Maximization algorithm & non-para. reg approach.

Key word : Expectation-Maximization
Hurdle model
Nonparametric regression
latent variable.
large-scaled data
Kernel smoother
smoothing spline.

Using data collected from various sources, however, pose few challenges.

① data structure is ~~often~~ complex.

② traditional assumptions on the sampling process are difficult to justify

③ the volume of data is larger than the data collected by traditional method.

* In this work
we propose a statistical framework to analyze the data merged from various sources.
The proposed model uses a "multilevel model" under non-para functional structural model assumption.

add modeling layer.

Nonparametric modeling approach provides a useful extention of ~~para~~ parametric models for hierarchical structure, by allowing an additional, flexible smoothness of para.

* our aim
to develop a statistical procedure for nonpara modeling approach. on the underlying reg fnc under the multilevel model structure, as well as deriving the theoretical properties of such procedure.

2. Model Methodology.

the location and date of each event of a group of musicians are observed over the entire USA for a given time period.

number of event, $i = 1, \cdots, n$
month, $j = 1, \cdots, J$

number of concert events in the jth month,

$$\mathbf{y}_j = (y_{1j}, \cdots, y_{nj})^T$$

with expactation vector

$$\eta_j = (\eta(X_1;\beta_j), \cdots, \eta(X_n;\beta_j))^T$$
$$= (\eta_{1j}, \cdots, \eta_{nj})^T$$

η is known link fnc.

$\beta_j = (\beta_{j1}, \cdots, \beta_{jp})^T$ is unknown time-varying para to be esti

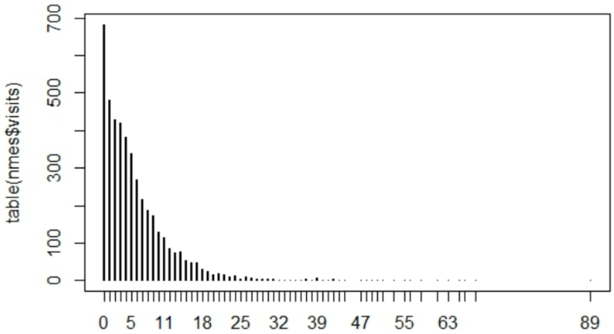
$X_i = (x_{i1}, \cdots, x_{ip})^T$ is CBSA specific regional covariates.

?

$\boxed{\text{assume}}$

$$y_{ij} \sim f(y_{ij}; X_i, \beta_j) \quad \cdots ①$$

where f is a parametric model (hurdle model)
아예 일어나지 않거나 일어나거나 하는 데이터.
(0 or positive)



The time varying para β_j's are $\boxed{\text{assume}}$

$$\beta_{jk} = \mu_k(z_j) + \varepsilon_{jk}, \quad k=1,\cdots,p$$

$\Downarrow$ p=1 로 간단하게 둔다.

$$\beta_j = \mu(z_j) + \varepsilon_j \quad \cdots ②$$

?

where months are re-scaled to $z_1, \cdots, z_J \in [0,1]$,

? $\mu(z_j)$ is mean fnc of β_j that belong to a class of smooth fnc with respect to z_j.

ε_j is random error $\sim N(0, \sigma^2)$, $\sigma^2 > 0$

* main interest.

to investigate how the relationship between y_{ij} and x_i changes over time,

This can be achived by finding the 2nd level model which capture the temporal change of β_j.

* The main challenge in this approach

there is no direct measurement of β_j

the smoothing splines approach & EM algorithm

- The joint pdf of $(X_1, \cdots, X_n, \theta)$

$$: g(X_1, \cdots, X_n; \theta) = \prod_1^n f(x_i|\theta)h(\theta)$$
$$= L(\theta|\underline{x})h(\theta)$$

- Marginal pdf of $X_1, \cdots, X_n$

$$: g_{\underline{X}}(\underline{x}) = \int_\Omega g(x_1, \cdots, x_n; \theta)\, d\theta$$

- Posterior distribution of θ.

Target.

given $x_1, \cdots, x_n$

$$k(\theta|\underline{x}) = \frac{g(x_1, \cdots, x_n; \theta)}{g_X(x_1, \cdots, x_n)}$$

2.2 Estimation

$\beta_j = \mu(z_j) + \varepsilon_j \Rightarrow$ consider the smoothing spline estimation!

$$\sum_{j=1}^J (\beta_j - \mu(z_j))^2 + \lambda \int_0^1 (\mu'')^2 \quad \cdots ③$$

$\mu \in W^{(2,2)}$, where $W^{(2,2)} = \{m \in L_2 : \int (m'')^2 < \infty\}$

Sobolev class of twice differentiable fnc

& λ is panalty constant

③ rewritten as

$$\min_{\theta \in \mathbb{R}^J} g(\theta : \beta) = (\beta - N\theta)^T(\beta - N\theta) + \lambda\theta^T W\theta \quad \cdots ④$$

where $\beta = (\beta_1, \cdots, \beta_J)$, $\theta = (\theta_1, \cdots, \theta_J)$

$N_{J\times J}^T = (N_{ij})$, $N_{ij} = N_i(z_j)$

$W = (W_{ij})$, $W_{ij} = \int N_i'' N_j''$ with natural cubic spline N_j having $z_1, \cdots, z_J$ knots

Power Basis for space

$$1, \quad x, \quad d_i(x) - d_{n-1}(x), \quad i = 1, \cdots, K-1$$

where $$d_i(x) = \frac{(x - X_i)_+^3 - (x - X_n)_+^3}{X_n - X_i}$$

일반적인

$$\underset{n\times K+1}{B} = \begin{pmatrix} 1 & x_1 & d_1(x_1) - d_{K-1}(x_1) & \cdots & d_{K-2}(x_1) - d_{K-1}(x_1) \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & d_1(x_n) - d_{K-1}(x_n) & \cdots & d_{K-2}(x_n) - d_{K-1}(x_n) \end{pmatrix}$$

$$\underset{J\times J}{N} = \begin{pmatrix} N_1(z_1) & \cdots & N_1(z_J) \\ \vdots & & \vdots \\ N_J(z_1) & \cdots & N_J(z_J) \end{pmatrix}$$

// ?

why J×J

information related to β_j

can be obtained from the obs (y_j, X)

$y_j = (y_{1j}, \cdots, y_{nj})$, j th month event count.

$X = (x_1, \cdots, x_n)$ CBSA specific information.

⟨Two step⟩

① obtain $\hat{\beta} = (\hat{\beta}_1, \cdots, \hat{\beta}_J)^T$ by individually fitting the model in $f(y_{ij}; \lambda_j, \beta_j)$ for each j, $j = 1, \cdots, J$

* Since β is not observable,

It is treated as a latent variable,

where $\hat{\beta}$ is treated as an observation for $\beta = (\beta_1, \cdots, \beta_J)^T$

Once all $\hat{\beta}_1, \cdots, \hat{\beta}_J$ are obtained,

define $h(\hat{\beta}_j | \beta_j)$

$\Rightarrow$ the prob density of the sampling dist of $\hat{\beta}_j$.

Maximum likelihood is used to obtain $\hat{\beta}_j$

$h(\hat{\beta}_j | \beta_j)$ can assume $AN(\beta_j, V_j)$

V_j is inverse of Fisher information matrix

② we estimate θ using EM, by minimizing

$$\sum_{j=1}^{J} (\beta_j - \mu(z_j))^2 + \lambda \int_0^1 (\mu'')^2$$

with a consideration of the $h(\hat{\beta}_j | \beta_j)$

Specifically, the smoothing spline criterian in

$$\min_{\theta \in \mathbb{R}^J} g(\theta; \beta) = (\beta - N\theta)^T (\beta - N\theta) + \lambda \theta^T W \theta$$

can be viewed as the negative complete loglikelihood with $(\beta | \theta, \sigma^2) \sim N(N\theta, \Sigma)$

$(\theta | \sigma^2) \sim N(0, \sigma^2 \lambda^{-1} W^{-1})$

?

where $\Sigma = \sigma^2 I_J$.

$$\max_{\theta \in \mathbb{R}^J} \int h(\hat{\beta} | \beta) \exp\left(-\frac{g(\theta; \beta)}{2}\right) d\beta \quad \cdots ⑤$$

— ⑤ 증명 — Line 1 → 2

$$\hat{\theta} = \arg\min E\left[(\hat{\beta} - \beta)^T V^{-1} (\hat{\beta} - \beta) + (\beta - N\theta)^T (\beta - N\theta) + \lambda \theta^T W \theta \,\middle|\, \hat{\beta}; \hat{\theta}^{(m)}, \sigma \right] \quad \cdots ⑥$$

It is straightforward to derive (유도하기 쉽다)

$$(\beta | \hat{\beta}; \theta, \sigma) \sim N\left((V^{-1} + \Sigma^{-1})^{-1} (V^{-1}\hat{\beta} + \Sigma^{-1} N\theta), (V^{-1} + \Sigma^{-1})^{-1} \right)$$

?

$$= N(\beta^*, V^*)$$

$$\frac{\partial}{\partial\theta} E[\cdot] = E\left[\frac{\partial}{\partial\theta} (\beta - N\theta)^T (\beta - N\theta) + \lambda \frac{\partial}{\partial\theta} \theta^T W \theta \,\middle|\, \hat{\beta}; \hat{\theta}^{(m)}, \sigma \right]$$

$$= E\left[-2N^T\beta + 2N^T N\hat{\theta} + 2\lambda W\hat{\theta} \,\middle|\, \hat{\beta}; \hat{\theta}^{(m)}, \sigma \right] \overset{\text{Let}}{=} 0$$

$$\hat{\theta}^{(m+1)} \cdot (N^T N + \lambda W) = N^T N \theta^{(m+1)} + \lambda W \theta^{(m+1)}$$

$$= N^T E[\beta | \hat{\beta}; \hat{\theta}^{(m)}, \sigma] = N^T \beta^*$$

$$= N^T (V^{-1} + \Sigma^{-1})^{-1} (V^{-1}\hat{\beta} + \Sigma^{-1} N\hat{\theta}^{(m)}) \quad \cdots ⑦$$

$$\therefore \hat{\theta}^{(m+1)} = (N^T N + \lambda W)^{-1} N^T (V^{-1} + \Sigma^{-1})^{-1} (V^{-1}\hat{\beta} + \Sigma^{-1} N\hat{\theta}^{(m)})$$

$m \to \infty$, $\hat{\theta}$

$$\hat{\theta} = (N^T N + \lambda W)^{-1} N^T (V^{-1} + \Sigma^{-1})^{-1} V^{-1}\hat{\beta} + (N^T N + \lambda W)^{-1} N^T (V^{-1} + \Sigma^{-1})^{-1} \Sigma^{-1} N\hat{\theta}$$

$$\left(I - (N^T N + \lambda W)^{-1} N^T (V^{-1} + \Sigma^{-1})^{-1} \Sigma^{-1} N\right) \hat{\theta} = (N^T N + \lambda W)^{-1} N^T (V^{-1} + \Sigma^{-1})^{-1} V^{-1} \hat{\beta}$$

$$(N^T N + \lambda W) \left(I - (N^T N + \lambda W)^{-1} N^T (V^{-1} + \Sigma^{-1})^{-1} \Sigma^{-1} N\right) \hat{\theta}$$

$$= N^T (V^{-1} + \Sigma^{-1})^{-1} V^{-1} \hat{\beta}$$

$$\therefore \hat{\theta} = \left((N^T N + \lambda W) - N^T (V^{-1} + \Sigma^{-1})^{-1} \Sigma^{-1} N \right)^{-1} N^T (V^{-1} + \Sigma^{-1})^{-1} V^{-1} \hat{\beta}$$

$$= \left[N^T (I - (V^{-1} + \Sigma^{-1})^{-1} \Sigma^{-1}) N + \lambda W \right]^{-1} N^T (V^{-1} + \Sigma^{-1})^{-1} V^{-1} \hat{\beta}$$

$$= [N^T D N + \lambda W]^{-1} N^T D \hat{\beta}$$

$$(\because D = I - V^* \Sigma^{-1} = V^* V^{-1})$$

$$\hat{\mu}(z) = \sum_{j=1}^{J} \hat{\theta}_j N_j(z) \quad \cdots ⑨$$

for smooth mean fnc.

$$\hat{\mu}(z) = \sum_{j=1}^{J} l_{\lambda,j}(z)\hat{\beta}_j$$

$\hat{\mu}(\cdot)$ is a linear smoother for $\hat{\beta}$

$$l_\lambda(z) = (l_{\lambda,1}(z), \cdots, l_{\lambda,J}(z))$$

Thus, the fitted response is

$$\hat{\mu} \equiv (\hat{\mu}(z_1), \cdots, \hat{\mu}(z_J))^T = S_\lambda \hat{\beta}$$

with smoothing matrix $S_\lambda = (l_\lambda(z_1), \cdots, l_\lambda(z_J))^T$

To choose the value of λ,

calculate Generalized Cross Validation (GCV)

$$GCV(\lambda) = J^{-1} \sum_{j=1}^{J} \left(\frac{\hat{\beta}_j - \hat{\mu}(z_j)}{1 - tr(S_\lambda)/J} \right)^2$$

effective degree of freedom is $df_\lambda = tr(S_\lambda)$

for a linear smoother.

where $h(\hat{\beta}_j | \beta_j) \sim AN(\beta_j, V_j)$

or $E(\beta | \hat{\beta}, \hat{\theta}^{(m)}, \sigma)$

$\beta_1^{(m)}, \cdots, \beta_B^{(m)}$ generated from dist.

$$\boxed{\begin{aligned}(\beta | \hat{\beta}; \theta, \sigma) &\sim N\left((V^{-1}+\Sigma^{-1})^{-1}(V^{-1}\hat{\beta} + \Sigma^{-1}N\theta), \ (V^{-1}+\Sigma^{-1})^{-1}\right) \\ &= N(\beta^*, V^*)\end{aligned}}$$

can be estimated by

$$\frac{\sum_{b=1}^{B} w_b \beta_b^{(m)}}{\sum_{b=1}^{B} w_b}$$

where w_b is weight for the $\beta_b^{(m)}$

$w_b = h(\hat{\beta} | \beta_b)$, straightforward choice

$$E(\beta | \hat{\beta}; \hat{\theta}^{(m)}, \sigma) = \int \beta \, h(\beta | \hat{\beta}; \hat{\theta}^{(m)}, \sigma) \, d\beta$$

$$= \int \beta \frac{h(\hat{\beta}, \beta | \hat{\theta}^{(m)}, \sigma)}{\int h(\hat{\beta}, \beta | \hat{\theta}^{(m)}, \sigma) d\beta} d\beta$$

$$= \frac{\int \beta h(\hat{\beta}|\beta) h(\beta | \hat{\theta}^{(m)}, \sigma) d\beta}{\int h(\hat{\beta}, \beta | \hat{\theta}^{(m)}, \sigma) \, d\beta}$$

$$= \frac{E_{\beta | \hat{\theta}^{(m)}, \sigma}[\beta h(\hat{\beta}|\beta)]}{E_{\beta | \hat{\theta}^{(m)}, \sigma}[h(\hat{\beta}|\beta)]} \quad \cdots ⑪$$

Thus, once a form of $h(\hat{\beta}|\beta)$ is given,

generate MC sample $\beta_b \sim (\beta_b ; \hat{\theta}^{(m)}, \sigma)$

$$E[\beta | \hat{\beta}, \hat{\theta}^{(m)}, \sigma] \approx \sum_{b=1}^{B} \beta_b h(\hat{\beta}|\beta_b) \Big/ \sum_{b=1}^{B} h(\hat{\beta}|\beta_b)$$

by WLLN large B.