Recent History Functional Linear Model

Kion Kim(kion@psu.edu)

Joint work
with
Damla Şentürk & Runze Li

The Pennsylvania State University

July. 29. 08

Functional Linear Models

Historical Functional Linear Models

Varying Coefficient Models

Recent History Functional Linear Model

Simulation

Kion Kim(kion@psu.edu)

Joint work with Damla Şentürk & Runze Li

Functional Linear Models

Let $x_i(s), s \in [0, T]$ be the predictor function and $y_i(t), t \in [0, T]$ be the response function, for $i = 1, \cdots, n$ subjects.
[Ramsay(1997)] considered the functional linear model,

$$y_i(t) = \alpha(t) + \int_0^T \beta(s,t) x_i(s) ds + \varepsilon_i(t), \quad s \in [0, T], t \in [0, T], \quad (1)$$

where $\alpha(t)$ is the intercept function, $\beta(s,t)$ is the bivariate regression function and $\varepsilon_i(t)$ is the error function with $E(\varepsilon_i(t)) = 0$, $Cov(\varepsilon_i(t), \varepsilon_j(t')) = \sigma_{tt'}$ for $i = j$ and 0 otherwise.

- We want to estimate the regression function, $\beta(s,t)$.
- We use regularization to estimate the function with infinite number of parameters.
- As a method of regularization, use tensor product basis function to estimate the surface due to rectangular support.

Historical Functional Linear Models (HFLM)

[Malfait and Ramsay(2003)] tailored the model in (1) to be applicable to situations where the response function $y(t)$ depends only on the predictors $x(s)$ at time $s \leq t$ as

$$y_i(t) = \alpha(t) + \int_{s_0(t)}^{t} \beta(s,t)x_i(s)ds + \varepsilon_i(t), \quad s \in [s_0(t), t], t \in [0, T], \tag{2}$$

where $s_0(t) = max(0, t-\delta)$ and $0 \leq \delta \leq t$.

- Here δ denotes the time lag back in time where the predictor function starts to have an impact on the response.
- The historical functional linear models have triangular support, considering only the effect of the past and present of the predictor process.
- Instead of the tensor product basis used in functional linear models, construct another two dimensional basis suitable for the triangular support, $\phi(s,t)$ to estimate the regression function $\beta(s,t)$. The technical method they employ is called Finite Element Method (FEM).

Varying Coefficient Models(VCM)

For n randomly selected subjects, each repeatedly measured over time, the longitudinal sample of $(Y(t), t, X(t))$ is denoted by $\{Y_{ij}, t_{ij}, X_{ij} : i = 1 \cdots, n, j = 1, \cdots, n_i\}$, where t_{ij} is the jth measurement time of the ith subject, Y_{ij} and $X_{ij} = (X_{0ij}, \cdots, X_{pij})^T$ are observed response and covariate vector at time t_{ij}. The varying coefficient models can be written as

$$Y_{ij} = X_{ij}^T \boldsymbol{\beta}(t_{ij}) + \varepsilon_i(t_{ij}), \tag{3}$$

where $\boldsymbol{\beta}(t_{ij})$ is $(p+1) \times 1$ vector of varying coefficients at time t_{ij} and the error term $\varepsilon_i(t_{ij})$ has mean 0 and is independent between subjects.

- Two popular methods of estimation in varying coefficient models are local polynomial fitting and basis approximation.

Recent History Functional Linear Model

In this new model, we allow the current response values to be affected by only the recent past of the predictor process,

$$y_i(t) = \alpha(t) + \int_{t-\delta_1}^{t-\delta_2} x_i(s)\beta(s,t)ds + \varepsilon_i(t), \quad t \in [\delta_1, T], s \in [0, T], \quad (4)$$

where $0 \leq \delta_2 \leq \delta_1 \leq t$. Here δ_1 allows for a lag for the predictor process to start affecting the response, while δ_2 denotes another lag. support

For a fixed time point t, $\beta(s,t)$ is a univariate function in s. we can expand this univariate function with respect to s using basis functions, $\phi_k(\cdot), k = 1, \cdots, K$, resulting in $\beta(s,t) \approx \sum_{k=1}^{K} b_k \phi_k(s)$, where $s \in [t - \delta_1, t - \delta_2]$. Repeating this univariate expansion for different t leads us to the following two dimensional expansion,

$$\beta(s,t) \approx \sum_{k=1}^{K} b_k(t)\phi_{k,t}(s), \quad s \in [t - \delta_1, t - \delta_2], \quad (5)$$

where $b_k(t)$ denotes the coefficient of the k^{th} basis function at time t.

The Estimation Steps

Defining $\psi_{i,k}(t)$ as $\psi_{i,k}(t) = \int_{t-\delta_1}^{t-\delta_2} x_i(s)\phi_{k,t}(s)ds$, the proposed model in (4) simplifies to

$$y_i(t) \approx \alpha(t) + \sum_{k=1}^{K} \psi_{i,k}(t)b_k(t) + \varepsilon_i(t), \quad t \in [\delta_1, T]. \tag{6}$$

The resulting model is a **varying coefficient model** and we are going to make use of the **two-step estimator**. (Fan & Zhang, 2000)

1. Estimation of $\psi_{ik}(t)$: In the case of functional data, assuming we have fine grid with no missing values, it is reasonable to use first order approximation for the integration.
2. Raw estimate of $b(t)$: Let $\hat{\mathbf{\Psi}}(t)$ be the $n \times K$ matrix, which has $\hat{\psi}_{i,j}(t)$ as $(i,j)^{th}$ element and $\tilde{\mathbf{\Psi}}(t) = [\mathbf{1}, \hat{\mathbf{\Psi}}(t)]$. If we define $\boldsymbol{y}(t)$ as an n dimensional vector, which contains the n realizations of the response variable on time point t, then, in the raw estimation step, we can get the raw estimates, $\tilde{\alpha}(t)$ and $\tilde{\boldsymbol{b}}(t) = [\tilde{b}_1(t), \cdots, \tilde{b}_K(t)]$, as

$$\begin{pmatrix} \tilde{\alpha}(t) \\ \tilde{\boldsymbol{b}}(t) \end{pmatrix} = \left[\tilde{\mathbf{\Psi}}(t)^T \tilde{\mathbf{\Psi}}\right]^{-1} \tilde{\mathbf{\Psi}}(t)^T \boldsymbol{y}(t), \quad t \in [\delta_1, 1] \tag{7}$$

The Estimation Steps

3. Refined estimate of $b(t)$: Let J be the collection of indices j that satisfy $t_j \in [\delta_1, 1]$ and τ' be the number of time points in J. Also let us define j_i as the i^{th} element in the index set J.
   In the refinement step, if a linear estimator with the weight function $w_k(t_j, t)$ is used for the estimation of the k^{th} coefficient function in the second step, the refined estimator, $\hat{b}_k(t), k = 1, \cdots, K$, can be given as

$$\hat{b}_k(t) = \sum_{i=1}^{\tau'} w_k(t_{j_i}, t)\tilde{b}_k(t_{j_i}). \tag{8}$$

Hence the bivariate coefficient function, $\beta(s,t)$, is directly estimated by

$$\hat{\beta}(s,t) = \sum_{k=1}^{K} \phi_k(s)\hat{b}_k(t). \tag{9}$$

Choosing the parameters-Strategy

In the model we propose, we need to select two time lags δ_1 and δ_2 and the number of basis functions used, K.

- **The choice of K for different δ_1 and δ_2 combination.**
  1. Decide the largest window size and the (δ_1, δ_2) candidate combinations that we want to consider.
  2. Choose the number of knots K for the largest window based on RSPE.
  3. For a given (δ_1, δ_2) combination, determine the number of knots, which is proportional to the window's relative length compared to the largest window considered at the beginning of the algorithm. Selection of K
- **The choice of δ_1 and δ_2 combination.**
  1. The different combination of δ_1 and δ_2 results in different number of basis functions K and gives us different varying coefficient model.$\Rightarrow$ *Model Selection Problem*.
  2. Since the varying coefficient model is a linear model for a given time point t, we can obtain AIC score for different combinations of (δ_1, δ_2) for a given time point.
  3. We pick the (δ_1, δ_2) combination of which AIC is minimized for most of the time points.

Criterions for the parameter selection

- For the selection of K, we used RSPE modified from [Muller and Zhang(2005)] and is defined by

$$RSPE(t) = \left[\frac{1}{n} \sum_{i=1}^{n} \left\{ \hat{y}_i^{(-i)}(t) - y_i(t) \right\}^2 \right]^{1/2},$$

  where $\hat{y}_i^{(-i)}(t)$ denotes the estimated response value for the i^{th} subject measured at time t from the data excluding the i^{th} subject.
- The problem of selecting δ_1 and δ_2 is a model selection rather than a variable selection for and we adopt a similar selection criterion to [Fan & Yao (2002)]. Using the observation that a different linear model holds at each time point in a varying coefficient model, we select the model that minimizes AIC for each time point. We pick δ_1, δ_2 such that AIC is minimized for most of the time points.

Simulation

- In the simulation study, we are mainly interested in two issues: the finite sample property of the proposed estimator in comparison with that of Malfait & Ramsay and the performance of the proposed parameter selection criterion.
- The two goals need to be addressed by two separate simulations.
- For the model assessment, we define NIE (normalized integrated error) as

$$NIE = \frac{\int \sqrt{\sum_{i=1}^{n} \{\hat{y}_i^{(-i)}(t) - y_i(t)\}^2 dt}}{\int \sqrt{\sum_{i=1}^{n} \{y_i(t)\}^2 dt}}. \qquad (10)$$

NIE is based on one subject leave out cross validation criterion.

Simulation Data

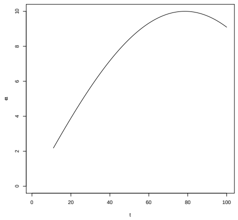

Figure: Intercept function $\alpha(t)$

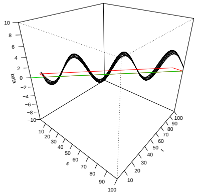

Figure: Regression function $\beta(s,t)$

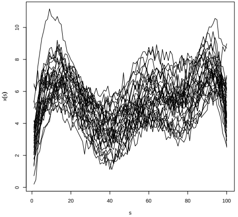

Figure: Predictor function

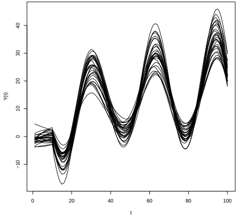

Figure: Response function

Simulation Result-The NIE comparison

- The true window size is set to be $\delta_1 = 10, \delta_2 = 0$.
- The average NIEs from 500 monte carlo runs are reported.
- The number in the parentheses is the stardard deviation of NIE.

Table: NIE for the proposed and the MR method.

			$n = 30$	$n = 100$	$n = 300$
Proposed method		$\sigma^2 = 3$	0.1191 (0.0075)	0.1152 (0.0041)	0.1139 (0.0019)
		$\sigma^2 = 5$	0.1188 (0.0082)	0.1145 (0.004)	0.1133 (0.0019)
MR method	N=10	$\sigma^2 = 3$	0.1462 (0.0109)	0.1385 (0.0058)	0.1360 (0.0033)
		$\sigma^2 = 5$	0.1667* (0.014)	0.1493 (0.0113)	0.1395 (0.077)
	N=20	$\sigma^2 = 3$	0.1528 (0.013)	0.1406 (0.0070)	0.1388* 0.0037
		$\sigma^2 = 5$	0.1753* (0.016)	0.159 (0.0919)	0.1548* (0.005)
	N=40	$\sigma^2 = 3$	0.1649* (0.0146)	0.1457* (0.0066)	0.1427* (0.0037)
		$\sigma^2 = 5$	0.188* (0.021)	0.1632* (0.0095)	0.159* (0.0056)

Simulation Result-The performance of model selection

- The true window size is set to be $\delta_1 = 10, \delta_2 = 5$.
- Six different combinations of δ_1 and δ_2 are considered as candidates and one of six combinations yielding the smallest AIC for most of the time points is chosen. The candidate combinations are given in Table 2.
- The ratio of correct choice of δ_1 and δ_2 from 500 Monte Carlo simulations for each sample size are reported in Table 3.

Table: The candidate combinates for δ_1 and δ_2.

δ_1	20	20	20	10	10	5
δ_2	0	5	10	5	0	0

Table: The ratio of the correct choice for δ_1 and δ_2 from 500 monte carlo simulation runs.

	$n = 30$	$n = 100$	$n = 300$
Ratio	0.74	0.83	0.99

Thank you!!

Supports for $\beta(s,t)$

Return

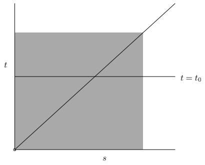

Figure: Suppport for FLM

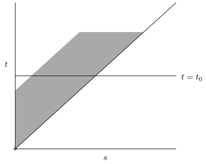

Figure: Support for HFLM

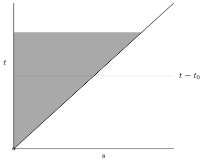

Figure: Support for HFLM

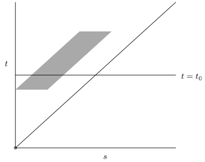

Figure: Support for the proposed model

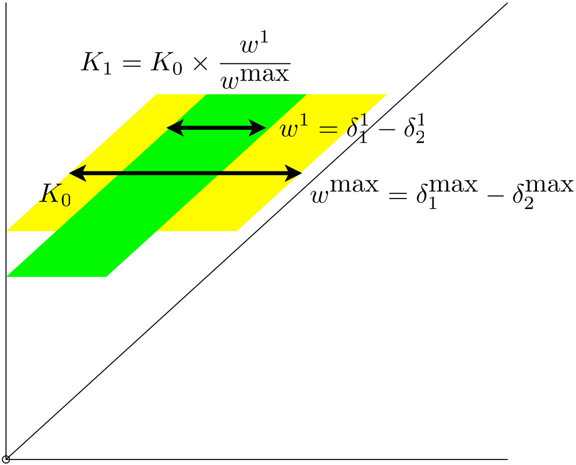

Figure: Selection of K

Return

Fan, J., Yao, Q., and Cai, Z.
"Adaptive varying-coefficient linear models."
Journal of the Royal Statistical Society Series B(. . . (2003).
URL http://www.blackwell-synergy.com/doi/abs/10.1111/1467-9868.00372

Fan, J. and Zhang, J.
"Two-step estimation of functional linear models with applications to longitudinal data."
J Roy Stat Soc B, 62:303–322 (2000).

Malfait, N. and Ramsay, J.
"The Historical Functional Linear Model."
The Canadian Journal of Statistics/La Revue Canadienne de . . . (2003).
URL http://links.jstor.org/sici?sici=0319-5724(200306)31%253A2%253C115%253ATHFLM%253E2.0.CO%253B2-L

Muller, H. and Zhang, Y.
"Time-varying functional regression for predicting remaining lifetime distributions from longitudinal trajectories."
Biometrics, 61:1064–1075 (2005).

Ramsay, J.
"0. and Silverman, BW (1997)."

Functional Data Analysis (1997).