Recent History Functional Linear Model

Kion Kim, Damla Şentürk, Runze Li
Department of Statistics, Pennsylvania State University

Summary

We propose a variant of historical functional linear models for cases where the current response is affected by the predictor process in a window into the past. Different from the rectangular support of functional linear models, the triangular support of the historical functional linear models and the point-wise support of the varying coefficient models, the current model has a sliding window support into the past. This idea leads to models that bridge the gap between varying coefficient models and functional linear (historic) models. The proposed estimation algorithm is shown to be fast involving one dimensional basis expansions and one dimensional smoothing procedures.

Functional Linear Model

Functional data refers to data that have dense repetitions on each subject. The basic philosophy of functional data is to think of the dense repetitions as realizations of a continuous function. The most general regression model for functional data is the functional linear model.

Let $x_i(s), s \in [0, S]$ be the predictor function and $y_i(t), t \in [0, T]$ be the response function, where $i = 1, \cdots, n$ denotes the number of subjects. Assuming the predictor function affects the response function through a linear model, Cardot, H., Ferraty, F. and Sarda, P. (1999) among many others considered the functional linear model

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

where $\alpha(t)$ is the intercept function, $\beta(s,t)$ is the bivariate coefficient 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.
Usually the estimation of coefficient function $\beta(s,t)$ is done by two-dimensional basis expansion or penalized least square method.

Historical Linear Model

In the case of non-periodical data, from the feed-forward point of view, it is not plausible to use future values of the covariate function in predicting the current value of the response function.

Malfait & Ramsay (2003) tailored the model in (1) to be applicable to the situation where the outcome $y(t)$ depends only on the predictors $x(s)$ at time $s \le t$ 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 $0 \le \delta \le t$.
While the functional linear model had rectangular support, considering past, present and future values of the predictor process, the historical functional linear model has triangular support, considering the effect of the past and present of the predictor process. They estimated the model using finite elements method (FEM) based on two dimensional basis functions on triangular elements.

Proposed 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,

$$y_i(t) = \alpha(t) + \int_{t-\delta_1}^{t-\delta_2} x_i(s)\beta(s,t)ds + \varepsilon_i(t), \quad t \in [\delta_1, T], s \in [0, T], 0 \le \delta_2 \le \delta_1 \le t. \tag{3}$$

Here δ_1 allows for a lag for the predictor process to start affecting the response, particularly useful for prediction models, while δ_2 denotes another lag, beyond which the predictor function does not affect the response. 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, $\phi_k(\cdot)$, resulting in $\beta(s,t) \approx \sum_{k=1}^K b_k\phi_k(s)$, where $s \in [t - \delta_1, t - \delta_2]$.

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 (3) simplifies to

$$y_i(t) = \alpha(t) + \sum_{k=1}^K \psi_{i,k}(t)b_k(t) + \varepsilon_i(t), \quad t \in [\delta_1, T]. \tag{4}$$

where $b_k(t)$ denotes the coefficient of the k^{th} basis function at time t.
The resulting model is a **varying coefficient model**, for which many estimation methods have been proposed. (Wu and Yu, 2002). We are going to make use of the **two-step estimator**. (Fan & Zhang, 2000)

Estimation

The estimation of model in equation (3) consists of three steps as follows. Note that we use cubic B-spline basis for the expansion of the regression function.

- **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.
- **Raw estimate** of $b(t)$: If we define $\boldsymbol{y}(t)$ as an n dimensional vector, which contains the n realizations of the response variable on time point t, and $\tilde{\boldsymbol{\Psi}}(t) = [1, \hat{\boldsymbol{\Psi}}(t)]$, where $\hat{\boldsymbol{\Psi}}(t)$ is a matrix with $\hat{\psi}_{ij}(t)$ as its $(i,j)^{th}$ value, the raw estimates, $\tilde{\alpha}(t)$ and $\tilde{\boldsymbol{b}}(t) = [\tilde{b}_1(t), \cdots, \tilde{b}_K(t)]$, are

$$\begin{pmatrix} \tilde{\alpha}(t) \\ \tilde{\boldsymbol{b}}(t) \end{pmatrix} = \left[\tilde{\boldsymbol{\Psi}}(t)^T\tilde{\boldsymbol{\Psi}}\right]^{-1}\tilde{\boldsymbol{\Psi}}(t)^T\boldsymbol{y}(t), \quad t \in [\delta_1, T] \tag{5}$$

- **Refined estimator** of $b(t)$: If a linear smoother with weight function $w_l^k(t)$ is used, the refined estimator, $\hat{b}_k, k = 1, \cdots, K$, can be given as

$$\hat{b}_k(t) = \sum_{l \in [\delta_1, T]} w_l^k(t)\tilde{b}_k(l) \tag{6}$$

Hence the bivariate regression function, $\beta(s,t)$, is recovered by $\hat{\beta}(s,t) = \sum_{i=1}^K \phi_k(s)\hat{b}_k(t)$.

Knots, window size and model selection

- We adapt root squared prediction error (RSPE) from Müller and Zhang (2005) to select the appropriate number of knots and model. A version RSPE 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 $\hat{y}_i^{(-i)}(t)$ is predicted value of the response function at time t without i^{th} subject.
- The error rate over time t is reported using normalized integrated error (NIE) defined by

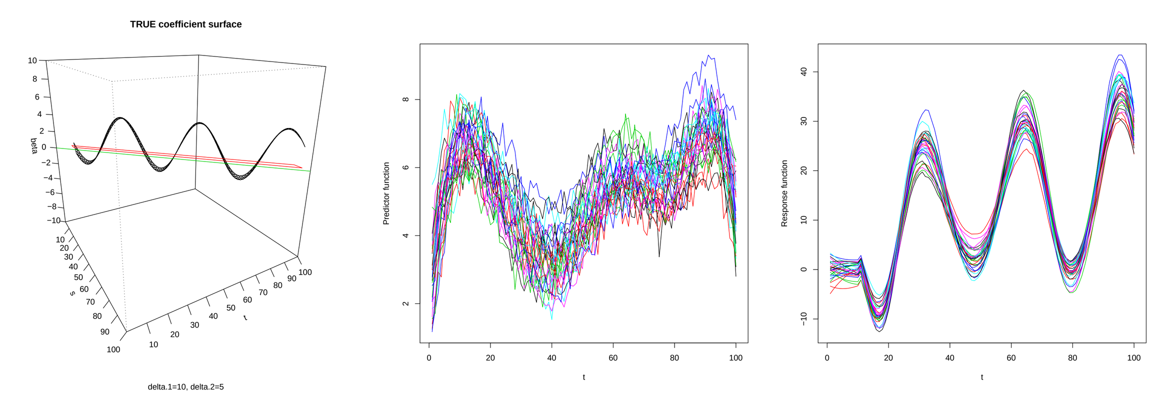
$$NIE = \frac{\int \sqrt{\frac{1}{n}\sum_{i=1}^n \left(\hat{y}_i^{(-i)}(t) - y_i(t)\right)^2}dt}{\int \sqrt{\frac{1}{n}\sum_{i=1}^n y_i(t)^2}dt}$$

- **Number of knots**: We select the number of knots for any given window size using the following two steps.
  - We decide optimal number of knots for the largest window we consider at first.
  - We average RSPE over t and pick the number of knots, K, which yields the smallest average RSPE.
  - we assume that the number of knots relevant to approximate the regression function for shorter windows is proportional to the windows relative length compared to the largest window considered at the beginning of the algorithm.
- **Window size**: The choice of δ_1 and δ_2 is equivalent to model selection in the varying coefficient models.
  - For the model selection of varying coefficient model, we adopt the idea from Fan et al. (2003).
  - We can compare different varying coefficient models at each fixed time point by using AIC.
  - We can pick the best combination of δ_1 and δ_2 by selecting (δ_1, δ_2) combination, which is chosen to be best in terms of AIC for the most of time points.

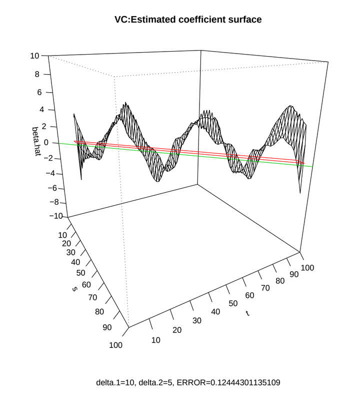
Preliminary Simulation

Data generation

- For this simulation, we generated 100 time points, i.e., $t = 1, \cdots, 100$.
- The true δ_1 and δ_2 are set as 10 and 5 respectively.
- The regression function, $\beta(s,t)$, is generated by the function, $\beta(s,t) = 1.5 \cdot \sin\left(\frac{s}{5} - 10\right) + .3 \cdot \cos\left(\frac{t}{5}\right) - 1$ for $s \in [t - \delta_1, t - \delta_2]$.
- The intercept function, $\alpha(t)$, is generated by the function, $\alpha(t) = 10\sin(2t) \cdot I\left(t > \frac{\delta_1}{100}\right)$.
- We generate the predictor function via $X_i(t) = \phi(t)\gamma_i + e_i, \gamma_i \sim MVN([3,8,6,2,7,4,9,5]^T, \boldsymbol{I})$, $i = 1, \cdots, n$, where $\phi(t)$ is a set of cubic B-spline basis function at time t and γ_i's are corresponding coefficients. The additive error e_i is distributed as $N(0, 0.1)$.
- The error function, $\varepsilon_i(t)$, is generated also as a linear combination of the above 8 bases with different mean, $\mu = \boldsymbol{0}$ and variance, $\sigma^2 = 5\boldsymbol{I}$.
- We generated different number of subjects: $n = 30, 100$ and 300.
- For different number of subjects, we simulated 100 dataset and reported average NIE and model selection results.



1. **Estimation result**: Here is one of the estimated regression function when $n = 30$.



2. **NIE**: The error results are

	30	100	300
Average NIE	0.1224	0.1205	0.1174

3. **Model selectiont result**

	30	100	300
Ratio	0.73	0.84	0.99

- The NIE is decreased as the number of subjects increases.
- The model selection criterion based on pointwise AIC works better as the number of subject increases.

Reference

Cardot, H., Ferraty, F. and Sarda, P. (1999), "Functional Linear Model. Statist," *Stat. & Prob. Letter*, 45, 11-22.

Fan, J., and Zhang, J. (2000), "Two-Step Estimation of Functional Linear Models With Applications to Longitudinal Data," *Journal of the Royal Statistical Society, Ser. B*, 62, 303-322.

Fan et al. (2003), "Adaptive varying-coecient linear models," *Journal of the Royal Statistical Society, Ser. B*, 65, 2245.

Malfait N., and Ramsay, J. O. (2003), "The Historical Functional Linear Model," *The Canadian Journal of Statistics*, 31, 115-128.

Müller, H. G., and Zhang, Y. (2005), "Time-varying functional regression for predicting remaining lifetime distributions from longitudinal tra jectories." *Biometrics*, 61, 10641075.

Wu, C. O. and Yu, K. F. (2002), "Nonparametric varying coefficient models for the analysis of longitudinal data," *International Statistical Review*, 70, 373-393.