Recent History Functional Linear Model for Longitudinal Data

PENNSTATE 1855

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 coeffcient models, the current model has a sliding window support into the past. This idea leads to models that bridge the gap between varying coeffcient models and functional linear (historic) models. We propose an algorithm for this model that can be applied to longitudinal data where the measurements are taken on irregular time points and missing values are allowed. The proposed estimation algorithm is shown to be fast, involving one dimensional basis expansions and one dimensional smoothing procedures.

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) [1] 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 $\beta(s,t)$ utilizes two-dimensional basis expansions and is fit by least squares or penalized least squares method.

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,

$$y_i(t) = \alpha(t) + \int_{t-\delta_1}^{t-\delta_2} x_i(s)\beta(s,t)ds + \varepsilon_i(t), \quad t \in [\delta_1, T], s \in [0,T], 0 \le \delta_2 \le \delta_1 \le T. \tag{2}$$

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 $s \in [t-\delta_1, t-\delta_2]$. Note that we use B-spline basis for the expansion of the regression function. Defining $\psi_{i,k}(t) = \int_{t-\delta_1}^{t-\delta_2} x_i(s)\phi_k(s-t+\delta_1)ds$, the proposed model in (2) simplifies to

$$y_i(t) = \alpha(t) + \sum_{k=1}^K \psi_{i,k}(t)b_k(t) + \varepsilon_i(t), \quad t \in [\delta_1, T], \; i = 1, \ldots, n. \tag{3}$$

t where $b_k(t)$ denotes the coefficient of the k^{th} basis function at time t.
The resulting model is a **varying coefficient model (VCM)**, for which many estimation methods have been proposed. (Wu and Yu, 2002)[7]. There are three approaches to estimate the varying coefficient functions: kernel-local polynomial smoothing, polynomial spline and smoothing spline. As explained in Fan and Zhang (2008)[3], kernel-local polynomial fit is the most natural approach since the VCM is a local model.

Estimation

The estimation method of the model in (2) for the functional data is proposed in Kim, Şentürk and Li (2009).[5] Here, we extend it to irregular longitudinal data where the location and total number of measurements vary among subjects. The estimation of the regression function, $\beta(s,t)$, in equation (2) consists of three steps.

- **Estimation of $\psi_{ik}(t)$**: The estimation of $\psi_{ik}(t)$ involves the numerical integration. For longitudinal data, due to the irregularity of data, one can obtain $\hat{\psi}_{ik}(t)$ only for the time points where a certain number of measurements are available. We use trapezoidal rule for the integration.
- **Estimation of $b(t)$**: For the estimation of VCM, the local linear fit proposed by Hoover et al. (1998)[4] is employed. The estimators, $\hat{b}_{lk}, \; k = 0, \ldots, K, \; l = 0, 1$, are chosen to minimize

$$Q(\boldsymbol{b}(t)) = \sum_{i=1}^n \sum_{j=1}^{n_i} \left\{ Y_{ij} - \sum_{k=1}^K \hat{\psi}_{ik}(T_{ij}) \left\{ \sum_{l=0}^1 b_{lk}(t)(T_{ij}-t)^l \right\} \right\}^2 K_h(T_{ij}-t), \tag{4}$$

where $K_h(\cdot)$ is a kernel function defined by $K_h(a) = K(a/h)/h$.

- **The recovery of $\beta(s,t)$**: The estimated bivariate regression function can be obtained by $\hat{\beta}(s,t) = \sum_{i=1}^K \phi_k(s)\hat{b}_k(t)$.

Choosing the degree of smoothness and window size

There are three sets of parameters that need to be chosen: the number of basis functions for the approximation, K, the bandwidth for the estimation of VCM, h, and the window where the predictors effect the response, (δ_1, δ_2). Note that K and h determine the resolution of the model.
The selection of those quantities can be performed via a two step procedure:

1. **Choose the optimal K and h combination for the maximum window size**.
2. **Choose the optimal (δ_1, δ_2) using K and h selected in step (1).**

The following **criteria** are used for each step.

- We adapt root squared prediction error (RSPE) from Müller and Zhang (2005)[6] to select the appropriate K and h. A version of RSPE is defined by

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

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 error rate over time t is reported using normalized integrated error (NIE) defined by

$$NIE = \frac{\sum_{i=1}^n \sqrt{\sum_{j=1}^{n_i} \{\hat{y}_i^{(-i)}(t_{ij}) - y_i(t_{ij})\}^2}}{\sum_{i=1}^n \sqrt{\sum_{j=1}^{n_i} y_i^2(t_{ij})}}.$$

The **selection strategy** is given as follows.

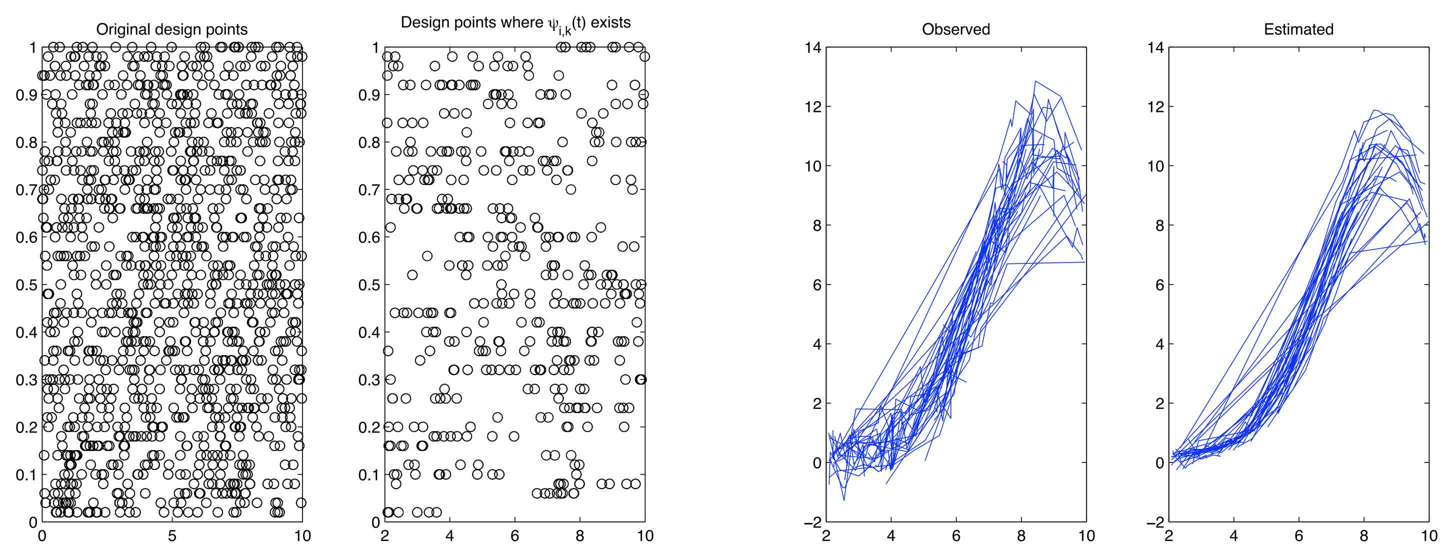
- ***K* and *h***: One can select K and h for any given window size based on K and h for the maximum window.
  - For the maximum window size of interest, we pick the K and h combination, which yields the smallest RSPE.
  - We assume that the number of knots relevant to approximate the regression function for shorter windows is proportional to the windows relative length compared to the largest window considered at the beginning of the algorithm.
  - It is assumed that the optimal bandwidth does not depend on the size and location of window.
- **(δ_1, δ_2)**: The choice of δ_1 and δ_2 is equivalent to model selection in the VCM.
  - For the model selection of VCM, we adopt the idea from Fan et al. (2003)[2].
  - Different VCMs are compared at each fixed time point by using AIC. The best (δ_1, δ_2)-combination is picked by selecting the one chosen to be the best in terms of AIC for the most of time points. For longitudinal data, due to the irregularity of data, we use bins and consider the measurements inside a bin as cross-sectional observations satisfying i.i.d. condition.

Simulation

Data generation

- The number of design points for the i^{th} subject, n_i, where realizations of the predictor and response stochastic processes take places were independently sampled from discrete uniform distribution, $DU[20, 25]$, for different subjects.
- Design points are generated from $U[0, 10]$ as many as n_i's.
- On those design points, n predictor functions are generated using the mean function $t + \sin(t)$ and the B-spline basis of order 4 need knot sequence $[2, 4, 6, 8]$ on the support of $[0, 10]$ via $x_i(t) = t + \sin(t) + \sum_{q=1}^8 b_{iq}B_q(t), \; i = 1, \ldots, n, \; t \in [0, 10]$.
  Note that 8 basis functions are used in predictor generation process. Coefficients, $b_{iq}, \; q = 1, \ldots, 8$, are independently generated from $N(0, 1)$.
- The regression function is generated by two basis functions, $\psi_1(t) = -\cos(\pi t/10)/\sqrt{5}$ and $\psi_2(t) = \sin(\pi t/10)/\sqrt{5}$, via the equation, $\beta(s,t) = \sum_{i=1}^2 \sum_{j=1}^2 c_{ij}\psi_i(s)\psi_j(t), \; s \in [t-\delta_1, t-\delta_2]$, where $c_{11} = 2, c_{12} = 2, c_{13} = 1$ and $c_{14} = 2$ and $(\delta_1, \delta_2) = (2, 1)$.
- Error functions are generated using the same basis functions as those for predictor functions with the coefficients of basis functions sampled from $N(0, \sigma_\varepsilon^2)$.
- The response functions are generate by the equation, $\int_{t-\delta_1}^{t-\delta_2} \beta(s,t)x_i(s)ds + \varepsilon(t)$.

Estimation result: One of the estimation results when $n = 50$ is given in the following figures. The B-spline basis of order 4 with 3 interior knots were used for the approximation of the regression function and 20% of data points were used for local linear fitting. The figure on the left shows design points and the one on the right shows the responses and their fitted value.



Window size selection: The proportion of the correct selection for window combination based on RSPE from 600 simulations are reported in the following table. The results are given for two different number of subjects, $n = 30, 50$ and three different error variances, $\sigma_\varepsilon^2 = 0.2^2, 0.5^2$ and 0.8^2.

n \ σ_ε^2	0.2^2	0.5^2	0.8^2
30	0.9448	0.8512	0.32
50	1	1	0.765

Table 1: (δ_1, δ_2)-Window selection.

Comparison: The proposed model is compared with VCM and Functional Principal Component Regression (FPCR) proposed by Yao, Müller and Wang (2005)[8]. The average NIEs from 200 simulation data are summarized in the following table. The standard error of NIE is also provided in the parenthesis.

	RHFLM	VCM	FPCR
Mean NIE	0.1088	0.2826	0.2282
(Std. dev.)	(0.0056)	(0.0382)	(0.0436)

Table 2: The model comparison

Bias and Variance: The Bias and variance of the estimator is calculated for $K = 1$ and $h = 1$, with which MISE is minimized. The result is reported based on 500 simulation.

$MISE$	$Bias^2$	Var
0.014	0.0068	0.0072

Table 3: Bias, variance and MISE.

References

[1] H. Cardot, F. Ferraty, and P Sarda. Functional linear model. *Statistics and Probability Letters*, Jan 1999.

[2] J. Fan, Q. Yao, and Z. Cai. Adaptive varying-coefficient linear models. *Journal of the Royal Statistical Society Series B*, Jan 2003.

[3] J. Fan and W. Zhang. Statistical methods with varying coefficient models. *Statistics and its interface*, 1(1):179, 2008.

[4] D. HOOVER, J. Rice, C. Wu, and L. YANG. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*, Jan 1998.

[5] K. Kim, D. Senturk, and R. Li. The recent history functional linear models. *Technical report*, 2009.

[6] H.G. Muller and Y. Zhang. Time-Varying Functional Regression for Predicting Remaining Lifetime Distributions from Longitudinal Trajectories. *Biometrics*, 61(4):1064–1075, 2005.

[7] C. Wu and K. Yu. Nonparametric varying coefficient models for the analysis of longitudinal data. *International Statistical Review*, Jan 2002.

[8] F. Yao, H. Müller, and W. Wang. Functional linear regression analysis for longitudinal data. *Annals of Statistics*, Jan 2005.