Statistica Sinica **34** (2024), 1-19
doi:https://doi.org/10.5705/ss.202022.0196

LOCALLY SPARSE ESTIMATOR OF GENERALIZED VARYING COEFFICIENT MODEL FOR ASYNCHRONOUS LONGITUDINAL DATA

Rou Zhong[1], Chunming Zhang[2] and Jingxiao Zhang*[1]

[1] *Renmin University of China and* [2] *University of Wisconsin-Madison*

Abstract: In longitudinal studies, it is common that the response and the covariate are not measured at the same time, which complicates the subsequent analysis. In this study, we consider the estimation of a generalized varying coefficient model with such asynchronous observations. We construct a penalized kernel-weighted estimating equation using the kernel technique in a functional data analysis framework. Moreover, we consider local sparsity in the estimating equation to improve the interpretability of the estimate. We extend the iteratively reweighted least squares algorithm in our computation, and establish the theoretical properties of the proposed method, including the consistency, sparsistency, and asymptotic distribution. Lastly, we use simulation studies to verify the performance of our method, and demonstrate the method by applying it to data from a study on women's health.

Key words and phrases: Asynchronous observation, functional data analysis, generalized varying coefficient model, kernel technique, local sparsity.

1. Introduction

A generalized varying coefficient model (Hastie and Tibshirani (1993); Cai, Fan and Li. (2000)) allows the coefficients to vary over time, significantly widening the application of regression models. Specifically, the model can be expressed as

$$E\{Y(t)|X(t)\} = g\{\beta_0(t) + \beta_1(t)X(t)\}, \quad t \in \mathcal{T}, \tag{1.1}$$

where $Y(t)$ is the response, $X(t)$ is the covariate, $g(\cdot)$ is a known strictly increasing and continuously twice-differentiable link function, $\beta_0(t)$ is the intercept function, $\beta_1(t)$ is the varying coefficient function, and $\mathcal{T}$ is a bounded and closed interval. Here, we propose a new estimating method for a generalized varying coefficient model with longitudinal measurements, from the perspective of functional data.

In practice, it often happens that the covariate and the response are not measured at the same time for each subject in longitudinal observations. Such asynchronous observations make the subsequent analysis more complicated. Two main types of approaches have been proposed to solve this problem. The first

*Corresponding author.