Statistica Sinica **33** (2023), 1-32
doi:https://doi.org/10.5705/ss.202020.0294

MODELING SPIKY FUNCTIONAL DATA WITH DERIVATIVES OF SMOOTH FUNCTIONS IN FUNCTION-ON-FUNCTION REGRESSION

Ruiyan Luo and Xin Qi

Georgia State University

Abstract: Smoothness penalties are efficient regularization and dimension reduction tools for functional regressions. However, for spiky functional data observed on a dense grid, the coefficient function in a functional regression can be spiky and, hence, the smoothness regularization is inefficient and leads to over-smoothing. We propose a novel approach to fit the function-on-function regression model by viewing the spiky coefficient functions as derivatives of smooth auxiliary functions. Compared with the smoothness regularization or sparsity regularization imposed directly on the spiky coefficient function in existing methods, imposing smoothness regularization on the smooth auxiliary functions can more efficiently reduce the dimension and improve the performance of the fitted model. Using the estimated smooth auxiliary functions and taking derivatives, we can fit the model and make predictions. Simulation studies and real-data applications show that compared with existing methods, the new method can greatly improve model performance when the coefficient function is spiky, and performs similarly well when the coefficient function is smooth.

Key words and phrases: Auxiliary function, derivative, function-on-function regression, smoothness regularization, spiky functional data

1. Introduction

The function-on-function (FOF) linear regression model is a useful tool for studying the association between functional variables. The past two decades have witnessed the development of methods to fit the FOF model for relatively smooth functional data observed on a moderately sized grid. With the development of technology, densely observed curves have been collected in different areas, and usually display complex local features. For example, spectrum curves contain a number of narrow and high peaks, whereas electroencephalography time series curves exhibit high local variations over the entire time interval. When applying the FOF model to these spiky curves, assuming the coefficient functions to be

Corresponding author: Ruiyan Luo, Department of Population Health Sciences, School of Public Health, Georgia State University, Atlanta, GA 30302-3965, USA. E-mail: rluo@gsu.edu.