Statistica Sinica **36** (2026), 1-20
doi:https://doi.org/10.5705/ss.202023.0356

INFERENCE ON LARGE-SCALE GENERALIZED FUNCTIONAL LINEAR MODEL

Kaijie Xue and Riquan Zhang*

Shanghai University of International Business and Economics

Abstract: In this work, we extend the classical generalized functional linear model to a large-scale generalized functional linear model to handle a variety of complex situations where the response (possibly discrete) can be nonlinearly linked to an ultra-high number of functional predictors. Unlike most existing requirements on functional data, we don't need to impose any conditions regarding eigenvalue-decay or square-integrability on those functional predictors, resulting in a more flexible but challenging model framework. Based on a penalized model estimator, we develop a general inferential method to assess the significance of an arbitrary group of regression curves. Concretely, a pseudo score function is adopted to construct the associated confidence region for the regression curves of interest. Notably, the proposed test is justified uniformly convergent to nominal level, without any demand on estimation consistency of the regression curves. Finally, numerical studies are carried out to show the empirical performance of the proposed test.

Key words and phrases: Eigenvalue-decay-free, estimation-consistency-relaxed, high dimensions, multiplier bootstrap, square-integrable-free.

1. Introduction

A series of work (Ramsay and Dalzell, 1991; Yuan and Cai, 2010; Malfait and Ramsay, 2003; Fan and Zhang, 2000; Cardot, Ferraty and Sarda, 1999) have been devoted to the study of classical functional linear model (FLM) containing a single functional predictor, focusing on either theoretical basis (Hall and Horowitz, 2007; Cai and Yuan, 2012; Ramsay and Silverman, 2005) or inferential methods (Cardot et al., 2003; Shang and Cheng, 2015; Lei, 2014; Hilgert, Mas and Verzelen, 2013; Zhang and Chen, 2007). As an important extension of FLM, the generalized functional linear model (GFLM) has been frequently employed to model the more complicated (possibly nonlinear) association between a response Y and a functional predictor $X(t) \in L^2(T)$, where the random process $X(\cdot)$ is defined and square-integrable on a compact subset $T \subseteq \mathbb{R}$. This model has been intensively studied by many articles (Müller and Stadtmüller, 2005; Shang and Cheng, 2015; Escabias, Aguilera and Valderrama, 2004). Concretely, given a sample of n i.i.d. pairs $\{Y_i, X_i(\cdot)\}$, the conditional density of the classical GFLM under the commonly-used canonical link belong to an exponential family, which takes the

*Corresponding author. E-mail: zhangriquan@163.com