

ESTIMATION AND VARIABLE SELECTION UNDER THE FUNCTION-ON-SCALAR LINEAR MODEL WITH COVARIATE MEASUREMENT ERROR

Yifan Sun and Grace Y. Yi*

University of Western Ontario

Abstract: Function-on-scalar linear regression has been widely used to model the relationship between a functional response and multiple scalar covariates. Its utility is, however, challenged by the presence of measurement error, a ubiquitous feature in applications. Naively applying usual function-on-scalar linear regression to error-contaminated data often yields biased inference results. Further, estimation of the model parameters is complicated by the presence of inactive variables, especially when handling data with a large dimension. Building parsimonious and interpretable function-on-scalar linear regression models is in urgent demand to handle error-contaminated functional data. In this paper, we study this important problem and investigate the measurement error effects. We propose a debiased loss function, combined with a sparsity-inducing penalty function, to simultaneously estimate functional coefficients and select salient predictors. An efficient computing algorithm is developed with tuning parameters determined by data-driven methods. Under mild conditions, the asymptotic properties of the proposed estimator are rigorously established, including estimation consistency, selection consistency, and the limiting distributions. The finite sample performance of the proposed method is assessed through extensive simulation studies, and the usage of the proposed method is illustrated by a real data application.

Key words and phrases: Estimation, function-on-scalar regression, functional data analysis, measurement error, variable selection.

1. Introduction

Functional data analysis has attracted extensive attention over the past two decades (Ramsay and Silverman, 2005; Horváth and Kokoszka, 2012). Typically, the function-on-scalar linear regression model (Ramsay and Silverman, 2005, Ch. 13) has been used widely to describe the relationship between a functional response and multiple scalar covariates, and many methods have been proposed for inference under this model. To name a few, see Chiang, Rice and Wu (2001), Ramsay and Silverman (2005, Sec. 13.4), Zhang and Chen (2007), Zhu, Li and Kong (2012), and the references therein.

Recently, research on variable selection has become increasingly interesting, which is paramount in the era of big data. With data of a large dimension,

*Corresponding author. E-mail: gyi5@uwo.ca