Chapter 26
Influence in the Functional Linear Model with Scalar Response

Manuel Febrero, Pedro Galeano and Wenceslao González-Manteiga

Abstract This paper studies how to identify influential curves in the functional linear model in which the response is scalar and the predictor is functional and how to measure their effects on the estimation of the model and on the forecasts, when the model is estimated by the principal components method. For that, we introduce and analyze two statistics that measure the influence of each curve on the functional slope estimate of the model, which are generalizations of the measures proposed for the standard regression model by Cook (1977) and Peña (2005), respectively.

26.1 Introduction

The collection of data which consists of repeated measurements of the same subject densely taken over an ordered grid of points belonging to a finite length interval is becoming progressively frequent. Data of these characteristics are usually called functional data, because even though the recording points are really discrete, we may assume that the entire function has been completely observed. It is well known that multivariate statistical methods are not well suited for functional data for several reasons. For instance, multivariate statistical methods ignore the time correlation structure of functional
data. Thus, there exists a demand for suitable procedures to analyze such data. The books of Ramsay and Silverman (2004, 2005) and Ferraty and Vieu (2006) are texts of reference and summarize several methods and case studies for handling functional data from different approaches.

In the recent literature, functional linear models in which the predictors and/or the response are of a functional nature have received considerable attention. This paper deals with the functional linear model with scalar response in which the predictor is functional. Several approaches have been proposed for estimating the functional parameter of the functional linear model with scalar response. For instance, Hastie and Mallows (1993), Marx and Eilers (1999), Cardot, Ferraty and Sarda (2003) and Ramsay and Silverman (2005) have analyzed the use of restricted basis functions and penalization methods. Ferraty and Vieu (2006) have proposed the use of nonparametric estimates based on kernels. Cardot, Ferraty and Sarda (1999) proposed a least-squares estimate based on functional principal components, which has been further analyzed in Cardot, Ferraty and Sarda (2003), Hall and Hosseini-Nasab (2006), Cai and Hall (2006) and Hall and Horowitz (2007), among others.

As any other statistical data, influential observations may be sometimes found in functional datasets. The aim of this paper is to analyze influence in the functional linear model with scalar response. In particular, we study how to identify curves with larger influence on the estimation of the functional parameter of the model and how to measure their effects on the estimation. For that, we propose two statistics that seem to be useful in detecting which curves have strong influence on the estimated slope. These statistics are the generalization to functional data of the measures proposed by Cook (1977) and Peña (2005) for the standard linear regression model. We use bootstrap methods to calibrate the distribution of these statistics, which allow us to detect the presence of influential observations. No much is known about influence in functional models. Only Shen and Xu (2007) and Chiu and Müller (2007) have introduced functional versions of the Cook distance in the case in which the predictors are real and the responses are functional, and in the case in which both the predictors and the responses are functional, respectively. Both models are different that the one considered here.

As mention previously, there are several ways to estimate the functional linear model with scalar response. The approach taken in this paper is based on the functional principal components technique, which has become very popular. Although it is well known that this estimator may be rough even for large sample sizes and alternative more smoother estimates have been proposed, we show that the estimator based on functional principal components provides a natural framework to analyze influence. Nevertheless, the results derived in this paper can be generalized to alternative smoothing estimators in a simple way.

The rest of this abstract is as follows. Section 2 presents the functional linear model with scalar response and reviews estimation based on the func-