Statistica Sinica **34** (2024), 1-23
doi:https://doi.org/10.5705/ss.202022.0086

AN RKHS APPROACH FOR PIVOTAL INFERENCE IN FUNCTIONAL LINEAR REGRESSION

Holger Dette* and Jiajun Tang

Ruhr-Universität Bochum

Abstract: We use a reproducing kernel Hilbert space approach to develop a methodology for testing hypotheses about the slope function in a functional linear regression for time series. In contrast to most existing studies, which tests for the exact nullity of the slope function, we are interested in the null hypothesis that the slope function vanishes only approximately, where deviations are measured with respect to the L^2-norm. We propose an asymptotically pivotal test that does not require estimating nuisance parameters or long-run covariances. The key technical tools that we use to prove the validity of our approach include a uniform Bahadur representation and a weak invariance principle for a sequential process of estimates of the slope function. Lastly, we demonstrate the potential of our methods using a small simulation study and a data example.

Key words and phrases: Functional linear regression, functional time series, m-approximability, relevant hypotheses, reproducing kernel Hilbert space, self-normalization.

1. Introduction

Numerous statistical methods exist for analyzing functional data; see Ramsay and Silverman (2005), Ferraty and Vieu (2010), Horváth and Kokoszka (2012), Hsing and Eubank (2015), and Wang, Chiou and Müller (2016). Because of its good interpretability, the functional linear regression model

$$Y_i = \int_0^1 X_i(s)\,\beta_0(s)\,ds + \varepsilon_i, \quad i \in \mathbb{Z}, \tag{1.1}$$

has become a useful tool for functional data analysis (e.g., see Cardot, Ferraty and Sarda (1999); Müller and Stadtmüller (2005); Yao, Müller and Wang (2005); Hall and Horowitz (2007); Yuan and Cai (2010)). In our study, $\{(X_i, \varepsilon_i)\}_{i\in\mathbb{Z}}$ denotes a strictly stationary time series, where X_i is a mean zero square-integrable random function on the interval $[0, 1]$, and ε_i is a real-valued centered random noise.

Because the slope function β_0 characterizes the dependence between the predictor and the response, many studies have focused on its estimation and corresponding statistical inference. A popular method for analyzing the slope function in model (1.1) is to use functional principle components (FPCs) (e.g.,

*Corresponding author.