Statistica Sinica **33** (2023), 1-21
doi:https://doi.org/10.5705/ss.202021.0078

A STRATIFIED PENALIZATION METHOD FOR SEMIPARAMETRIC VARIABLE LABELING OF MULTI-OUTPUT TIME-VARYING COEFFICIENT MODELS

Ting Zhang[1], Weiliang Wang[2] and Yu Shao[2]

[1] *University of Georgia and* [2] *Boston University*

Abstract: In a time-varying coefficient model, the regression coefficient is allowed to change over time as a nonparametric function to capture the time-varying feature. Owing to its popularity in time series applications, where the assumption of independence typically does not hold, it is desirable to allow dependent and nonstationary observations. We consider the problem of semiparametric variable labeling and estimation for multi-output time-varying coefficient models in a time series setting, where a variable can be labeled as time-varying, time-constant, or irrelevant, in a nested structure. We first show that the natural approach of imposing separate penalties on the local linear estimator and its derivative do not work as intended for semiparametric labeling, owing to the lack of connection between the coefficient and the derivative estimators in the popular local linear method. We then propose a stratified fix that borrows information from the coefficient estimator and combines it with the derivative into the same stratum that simultaneously achieves successful labeling and estimation. We establish the theoretical properties of the proposed method, including its estimation and labeling consistency, for a general class of nonstationary processes. Numerical examples, including a Monte Carlo simulation study and a real-data application, are presented to illustrate the proposed method.

Key words and phrases: Kernel smoothing, local linear estimation, nonstationary time series, time-varying coefficient model, variable selection.

1. Introduction

Linear regression models have been recognized as powerful and popular statistical tools for studying the relationship between a response variable and a set of explanatory variables. However, for applications to time series data, numerous empirical examples have suggested that the regression coefficient does not necessarily stay as a constant, and can change over time with other aspects of the data, making the observed time series nonstationary. For example, Fan and Zhang (1999) studied the relationship between the number of daily hospital admissions and the level of multiple pollutants in Hong Kong, finding a time-varying

Corresponding author: Ting Zhang, Department of Statistics, University of Georgia, Athens, GA 30602, USA. E-mail: tingzhang@uga.edu.