Statistica Sinica **36** (2026), 1-22
doi:https://doi.org/10.5705/ss.202023.0212

SIMULTANEOUS CHANGE POINT DETECTION AND IDENTIFICATION FOR HIGH-DIMENSIONAL LINEAR MODELS

Bin Liu[1], Xinsheng Zhang[1] and Yufeng Liu*[2]

[1]*Fudan University and* [2]*University of North Carolina at Chapel Hill*

Abstract: In this article, we consider change point inference for high-dimensional linear models. For change point detection, given any subgroup of variables, we propose a new method for testing the homogeneity of corresponding regression coefficients across the observations. Under some regularity conditions, the proposed new testing procedure controls the type I error asymptotically and is powerful against sparse alternatives and enjoys certain optimality. For change point identification, an "argmax" based change point estimator is proposed which is shown to be consistent for the true change point location. Moreover, combining with the binary segmentation technique, we further extend our new method for detecting and identifying multiple change points. Extensive numerical studies justify the validity of our new method and demonstrate its competitive performance.
Key words and phrases: Change point inference, high dimensions, linear regression, multiplier bootstrap, subgroups.

1. Introduction

Driven by the great improvement of data collection and storage capacity, high-dimensional linear regression models have attracted a lot of attentions because of its simplicity for interpreting the effect of different variables in predicting the response. Specifically, we are interested in the following model:

$$Y = \boldsymbol{X}^\top \boldsymbol{\beta} + \epsilon,$$

where $Y \in \mathbb{R}$ is the response variable, $\boldsymbol{X} = (X_1, \ldots, X_p) \in \mathbb{R}^p$ is the covariate vector, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is a p-dimensional unknown vector of coefficients, and $\epsilon \in \mathbb{R}$ is the error term.

For high-dimensional linear regression, the L_1-penalized technique lasso (Tibshirani, 1996) is a popular method for estimating $\boldsymbol{\beta}$. In the past decades, lots of research attentions both in machine learning and statistics have been focused on studying theoretical properties of lasso and other penalized methods. Most of the existing literature on high-dimensional linear regression focuses on the case with a homogeneous linear model, where the regression coefficients are assumed invariant across the observations. With many modern complex datasets for analysis in

*Corresponding author. E-mail: yfliu@email.unc.edu