Statistica Sinica **36** (2026), 1-23
doi:https://doi.org/10.5705/ss.202023.0288

KERNEL MODE-BASED REGRESSION UNDER RANDOM TRUNCATION

Tao Wang* and Weixin Yao

University of Victoria and University of California Riverside

Abstract: We propose to estimate a parametric regression with truncated data built on the mode value, where the dependent variable is subject to left truncation by another random variable. We construct a kernel mode-based objective function with a constant bandwidth for estimation and suggest a modified mode expectation-maximization algorithm to numerically estimate the model. The asymptotic normal distribution of the proposed estimator is derived under mild conditions. To efficiently construct confidence intervals for the resulting estimator, we develop a mode-based empirical likelihood method, where the asymptotic distribution of the empirical log-likelihood ratio is shown to follow a chi-square distribution. Furthermore, by combining the kernel mode-based objective function with the SCAD penalty, a variable selection procedure for the parameters is introduced and its oracle property is established. Monte Carlo simulations and real data analysis related to housing market are presented to show the finite sample performance of the developed estimation and variable selection procedures.

Key words and phrases: Empirical likelihood, mode-based regression, random truncation, robust estimation, variable selection.

1. Introduction

The concept of the mode is attractive as the value of highest probability density. The mode can be defined without moment conditions and could provide another understanding of the data, i.e., capture the "most likely" values. Owing to these appealing features, regression models based on the mode value, denoted as $\text{Mode}(Y \mid \mathbf{X})$ for random variables $(Y, \mathbf{X})$ (modal regression), have received significant attention recently (Yao and Li, 2014; Ullah, Wang and Yao, 2022, 2023), which can provide a valuable alternative to existing regressions. In addition, mode-based regression can be utilized as an alternative to robust regression to achieve robustness and efficiency in the presence of outliers or heavy-tailed distributions (Wang and Li, 2021; Wang, 2024). Due to space constraints, we provide more explanations on the distinction between modal regression and mode-based regression, as well as their respective advantages, in the online Supplementary Material. However, all existing methods related to mode-based regression assume that the data are fully observed, which may be unrealistic in

*Corresponding author. E-mail: taow@uvic.ca