

## Estimation and model selection in a class of semiparametric models for cluster data

Yan Sun · Jialiang Li · Wenyang Zhang

Received: 27 August 2010 / Revised: 6 June 2011 / Published online: 19 November 2011  
© The Institute of Statistical Mathematics, Tokyo 2011

**Abstract** Stimulated by a study in Bangladesh about the first birth interval, we propose a semivarying-coefficient model for cluster data analysis. We consider the estimation procedure for the proposed model and establish the asymptotic results of the proposed estimators. Furthermore, we employ the cross-validation (CV) to identify the constant coefficients. The associated asymptotic properties are rigorously examined. Simulation studies are conducted to investigate the performance of the proposed estimation and the CV-based model selection procedure for finite sample size. Finally, our methods are used to analyse the aforementioned data set to explore how several factors affect the first birth interval in Bangladesh.

**Keywords** Cluster data · Cross-validation · Local linear modelling · Semiparametric inference · Varying-coefficient models