Accepted Manuscript

Generalized nonparametric smoothing with mixed discrete and continuous data

Degui Li, Léopold Simar, Valentin Zelenyuk

PII: S0167-9473(14)00174-1
DOI: http://dx.doi.org/10.1016/j.csda.2014.06.003
Reference: COMSTA 5885

To appear in: *Computational Statistics and Data Analysis*

Received date: 2 January 2014
Revised date: 7 May 2014
Accepted date: 2 June 2014

Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics and Data Analysis 00 (2014) 1–26
Computational Statistics and Data Analysis

Generalized nonparametric smoothing with mixed discrete and continuous data

Degui Li[a], Léopold Simar[b], Valentin Zelenyuk[c1]

[a] Department of Mathematics, University of York, The United Kingdom
[b] Institut de Statistique, Biostatistique et Sciences Actuarielles, Université Catholique de Louvain, Belgium
[c] School of Economics and Centre for Efficiency and Productivity Analysis, University of Queensland, Australia

Abstract

The nonparametric smoothing technique with mixed discrete and continuous regressors is considered. It is generally admitted that it is better to smooth the discrete variables, which is similar to the smoothing technique for continuous regressors but using discrete kernels. However, such an approach might lead to a potential problem which is linked to the bandwidth selection for the continuous regressors due to the presence of the discrete regressors. Through the numerical study, it is found that in many cases, the performance of the resulting nonparametric regression estimates may deteriorate if the discrete variables are smoothed in the way previously addressed, and that a fully separate estimation without any smoothing of the discrete variables may provide significantly better results both for bias and variance. As a solution, it is suggested a simple generalization of the nonparametric smoothing technique with both discrete and continuous data to address this problem and to provide estimates with more robust performance. The asymptotic theory for the new nonparametric smoothing method is developed and the finite sample behavior of the proposed generalized approach is studied through extensive Monte-Carlo experiments as well an empirical illustration.

Keywords: Discrete regressors, Nonparametric regression, Kernel smoothing, Cross-validation, Local linear smoothing

1. Introduction

Rapid advance of computing power and wide availability of large data sets encourage many researchers to substantially increase their attention to various nonparametric methods for estimating nonlinear regression relationship. One of the most popular nonparametric methods appears to be the local polynomial least squares method, considered by Stone (1977), Cleveland (1979), Cleveland and Delvin (1988), Fan (1992, 1993), Fan and Gijbels (1992), Ruppert and Wand (1994) and popularized by Fan and Gijbels (1996). This method has received even greater appeal when it is substantially empowered by the seminal work of Racine and Li (2004), which suggests a neat way to deal with discrete regressors in the context of nonparametric regression. Racine and Li's work has inspired many interesting applications in a wide range of areas, for example, by Stengos and Zacharias (2006), Maasoumi et al. (2007), Parmeter et al. (2007), Eren and Henderson (2008), Walls (2009), Hartarska et al. (2010), Henderson (2010), to mention just a few. In these works where Racine and Li (2004)'s approach is used, researchers are able to obtain new insights with much more confidence, as their approach is free from imposing any parametric form on the regression relationship, while using both continuous and discrete regressors without splitting the sample into sub-samples for each value of

[1]Corresponding author. E-mail address: v.zelenyuk@uq.edu.au.