Available online at www.sciencedirect.com
ScienceDirect
Journal of Multivariate Analysis 99 (2008) 2016–2038
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

Asymptotic results in segmented multiple regression

Jeankyung Kim[a], Hyune-Ju Kim[b,*]

[a] *Department of Statistics, Inha University, 253 Yonghyundong, Namgu, Incheon, 402-751, Republic of Korea*
[b] *Department of Mathematics, 215 Carnegie Building, Syracuse University, Syracuse, NY 13244-1150, USA*

Received 24 April 2007
Available online 17 February 2008

Abstract

This paper studies the asymptotic behavior of the least squares estimators in segmented multiple regression. For a model with more than one partitioning variable, each of which has one or more change-points, we study the asymptotic properties of the estimated change-points and regression coefficients. Using techniques in empirical process theory, we prove the consistency of the least squares estimators and also establish the asymptotic normality of the estimated regression coefficients. For the estimated change-points, we obtain their consistency at the rates of $1/\sqrt{n}$ or $1/n$, with or without continuity constraints, respectively. The change-points estimated under the continuity constraints are also shown to asymptotically have a multivariate normal distribution. For the case where the regression mean functions are not assumed to be continuous at the change-points, the asymptotic distribution of the estimated change-points involves a step function process, whose distribution does not follow a well-known distribution.

AMS 2000 subject classifications: primary 62J02; secondary 62E20

Keywords: Empirical process; Change-point; Consistency; Asymptotic normality

1. Introduction

A segmented line regression model has been used in many applications to describe changes in linear trends. A great amount of work has been done on segmented line regression, dealing with topics like computational issues in fitting a segmented line regression model, inference on parameters including the change-points, and the selection of the number of change-points. For a model with a given number of change-points, distributional properties of the least squares

* Corresponding author.
E-mail address: hjkim@syr.edu (H.-J. Kim).

0047-259X/$ - see front matter
doi:10.1016/j.jmva.2008.02.028

estimators have been studied by many authors. Feder [4] considered a segmented regression model with one independent variable under the continuity constraints and studied the asymptotic distributions of the estimated regression coefficients and change-points. Under some technical assumptions on the independent variable, Feder [4] proved the consistency and asymptotic normality of the least squares estimators, first for the pseudo-sample, which is obtained by removing some data points around the estimated change-points and then obtained the desired results for the full sample using the asymptotic equivalence between the pseudo-sample estimates and the full sample estimates. Hinkley [6,7] and Hŭsková [9] considered some special cases of the model studied in Feder [4], and provided more details on distributional properties of the estimators. Bai [1] and Bai and Perron [2,3] considered a multiple regression model with structural changes, the model without the continuity constraints at the change-points, and studied the asymptotic properties of the estimators. For a model where the change-points are index points in the discrete scale, Bai [1] proved, under mild assumptions on the independent variables, that the estimated break fraction is consistent at the rate of $1/n$ and showed that the limiting distribution of the estimated change-point involves an argument that maximizes a shifted Wiener processes. Liu et al. [13] proposed an information-based criterion to select the number of change-points, and studied the asymptotic behavior of the estimators of multivariate regression models, with or without the continuity constraints. Under some assumptions on the independent variables, similar to those of Bai [1], Liu et al. [13] proved the consistency of the estimated change-points at the rates of $1/\sqrt{n}$ or $1/n$ for a continuous or discontinuous model, respectively, and obtained the asymptotic normality for the distributions of the estimated regression coefficients. Most of these work, however, focus on the case with only one partitioning variable through which changes occur, and it is our interest in this paper to consider a segmented regression model with more than one partitioning variable, with or without the continuity constraints at the change-points.

Fitting the segmented regression model with two partitioning variables can be done by a grid search, where a two-dimensional grid is searched over for a minimum residual sum of squares or by using nonlinear regression such as SAS PROC NLIN. Our focus in this paper is not on computational issues of fitting algorithms, but on the asymptotic behavior of the least squares estimators, which would help one to conduct asymptotic inference on the parameters. We first establish the consistency of the least squares estimators by using a maximal inequality for partial sums. Then, we study the asymptotic properties of the least squares estimators for a simplest model with two partitioning variables, each of which has one change-point, and with one additional independent variable. We note that its extension to models with more than two partitioning variables, each with multiple change-points, and/or more than one additional variable is straightforward. When we have the continuity constraints at the change-points, the asymptotic joint distribution of the estimated regression coefficients is multivariate normal and the estimated change-points are consistent at the $1/\sqrt{n}$-rate. The distribution of the estimated change-points also can be approximated by a bivariate normal distribution. When the segments are not assumed to be continuous at the change-points, we prove the asymptotic normality of the estimated regression coefficients and the consistency of the estimated change-points at the $1/n$-rate and discuss the distribution of the estimated change-points.

This paper is organized as follows. In Section 2, we present the model and summarize some preliminary results such as the consistency of the least squares estimators. Section 3 considers the segmented regression model with the continuity constraints at the change-points. The model without the continuity constraints is studied in Section 4. Section 5 summarizes simulation results and also includes an example. Concluding remarks are given in Section 6.