Statistical Methodology 19 (2014) 4–24

Contents lists available at ScienceDirect

Statistical Methodology

journal homepage: www.elsevier.com/locate/stamet

Statistical tests in the partially linear additive regression models

Salim Bouzebda [a], Khalid Chokri [b,*]

[a] *Laboratoire de Mathématiques Appliquées de Compiègne-L.M.A.C., Université de Technologie de Compiègne, B.P. 529, 60205 Compiègne Cedex, France*

[b] *L.S.T.A., Université Pierre et Marie Curie, 4 place Jussieu, 75252 Paris Cedex 05, France*

A R T I C L E I N F O

Article history:
Received 1 December 2012
Received in revised form
3 February 2014
Accepted 4 February 2014

Keywords:
Additive model
Asymptotic normality
Curse of dimensionality
Kernel-type estimators
Law of iterated logarithm
Marginal integration
Partially linear models
Hypothesis testing

A B S T R A C T

In the present paper, we are mainly concerned with statistical tests in the partially linear additive model defined by

$$Y_i = \mathbf{Z}_i^\top \boldsymbol{\beta} + \sum_{\ell=1}^{d} m_\ell(X_{i,\ell}) + \varepsilon_i, \quad 1 \le i \le n,$$

where $\mathbf{Z}_i = (Z_{i,1}, \ldots, Z_{ip})^\top$ and $\mathbf{X}_i = (X_{i,1}, \ldots, X_{id})^\top$ are vectors of explanatory variables, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is a vector of unknown parameters, $m_1, \ldots, m_d$ are unknown univariate real functions, and $\varepsilon_1, \ldots, \varepsilon_n$ are independent random errors with mean zero and finite variances σ_ε^2. More precisely, we first consider the problem of testing the null hypothesis $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. The second aim of this paper is to propose a test for the null hypothesis $\mathcal{H}_0^\sigma : \sigma_\varepsilon^2 = \sigma_0^2$, in the partially linear additive regression models. Under the null hypotheses, the limiting distributions of the proposed test statistics are shown to be standard chi-squared distributions. Finally, simulation results are provided to illustrate the finite sample performance of the proposed statistical tests.

* Corresponding author. Tel.: +33 609504606.
E-mail addresses: salim.bouzebda@utc.fr (S. Bouzebda), khalid_chokri@yahoo.fr (K. Chokri).

http://dx.doi.org/10.1016/j.stamet.2014.02.001
1572-3127/

1. Introduction

Regression analysis is proved to be a flexible tool and that provided a powerful statistical modeling framework in a variety of applied and theoretical contexts, where one intends to model the predictive relationship between related responses and predictors. It is worth noticing that the parametric regression models provide useful tools for analyzing practical data when the models are correctly specified, but may suffer from large modeling biases when the structures of the models are misspecified, which is the case in many practical problems. As an alternative, nonparametric smoothing methods ease the concerns on modeling biases. However, it is well known that unrestricted multivariate nonparametric regression models are subject to the *curse of dimensionality*, in multivariate settings, and fail to take advantage of the flexibility structure in modeling phenomena with *moderate* set of data, see [58,59,26,29] among others. To overcome that difficulty, Engle et al. [24] modeled the covariate effects via a partially linear structure, which is a special semiparametric model. Since the partially linear models contain both parametric and nonparametric components, they offer a compromise between the flexibility of a full nonparametric regression and reasonable asymptotic behavior. For several examples of practical problems that can be solved with partial linear models, the interested reader may refer to Härdle et al. [30] for more details. To be more precise, the partially linear regression models are defined as follows

$$Y = \mathbf{Z}^\top \boldsymbol{\beta} + m(\mathbf{X}) + \varepsilon, \tag{1.1}$$

where $\boldsymbol{\beta} \in \mathbb{R}^p$ is a vector of unknown parameters, m is the nonlinear part of the model and ε is the modeling error with

$$\mathbb{E}(\varepsilon|\mathbf{X}, \mathbf{Z}) = 0 \quad \text{and} \quad \mathrm{Var}(\varepsilon|\mathbf{X}, \mathbf{Z}) = \sigma^2(\mathbf{X}, \mathbf{Z}) := \sigma_\varepsilon^2.$$

Here and in the sequel, $\mathbf{V}^\top$ stands for the transpose of the vector $\mathbf{V}$. The partially linear regression model has a broad applicability in the fields of biology, economics, education and social sciences. This model and various associated estimators, test statistics, and generalizations have generated a substantial body of literature, which includes, among many others, the works of Rice [51], Chen [11], Robinson [52], Chen and Shiau [13], Eubank and Speckman [25], Donald and Newey [21], Shi and Li [56,57], Bhattacharya and Zhao [2], Hamilton and Truong [28], Liang et al. [38], Shen et al. [54], Yu et al. [63], Bouzebda et al. [5] and the reference therein. To reduce the dimension impact of the nonparametric part in the partially linear regression model (1.1), we consider the partially linear additive model that imposes an additive structure to the nonparametric function m

$$Y = \mathbf{Z}^\top \boldsymbol{\beta} + \sum_{j=1}^{d} m_j(\mathrm{x}_j) + \varepsilon, \tag{1.2}$$

where X_j is the jth component of the vector $\mathbf{X}$ and m_j is a real univariate function. In the partially linear additive regression models, the functions m_j can be estimated with the one-dimensional rate. Hence the curse of dimensionality can be treated in a satisfactory manner and another advantage is that the lower-dimensional curves are easier to visualize and to interpret than a higher-dimensional function, see [46] for further discussions. We may also cite the paper of Yu et al. [63], that analyzes efficiency gains in semiparametric models from imposing additional structure on the nonparametric component.

In practice, investigators often want to know the impact of the covariates $\mathbf{Z}$ on the response Y, under the model (1.2), which requires testing the null hypothesis

$$\mathcal{H}_0^{\boldsymbol{\beta}} : \boldsymbol{\beta} = \boldsymbol{\beta}_0,$$

versus the alternative

$$\mathcal{H}_1^{\boldsymbol{\beta}} : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0.$$

In this paper we construct a statistical test for this end. The second aim of the present paper is to test the null hypothesis

$$\mathcal{H}_0^{\sigma} : \sigma_\varepsilon^2 = \sigma_0^2,$$