Journal of Computational and Applied Mathematics 235 (2011) 1418–1428

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

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

Efficiency of a Liu-type estimator in semiparametric regression models

Esra Akdeniz Duran [a], Fikri Akdeniz [b,*], Hongchang Hu [c]

[a] Department of Statistics, Gazi University, Ankara, Turkey
[b] Department of Statistics, Çukurova University, Adana, Turkey
[c] School of Mathematics and Statistics, Hubei Normal University, Huangshi 435002, PR China

ARTICLE INFO

Article history:
Received 9 April 2009
Received in revised form 5 March 2010

MSC:
primary 62G05
secondary 62J05
62J07

Keywords:
Almost unbiased Liu-type estimator
Liu-type estimator
Mean square error
Semiparametric regression model
Two-step estimation

ABSTRACT

In this paper we consider the semiparametric regression model, $y = X\beta + f + \varepsilon$. Recently, Hu [11] proposed ridge regression estimator in a semiparametric regression model. We introduce a Liu-type (combined ridge-Stein) estimator (LTE) in a semiparametric regression model. Firstly, Liu-type estimators of both β and f are attained without a restrained design matrix. Secondly, the LTE estimator of β is compared with the two-step estimator in terms of the mean square error. We describe the almost unbiased Liu-type estimator in semiparametric regression models. The almost unbiased Liu-type estimator is compared with the Liu-type estimator in terms of the mean squared error matrix. A numerical example is provided to show the performance of the estimators.

1. Introduction

Consider the following semiparametric regression model

$$y_i = x_i^T\beta + f(t_i) + \varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{1}$$

where y_i's are observations, $x_i^T = (x_{i1}, x_{i2}, \ldots, x_{ip})$ and $x_1, x_2, \ldots, x_n$ are known p-dimensional vectors with $p \le n$, t_i's are values of an extra univariate variable such as the time at which the observation is made, $\beta = (\beta_1, \beta_2, \ldots, \beta_p)^T$ is an unknown p-dimensional parameter vector, $f(.)$ is an unknown smooth function, and ε's are random errors assumed to be i.i.d. $N(0, \sigma^2)$ distributed.

This model, also called a partial linear model or a partial spline model introduced in [1]. We shall call $f(t)$ the smooth part of the model and assume that it represents a smooth unparametrized functional relationship. The goal is to estimate the unknown parameter vector β and nonparametric function $f(t)$ from the data $\{y_i, x_i, t_i\}$. In matrix–vector notation, model (1) is written as

$$y = X\beta + f + \varepsilon, \tag{2}$$

where $y = (y_1, \ldots, y_n)^T$, $X^T = [x_1, \ldots, x_n]$, $f = (f(t_1), \ldots, f(t_n))^T$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$.

As in most literature, we assume that X has full column rank. This model has received a considerable amount of research in the past two decades. Engle et al. [1] proposed these models and used them to analyze the relation between electricity

* Corresponding author. Tel.: +90 3223386548; fax: +90 3223386070.
E-mail addresses: esraakdeniz@gmail.com (E.A. Duran), akdeniz@cu.edu.tr (F. Akdeniz), retutome@163.com (H. Hu).

0377-0427/$ – see front matter
doi:10.1016/j.cam.2010.08.028

usage and average daily temperature. Since then the models have been widely studied in the literature. One reason is that it is much more flexible than the standard linear model since it combines both parametric and nonparametric components. Another reason is that it allows easier interpretation of the effect of each variable compared to a completely nonparametric regression. Because of its relation to the classical linear regression model, this model is called 'partially linear model' in the literature. Model (2) is devised of the form

$$E(y|X, t) = X\beta + f, \tag{3}$$

where the explanatory variables are represented separately in two parts: the nonparametric part ($f(t)$) and the parametric linear part ($X\beta$). In this semiparametric regression model, both the functional form of the nonparametric part of the model and the parameters will be estimated.

All the existing approaches for the partially linear model are based on different nonparametric regression procedures. There have been several approaches for estimating β and f. Among the most important approaches are the spline methods in [1–5]; kernel methods in [6,7]; piecewise polynomials method in [8,9] constructed a feasible least squares estimator of β by estimating the nonparametric component by a Nadaraya–Watson kernel estimator. Speckman [7] introduced the idea of a profile least squares method and studied local constant smoother. Liang [10] proposed a penalized spline method with a linear mixed effects framework in partially linear models. Hu [11] used two-step estimation method for ridge estimators of both the parametric and the nonparametric component of the semiparametric regression model. The essential thought of two-step estimation is the following: the first step, $f(t, \beta)$ is defined with supposition where β is supposed to be known; the second step, the estimator of parametric β is attained by least squares method; accordingly, $\hat{f} = f(t, \hat{\beta})$ is gained.

In regression analysis, researchers often encounter the problem of multicollinearity. The remedies for the problem of multicollinearity depend on the objective of the regression analysis. The multicollinearity is a problem when the primary interest is in the estimation of the parameters in a regression model. In the case of multicollinearity we know that when the correlation matrix has one or more small eigenvalues, the estimates of the regression coefficients can be large in absolute value. The least squares estimator performs poorly in the presence of multicollinearity. Some biased estimators have been suggested as a means to improve the accuracy of the parameter estimate in the model when multicollinearity exists. For the purposes of this paper we will employ the biased estimator proposed in [12] to combat the multicollinearity. Liu [12] combined the Stein [13] estimator with the ordinary ridge regression estimator to obtain what we call the Liu estimator or the Liu-type estimator (see [14,15]).

Condition number is a measure of the presence of multicollinearity. If $X'X$ is ill conditioned with a large condition number, ridge regression estimator Hoerl and Kennard [16] or a Liu-type estimator (see [12,17]) can be used to estimate β. In this paper we will examine a biased estimation technique to be followed when the matrix $X'X$ appears to be ill conditioned in the semiparametric regression model. We assume that the condition number of the parametric component is large indicating that a biased estimation procedure is desirable.

Traditionally, real-valued squared error loss and the matrix-valued squared error loss are considered for investigation of the performance of an estimator $\phi(y)$ for parameter vector $\theta \in \Theta$ (parameter space). Nonetheless, it is sometimes reasonable to take alternative loss functions into account. For example, asymmetric loss function (the so-called LINEX loss function)

$$L(\theta, \phi(y)) = b[e^{a\Phi} - a\Phi - 1], \quad a \neq 0,\ b > 0,\ \Phi = \phi(y) - \theta$$

(see Zellner [18]). Zellner [19] proposed a balanced loss function

$$L(\theta, \phi(y)) = \alpha \left\| y - \phi(y) \right\|^2 + (1 - \alpha) \left\| \theta - \phi(y) \right\|^2, \quad 0 < \alpha < 1.$$

The balanced loss function consists of two components. First, a measure of goodness of fit $\|y - \phi(y)\|^2$ with a relative weight α, and, second, a measure of goodness of estimation $\|\theta - \phi(y)\|^2$ with a relative weight $1 - \alpha$.

In this paper we consider two competing estimators b_d (Liu-type estimator) and $\hat{\beta}_{TS}$ (two-step estimator). We measure the "closeness" of b_d and $\hat{\beta}_{TS}$ to β in terms of squared Euclidean distance by the trace of the matrix mean square error (*MSEM*). We also consider the superiority of Almost Unbiased Liu-type estimator (AULTE) and Liu-type Estimator (LTE) in semiparametric regression model.

The paper is organized as follows. In Section 2, the estimators in the semiparametric model are introduced. The Liu-type estimator is compared with the two-step estimator in terms of the mean square error in Section 3. In Section 4, we describe the almost unbiased combined Liu-type estimator in semiparametric regression models. In Section 5, the almost unbiased Liu-type estimator is compared with the Liu-type estimator in terms of the mean squared error matrix. Section 6 gives a numerical example to show the performance of the estimators.

2. Liu-type estimator

Consider the multiple linear regression model

$$y = X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I), \tag{4}$$

where y is an $n \times 1$ vector of observations on the dependent variable, X is an $n \times p$ matrix of rank p, β is a $p \times 1$ vector of regression coefficients, and ε is an $n \times 1$ vector of error terms with $E(\varepsilon) = 0$, $E(\varepsilon\varepsilon') = \sigma^2 I_n$, where σ is constant but unknown.