Statistics and Probability Letters 83 (2013) 100–114

Contents lists available at SciVerse ScienceDirect

Statistics and Probability Letters

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

Semiparametric analysis of additive isotonic errors-in-variables regression models

Zhimeng Sun [a,b,*], Zhongzhan Zhang [b]

[a] School of Statistics, Central University of Finance and Economics, Beijing 100081, China
[b] College of Applied Sciences, Beijing University of Technology, Beijing 100124, China

ARTICLE INFO

Article history:
Received 26 March 2012
Received in revised form 30 August 2012
Accepted 30 August 2012
Available online 7 September 2012

MSC:
62G05
62G20

Keywords:
Measurement errors
Semiparametric
Isotonic regression
Asymptotic properties
Estimation

ABSTRACT

We consider the estimation of a semiparametric additive isotonic regression model with error-prone covariates. We show the limiting distributions of the proposed estimators of the parametric component as well as the functional component. A simulation study is carried out to investigate the performance of the proposed estimators.

1. Introduction

Consider the nonparametric isotonic regression model

$$Y = h(W) + \varepsilon \tag{1}$$

where Y is a response, W is a random covariate valued in R^K, $K \geq 1$, ε is the model error, and $h(\cdot)$ is a regression function which is monotonic in each coordinate of W. In epidemiology area, model (1) is often used to model the relationship between risk and exposure, which is unknown but believed non-decreasing with increasing exposure. Traditional nonparametric models usually assume that $h(\cdot)$ is a smooth function and use the polynomial spline or the penalized least squares method to estimate $h(\cdot)$. Compared with these nonparametric approaches, an attractive property of the isotonic estimation approach is that it is entirely data-driven and does not need to specify the smoothing parameter or the penalty parameter. This is because the built-in monotonicity constraints allow for choosing the smoothing parameter automatically. The nonparametric isotonic regression model has a long history and has been studied by Brunk (1958), Wright (1981), Hall and Huang (2001), Mammen et al. (2001), Dette et al. (2006), among others.

It is well known that there may exist curse of dimensionality in the estimation of high dimensional $h(\cdot)$ in model (1). One effective way to overcome difficulties caused by curse of dimensionality is that we employ an additive structure on the high

* Corresponding author at: School of Statistics, Central University of Finance and Economics, Beijing 100081, China.
E-mail address: zmsun82@163.com (Z. Sun).

0167-7152/$ – see front matter
doi:10.1016/j.spl.2012.08.028

dimensional function $h(\cdot)$. Model (1) thus becomes

$$Y = \sum_{k=1}^{K} h_k(W_k) + \varepsilon \tag{2}$$

which is known as the nonparametric additive isotonic regression model and has been studied by Mammen and Yu (2007). A natural extension of model (2) incorporating a linear part is the semiparametric additive isotonic regression model

$$Y = X^\top \beta + \sum_{k=1}^{K} h_k(W_k) + \varepsilon \tag{3}$$

where $X = (X_1, \dots, X_p)^\top$ is a p-dimensional covariate, β is a p-dimensional parameter of interest, $\top$ represents transpose. Under model (3) with $K = 1$, Huang (2002) proposed least-square estimators for β and $h_1(\cdot)$ and showed the limiting distribution of the estimators. Cheng (2009) extended Huang (2002)'s results to the situation that $K > 1$. For other studies on semiparametric isotonic regression model see Sun et al. (2011, 2012); Sun and Zhang (2012), among others. In applications, X can usually not be measured exactly. It is often measured with errors. Instead of observing X, we observe $Z = X + U$, where $U = (U_1, \dots, U_p)^\top$ is the p-dimensional vector of measurement errors. For studies on observations with additive errors see Cui and Li (1998), Liang et al. (1999), Cui and Kong (2006), among others.

Above all, we are interested in the semiparametric additive isotonic errors-in-variables regression model which takes the form

$$Y = X^\top \beta + \sum_{k=1}^{K} h_k(W_k) + \varepsilon, \qquad Z = X + U. \tag{4}$$

Without loss of generality, we assume that ε is independent of $(X, W_1, \dots, W_K)$ with mean 0 and variance σ^2, $W_k(k = 1, \dots, K) \in \mathscr{W}$ are pairwise independent, $\mathscr{W}$ is a closed interval $[\underline{\mathscr{W}}, \overline{\mathscr{W}}]$ on the real line R, where $\underline{\mathscr{W}}$ and $\overline{\mathscr{W}}$ are the left endpoint and the right endpoint respectively, U is independent of $(Y, X, W_1, \dots, W_K, \varepsilon)$ with mean 0 and covariance Σ_{UU}, $h_k(\cdot)(k = 1, \dots, K)$ are non-decreasing on $\mathscr{W}$. For identifiability of the regression parameter, we assume that $E(h_k(W_k)) = 0, k = 1, \dots, K$. We also center Y and X around zero; thus $E(X)$ is set to be zero. For identifiability of the model, we assume that Σ_{UU}, which may be determined by machines in experiments, is already known. Let $\{(Y_i, W_{1i}, \dots, W_{Ki}, Z_i), i = 1, \dots, n\}$ be independent identically distributed observations of $(Y, W_1, \dots, W_K, Z)$ from model (4). That is $Y_i = X_i^\top \beta + \sum_{k=1}^{K} h_k(W_{ki}) + \varepsilon_i, Z_i = X_i + U_i, i = 1, \dots, n$. A simple estimation approach of model (4) is that we ignore the measurement errors and replace X by Z. Unfortunately, the resulting estimator of β is inconsistent, and hence results in inconsistent estimate curves for $h_k(W_k)(k = 1, \dots, K)$, as can be seen from both the form of the estimator of β and the simulation results in the following parts of this paper. Different from the estimation procedure in Huang (2002) and Cheng (2009) in which the estimators of the parametric and nonparametric components were defined synchronously and a reasonable initial estimator of the parameter β is required in the computation algorithm, in this paper, we derive root-n consistent estimators of β and σ^2 first. Then, based on the estimator of β, we define the least-square estimators of $h_k(W_k)(k = 1, \dots, K)$ under monotone constraints. We prove that both the estimators of β and σ^2 are asymptotically normally distributed. We show the asymptotic distributions of estimators of the nonparametric components, as well as the estimator's oracle property, which means that $h_k(\cdot)$ can be estimated as well as it could be by an isotonic estimator as if the other nonparametric and parametric components were known. The outline of the paper is as follows. In Section 2, we show the estimation procedure and give the asymptotic distributions of the estimators. In Section 3, we conduct a simulation to detect the behavior of the estimators. The proofs of the theorems are given in the Appendix.

2. Estimation procedure and main results

2.1. Estimation procedure for the parametric components

Our aim is to construct estimators for the parametric components β, σ^2 and the nonparametric components $h_k(\cdot)(k = 1, \dots, K)$ of model (4). From model (4), it is easy to get that

$$Y_i = Z_i^\top \beta + \sum_{k=1}^{K} h_k(W_{ki}) + \varepsilon_i - U_i^\top \beta, \quad i = 1, \dots, n. \tag{5}$$

Taking conditional expectation on $W_{1i}, \dots, W_{Ki}$ respectively on both sides of (5), we have

$$E[Y_i|W_{ki}] = E[Z_i|W_{ki}]^\top \beta + h_k(W_{ki}), \quad k = 1, \dots, K. \tag{6}$$

Subtracting the sum of the each side of (6) from the same side of (5) respectively, we get

$$Y_i - \sum_{k=1}^{K} E[Y_i|W_{ki}] = \left[Z_i - \sum_{k=1}^{K} E[Z_i|W_{ki}] \right]^\top \beta + \varepsilon_i - U_i^\top \beta, \quad i = 1, \dots, n. \tag{7}$$