Computational Statistics and Data Analysis 128 (2018) 58–72

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Quasi-likelihood estimation of the single index conditional variance model

Hongfan Zhang

School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, Sichuan, 611731, PR China

A R T I C L E I N F O

Article history:
Received 17 July 2017
Received in revised form 18 March 2018
Accepted 18 June 2018
Available online 28 June 2018

Keywords:
Conditional variance
Single index model
Heteroscedasticity
Quasi-likelihood

A B S T R A C T

This paper investigates estimation methods for the conditional variance function with a single index structure. We introduce two estimators of the single index parameter vector through maximizing local linear quasi-likelihood functions. The resulting parameter index estimators can achieve root-n consistency and the variance function estimator can maintain positivity. We show that the proposed methods can estimate the conditional variance with the same asymptotic efficiency as if the conditional mean function is given. Asymptotic distributions of the proposed estimators are also derived. Simulation studies and a real data application demonstrate our estimation approaches.

1. Introduction

Exploring estimation methods for conditional variance functions has been an interesting topic in many disciplines of research. Consider the nonparametric heteroscedastic model

$$Y = M(\mathbf{X}) + V^{1/2}(\mathbf{X})\varepsilon, \tag{1}$$

where $Y \in \mathbb{R}^1$ represents the univariate response and X is the explanatory variable which can be univariate or multivariate, with the latter being the focus of this article. The term ε is generally assumed to satisfy $E(\varepsilon|\mathbf{X}) = 0$ and $E(\varepsilon^2|\mathbf{X}) = 1$. Note that we assume the variance of ε conditioning on $\mathbf{X}$ is unit. If otherwise, the variance can be absorbed into the nonparametric function $V(\mathbf{X})$. Under these assumptions we have $M(\mathbf{X}) = E(Y|\mathbf{X})$ and $V(\mathbf{X}) = Var(Y|\mathbf{X})$, both of which are unknown and need to be estimated.

For $\mathbf{X} \in \mathbb{R}^1$, methods of estimating $V(\mathbf{X})$ are commonly based on local linear smoothing techniques; see, for example, Fan and Gijbels (1996), Ruppert et al. (1997) and Fan and Yao (1998). These methods first estimate the conditional mean and then the conditional variance by regressing the squared residuals. The resulting estimator is proved to be asymptotically equivalent to that with $M(\mathbf{x})$ known. However, the local linear conditional variance estimator cannot guarantee positive values, hence making this approach deficient. Xu and Phillips (2011) proposed a re-weighted local constant estimator which inherits the nonnegative property of the variance function and retains the asymptotic bias and variance of the local linear estimator. Some other estimators based on maximizing the local likelihood function were also proposed to tackle the nonnegative problem, see Avramidis (2002), Ziegelmann (2002) and Yu and Jones (2004). Compared with the conventional local linear estimator, the resulting maximum likelihood estimators have more promising properties and better numerical performance. Actually in this paper this idea is generalized to handle our single index variance model.

If model (1) is extended to the multivariate case that $\mathbf{X} \in \mathbb{R}^p$, the "curse of dimensionality" urges us to restrict the form of the conditional variance function to a lower dimension without any loss of information. In this paper we concentrate on

E-mail address: zhfbeyond@126.com.

https://doi.org/10.1016/j.csda.2018.06.008
0167-9473/

the following single index structure for both conditional variance and conditional mean:

$$Y = m(\beta_0^\top \mathbf{X}) + v^{1/2}(\alpha_0^\top \mathbf{X})\varepsilon, \tag{2}$$

where β_0 and α_0 are two p-dimensional vectors, and ε is also assumed to satisfy $E(\varepsilon|\mathbf{X}) = 0$ and $E(\varepsilon^2|\mathbf{X}) = 1$. Usually the error term ε is assumed to be Gaussian, which facilitates maximum likelihood function. However, we will show the Gaussian assumption is not essential, leading to our quasi-likelihood estimation. For model identification, we further assume that $|\beta_0| = |\alpha_0| = 1$ where $|\cdot|$ denotes the usual Euclidian norm, and their first elements are positive. This paper focuses on the estimation of α_0.

For the single index model, various estimation methods have been proposed, for example, Härdle and Stoker (1989), Ichimura (1993), Härdle et al. (1993), Horowitz and Härdle (1996), Yin and Cook (2005) and Xia (2006). The refined OPG method and refined MAVE method proposed by Xia (2006) outperform many other existing methods on account of their being free of strong restrictions on $\mathbf{X}$ design, root-n consistency of the parameter index estimators and good estimation accuracy. However, the methods of Xia (2006) are mainly intended for the conditional mean function $m(\beta_0^\top \mathbf{X})$ and the parameter index β_0 therein, rather than the single index conditional variance model $v(\alpha_0^\top \mathbf{X})$. To indicate the difference, we shall denote Xia's methods by R-OPGm and R-MAVEm respectively.

Note that if the conditional mean function is given, then model (2) reduces to the single index volatility model $r = \sigma(\alpha_0^\top \mathbf{X})\varepsilon$, where $r = Y - m(\beta_0^\top \mathbf{X})$ and $\sigma(\cdot) = v^{1/2}(\cdot)$; see Xia et al. (2002a) and Engle (1982). Our methods for estimating the single index volatility model are based on maximizing local quasi-likelihood functions. The formulated objective function, although derived from the error term being assumed Gaussian, can be deemed as an "average variance" with a "penalty" term, the latter removing trivial solutions. Like R-OPGm and R-MAVEm, we also use techniques such as single dimensional kernels and alternating iteration procedures for the single index volatility model. Similarly, we denote our two estimation methods by L-OPGv and L-MAVEv. When dealing with model (2) with the conditional mean function unknown, we comply with the residual based methods as for model (1), namely, first to use R-MAVEm of Xia (2006) to estimate β_0 and $m(\cdot)$, and second to use L-OPGv and L-MAVEv as stated above to estimate α_0 and $v(\cdot)$ by regressing the estimated residuals. We will show that, in this case, our proposed methods can estimate the conditional variance with the same asymptotic efficiency as if the conditional mean function is given.

It is worth mentioning that model (2) is a dimension reduction model. If we denote $B = (\beta_0, \alpha_0)$, then the space spanned by the columns of B, $\mathcal{S}(B)$ say, is the central dimension reduction subspace (CS); see, e.g., Cook (1998b), Cook and Li (2002), Li (1991) and Xia et al. (2002b). Further, it can be seen that $\beta_0 \in \mathcal{S}_{E(Y|\mathbf{X})}$ (CMS), the central mean subspace; see, e.g., Cook (1998a), Cook and Li (2002), Li (1992), and $\alpha_0 \in \mathcal{S}_{Var(Y|\mathbf{X})}$ (CVS), the central variance subspace; see, e.g., Yin and Cook (2002). A number of dimension reduction methods are capable of recovering β_0 and α_0; see Ma and Zhu (2013) for detailed discussion. The well known dMAVE, the density based MAVE method of Xia (2007), can handle model (2). In fact in the experiment section of Xia (2007), a simulation for model (2) has been conducted to demonstrate the performance of dMAVE. Zhu et al. (2013) address the estimate of model (2) through estimating equations. This approach is an extension of the so called semiparametric method of Ma and Zhu (2012), which also pertains to a general class of dimension reduction methods. Another emerging approach to single index models is based on maximizing a kernel version of the Hilbert–Schmidt Independence Criterion. For example, see Zhang and Yin (2015). This approach can also handle model (2). Owing to the popularity of the above three methods in respect of dimension reduction, we shall compare them on finite sample performance with our estimation methods developed in this paper.

The rest of the article is organized as follows. Section 2 describes the details of the quasi-likelihood estimation methods of the single index variance model, including the derivation and the algorithms. Section 3 studies the asymptotic normality. In Section 4, some numerical simulations are carried out to evaluate the finite sample performance of our methods. We also report some results from an empirical study of the Hitters' salary data in Section 5. Section 6 gives a short discussion. Regularity conditions and proofs of results are presented in the Appendix.

2. Quasi-likelihood estimation

Let $(Y_i, \mathbf{X}_i, i = 1, \ldots, n)$ with $Y_i \in \mathbb{R}^1$ and $\mathbf{X}_i \in \mathbb{R}^p$ be an i.i.d. random sample from the population $(Y, \mathbf{X})$. Under model (1), if the error terms $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ are assumed to be independent and each $\varepsilon_i|\mathbf{X}_i = \mathbf{x}_i \sim N(0, 1)$, we have

$$Y_i|\mathbf{X}_i = \mathbf{x}_i \sim N(M(\mathbf{x}_i), V(\mathbf{x}_i)). \tag{3}$$

Then the sample log-likelihood function of $Y|\mathbf{X} = \mathbf{x}$, with some constant terms ignored, can be written as

$$L = -\frac{1}{2}\sum_{i=1}^{n}\left\{\frac{r_i^2}{V(\mathbf{X}_i)} + \log V(\mathbf{X}_i)\right\}, \tag{4}$$

where $r_i = Y_i - M(\mathbf{X}_i)$. The true residual r_i here corresponds to the single index volatility model with the conditional mean function known. By model (2) we have $M(\mathbf{X}_i) = m(\theta_0^\top \mathbf{X}_i)$ and $V(\mathbf{X}_i) = v(\alpha_0^\top \mathbf{X}_i)$. The R-OPGm and the R-MAVEm methods are capable of estimating β_0 and $m(\cdot)$, and hence r. At this stage we retain r_i in (4) instead of its estimator $\hat{r}_i$ for the sake of easy elaboration. Substituting α_0 by α in (4) and writing L as $L(\alpha)$, $L(\alpha)$ is a loss function. If the function $v(\cdot)$ is given, maximizing $L(\alpha)$ with respect to α will result in the classic nonlinear least squares estimator of α_0. However, the conditional variance function is unknown here. Local linear smoothing technique should be employed, which can be expounded in two separate ways as follows.