COMSTA: 6609 Model 3G pp. 1–18 (col. fig: NIL)

ARTICLE IN PRESS

Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Time-varying quantile single-index model for multivariate responses

Weihua Zhao [a], Yan Zhou [b,*], Heng Lian [c,*]

[a] School of Science, Nantong University, Nantong, PR China
[b] College of Mathematics and Statistics, Institute of Statistical Sciences, Shenzhen University, PR China
[c] Department of Mathematics, The City University of Hong Kong, Kowloon Tong, Hong Kong

ARTICLE INFO

Article history:
Received 11 December 2017
Received in revised form 15 March 2018
Accepted 9 May 2018
Available online xxxx

Keywords:
B-splines
Multiple responses
Quantile regression
Single-index model

ABSTRACT

We consider simultaneous semiparametric estimation of conditional quantiles for multiple responses using a dynamic single-index structure. Motivated by a financial application, a market factor index is constructed that is shared among different portfolios which results in a more interpretable and efficient model, compared to separately building multiple conditional quantiles. On the other hand, the link functions are allowed to be different across portfolios. The asymptotic normality of the index parameter is established, as well as the convergence rate of the nonparametric functions. Monte Carlo studies demonstrated the advantages of the proposed estimator and an application to financial data is used to illustrate the method.

1. Introduction

Linear quantile regression, proposed in Koenker and Bassett Jr (1978), has received much attention due to its ability to produce a more complete picture of the conditional distribution of the response, compared to mean regression. Due to this nature, quantile regression is frequently applied to finance, medicine and biology in which researchers are often interested in the tail of the distribution to control certain risks (Tsai, 2012; Madadizadeh et al., 2016; He et al., 2016).

Although frequently used in the literature, parametric models impose stringent structural assumptions and lack the flexibility to deal with various nonlinearity present in some data sets. Semiparametric models, on the other hand, are more flexible while retaining some interpretability and efficiency of parametric models (Stone, 1985; Hastie and Tibshirani, 1993; Fan and Zhang, 1999; Cai et al., 2000; Cai and Xiao, 2012). In this work, we are focusing on the single-index models (Ichimura, 1993) which can be roughly regarded as a dimension reduction approach and the response is modelled as a nonparametric function of the constructed index. More specifically, the single-index model is given by

$$Y = g(\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}) + e,$$

where Y is the response, $\mathbf{X} = (X_1, \ldots, X_p)^{\mathrm{T}}$ is the p-dimensional predictor, g is the unknown link function, $\boldsymbol{\beta}$ is the index parameter, and e represents the error term. For quantile regression at quantile level $\tau \in (0, 1)$, we assume $P(e \leq 0|\mathbf{X}) = \tau$. In the single-index model (SIM), the reduction in dimension is achieved by the index $\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}$ and the response is only related to the predictor through the index value. The index can often be easily interpreted for various applications. The quantile single-index models have been investigated recently in Wu et al. (2010), Kong and Xia (2012) and Ma and He (2016).

* Corresponding authors.
E-mail addresses: zhouy1016@szu.edu.cn (Y. Zhou), henglian@cityu.edu.hk (H. Lian).

https://doi.org/10.1016/j.csda.2018.05.006
0167-9473/

Please cite this article in press as: Zhao W., et al., Time-varying quantile single-index model for multivariate responses. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.05.006.

Another important study worth mentioning is Christou and Akritas (2016) in which the authors proposed a novel estimation approach for quantile single-index regression which avoided the iterative steps typically used in previous approaches.

In this work, we consider multivariate responses $Y = (Y_1, \ldots, Y_q)^{\mathrm{T}}$ and we are interested in the τ conditional quantiles for all Y_l, $1 \le l \le q$, given the common predictors $\mathbf{X} = (X_1, \ldots, X_p)^{\mathrm{T}}$. Under the modelling framework of SIM, we can construct the estimators for each response separately using the specification

$$Y_l = g_l(\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_l) + e_l, \tag{1}$$

and as such the q regression problems are totally unrelated. Throughout the paper we assume p and q are fixed.

Our study is motivated by a financial application regarding daily returns of multiple portfolios, which compose our multiple responses. Nobel laureate Eugene Fama and Kenneth French have created a 5-factor model (Fama and French, 2015) to describe stock returns. This was proposed in the framework of classical linear regression model. However, linear models can be too restrictive for many data sets which do not model possible nonlinear relationships appropriately. It is thus interesting to extend this using SIM, given its flexibility and thus a potential for better fit. Using a SIM, we effectively construct an index that is a linear combination of the five factors, much as in linear models, thus retaining largely the interpretability of the "factor index". However, in such applications it may be problematic to use a different index for different portfolios. That is, it is not so intuitive to say that portfolio 1 is determined by a certain linear combination of factors, while portfolio 2 is determined by another different linear combination of factors. Thus we propose to change the multivariate regression model to

$$Y_l = g_l(\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}) + e_l,$$

where $\boldsymbol{\beta}$ is the *common* index parameter for all portfolios. We retain multiple portfolio-dependent link functions g_l to represent the variability between portfolios, which certainly have different return profiles.

Furthermore, a second extension adopted here is motivated from the literature that demonstrated the time-varyingness of similar prediction models (Zhang et al., 2013), leading to

$$Y_l = g_l(\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}, T) + e_l,$$

where T is the time variable. In the previous theoretical studies of quantile SIM, the data are assumed to be independently and identically distributed (i.i.d.), which are not suitable for financial data. Thus a final contribution in this work is to extend the theoretical results to stationary time series data under appropriate geometric mixing conditions, similar to that used in Cai et al. (2000), Cai and Wang (2008), Cai and Xiao (2012) and Cai et al. (2015).

The rest of the article is organized as follows. In Section 2, we detail the model and the estimator based on spline approximation for link functions, and the asymptotic properties are also presented. Section 3 contains some Monte Carlo studies and the results of the analysis of the portfolio daily return data. The paper concludes with a discussion in Section 4. The technical details of the proofs are relegated to the Appendix.

2. Multivariate quantile single-index models

2.1. Estimation method

We consider the single-index model

$$Y_{il} = g_l(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}, T_i) + e_{il},$$

where $(\mathbf{X}_i, T_i, Y_i = (Y_{i1}, \ldots, Y_{iq})^{\mathrm{T}}, e_i = (e_{i1}, \ldots, e_{iq})^{\mathrm{T}})$ are strictly stationary, $g_1, \ldots, g_q$ are the q link functions, $P(e_{il} \le 0|\mathbf{X}_i, T_i) = \tau$, $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})^{\mathrm{T}}$ is the p-dimensional vector of covariates, and T_i is the time variable. Besides, we always assume $\|\boldsymbol{\beta}\| = 1$ and $\beta_1 > 0$ for identifiability. We assume the joint density of $(T_i, \mathbf{X}_i)$ exists on its support, which in particular excludes the degenerate case that T_i is a deterministic function of $\mathbf{X}_i$. Other than that, the dependence between T_i and $\mathbf{X}_i$ can be arbitrary.

We use polynomial splines to approximate the components. Assuming $\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}$ is supported on $[a, b]$ (in practice, given a current estimate of $\boldsymbol{\beta}$, we can set a and b to be the minimum and maximum value of $\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}$, respectively), we use polynomial splines to approximate the link function. Let $\tau_0 = a < \tau_1 < \cdots < \tau_{K_1'} < b = \tau_{K_1'+1}$ be a partition of $[a, b]$ into subintervals $[\tau_l, \tau_{l+1})$, $l = 0, \ldots, K_1'$ with K_1' internal knots. We only restrict our attention to equally spaced knots although data-driven choice can be considered such as putting knots at certain sample quantiles of the observed covariate values. A polynomial spline of order s_1 is a function whose restriction to each subinterval is a polynomial of degree $s_1 - 1$ and globally $s_1 - 2$ times continuously differentiable on $[a, b]$. The collection of splines with a fixed sequence of knots has a B-spline basis $\mathbf{b}_1(x) = \{b_{1,1}(x), \ldots, b_{1,K_1}(x)\}$ with $K_1 = K_1' + s_1$. Similarly, assuming without loss of generality that T_i is supported on $[0, 1]$, we can construct B-spline basis $\mathbf{b}_2(t) = \{b_{2,1}(t), \ldots, b_{2,K_2}(t)\}$. We assume B-spline basis is normalized to have $\sum_{l=1}^{K_1} b_{1,l}(x) = \sqrt{K_1}$ and $\sum_{l=1}^{K_2} b_{2,l}(x) = \sqrt{K_2}$. Such normalization is not essential and is only imposed to simplify some expressions in theoretical derivations later. Finally, for a bivariate function supported on $[a, b] \times [0, 1]$, we construct the tensor basis $\mathbf{B}(x, t) = (B_1(x, t), \ldots, B_K(x, t))^{\mathrm{T}} = (b_{1,1}(x)b_{2,1}(t), \ldots, b_{1,K_1}(x)b_{2,K_2}(t))^{\mathrm{T}}$ where $K = K_1K_2$.

Please cite this article in press as: Zhao W., et al., Time-varying quantile single-index model for multivariate responses. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.05.006.