ARTICLE IN PRESS

Journal of the Korean Statistical Society ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

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

Quantile regression for robust inference on varying coefficient partially nonlinear models

Jing Yang [a,*], Fang Lu [a], Hu Yang [b]

[a] Key Laboratory of High Performance Computing and Stochastic Information Processing (Ministry of Education of China), College of Mathematics and Computer Science, Hunan Normal University, Changsha, 410081, China

[b] College of Mathematics and Statistics, Chongqing University, Chongqing, 401331, China

ARTICLE INFO

Article history:
Received 25 March 2017
Accepted 3 December 2017
Available online xxxx

AMS 2000 subject classifications:
62G05
62G08

Keywords:
Varying coefficient partially nonlinear models
Quantile regression
Asymptotic properties
Robustness

ABSTRACT

In this paper, we propose a robust statistical inference approach for the varying coefficient partially nonlinear models based on quantile regression. A three-stage estimation procedure is developed to estimate the parameter and coefficient functions involved in the model. Under some mild regularity conditions, the asymptotic properties of the resulted estimators are established. Some simulation studies are conducted to evaluate the finite performance as well as the robustness of our proposed quantile regression method versus the well known profile least squares estimation procedure. Moreover, the Boston housing price data is given to further illustrate the application of the new method.

1. Introduction

Varying coefficient partially linear models (VCPLMs) have been widely studied due to their merits including the interpretability of parametric models and the flexibility of nonparametric models. Related literature on VCPLMs include but not be restricted to Ahmad, Leelahanon, and Li (2005), Fan and Huang (2005), Kai, Li, and Zou (2011), Wang, Zhu, and Zhou (2009) and Xia, Zhang, and Tong (2004). To capture some more complicated potential relationship between the response variable and the covariates, Li and Mei (2013) proposed the so-called varying coefficient partially nonlinear models (VCPNLMs) via replacing the linear component of VCPLMs by a nonlinear function of the covariates. Specifically, the VCPNLMs can be written as

$$Y = X^T\alpha(U) + g(Z, \beta) + \varepsilon, \tag{1}$$

where "T" represents the transpose of a vector or matrix throughout this paper, $(X, Z) \in R^p \times R^s$ and $U \in R$ are the associated covariates, $\alpha(\cdot) = (\alpha_1(\cdot), \ldots, \alpha_p(\cdot))^T$ is a p-dimensional vector of unknown coefficient functions, $g(\cdot, \cdot)$ is a previously known nonlinear function, $\beta = (\beta_1, \ldots, \beta_q)^T$ is a vector of unknown coefficients that do not necessarily have the same dimension with Z, and ε is the random error. In addition, we assume the first component of X is 1 and thus no separate intercept is explicitly written.

Model (1) is a very flexible model that contains many submodels such as nonlinear model, partially nonlinear model, varying coefficient model and varying coefficient partially linear model as special cases, then it is valuable and meaningful

* Corresponding author.
E-mail address: yang2009jing@163.com (J. Yang).

https://doi.org/10.1016/j.jkss.2017.12.002
1226-3192/

to do some statistical inference on VCPNLMs. Regretfully, there exists little literature about this study up to now. Li and Mei (2013) proposed a profile nonlinear least squares estimation approach for the parameter vector β and coefficient function vector $\alpha(\cdot)$, and further established the asymptotic properties of the corresponding estimates. Yang and Yang (2016) presented a two-stage estimation algorithm by employing an orthogonality-projection-based method for the estimation of parametric coefficient, and developed a novel variable selection procedure via smooth-threshold estimating equations for the coefficient functions. Qian and Huang (2016) considered the corrected profile least-squared estimation procedure for the VCPNLMs with measurement errors in the nonparametric part, and proposed a generalized likelihood ratio test for the varying coefficient part to check whether the coefficient functions are a constant or not.

It is worth noting that the above mentioned literature on model (1) are closely related to the least squares (LS) approach. As we known, although LS method is the best one in the case of normal distributed data set, it is not robust and probably produce large bias when there exist outliers in the data set or the error follows some heavy-tailed distribution. Worse still, the LS based estimator will break down under the error of Cauchy distribution because it is no longer a consistent estimator now. Therefore, it is of interest and desirable to develop some robust statistical methodology for VCPNLMs and the quantile regression (QR) proposed by Koenker and Bassett (1978) has shown to be the most popular alternative way to overcome the drawback of LS method. Related literature on QR method can be referred to, for instance, Honda (2004), Kai et al. (2011), Koenker (2005), Koenker and Bassett (1978), Kong and Xia (2012), Wu, Yu, and Yan (2010) Yu and Jones (1998), and Yu and Lu (2004), for a comprehensive review. Consequently, motivated by these observations, we devoted to extending the QR method to VCPNLMs in this paper, and further illustrate its empirical success through some simulation studies. To the best of our knowledge, this is the first attempt at providing robust estimate via quantile regression for model (1).

The rest of this paper is organized as follows. In Section 2, we first present the detailed estimation procedure of model (1) based on quantile regression approach. Asymptotic properties of the resulted estimators are established in Section 3 under some suitable conditions. Some simulation studies and a real data analysis are conducted in Section 4 to evaluate the performance of our proposed methods, and then followed by a short conclusion in Section 5. All technical proofs are collected in the Appendix.

2. Quantile regression for VCPNLMs

Quantile regression is often used to estimate the conditional quantile of the response variable Y, which has the following definition

$$Q_\tau(x,z,u) = \arg\min_a E\{\rho_\tau(Y-a) \mid (X,Z,U)=(x,z,u)\},$$

where $\rho_\tau(t) = t[\tau - I(t<0)]$ is the check loss function for $\tau \in (0,1)$. For model (1), $Q_\tau(x,z,u)$ can be expressed as $x^T\alpha_\tau(u) + g(z,\beta_\tau)$, where a baseline function is included in $\alpha_\tau(u)$. Thus, for any given τ, quantile regression can be applied to estimate the varying coefficient functions $\alpha_\tau(\cdot)$ and the parameter β_τ.

Suppose that $\{X_i, Y_i, Z_i, U_i\}_{i=1}^n$ are independent identically distributed (i.i.d.) random samples from

$$Y = X^T\alpha_\tau(U) + g(Z,\beta_\tau) + \varepsilon_\tau, \tag{2}$$

where ε_τ is the random error whose conditional τth quantile equals to zero. The QR estimates of $\alpha_\tau(\cdot)$ and β_τ can be obtained by minimizing the following quantile loss function

$$\sum_{i=1}^n \rho_\tau\{Y_i - g(Z_i,\beta) - X_i^T\alpha(U_i)\}. \tag{3}$$

For the convenience of presentation, we will omit τ from β_τ, $\alpha_\tau(\cdot)$ and ε_τ in model (2) wherever clear from the context, but we should keep in mind that those quantities are τ-specific.

Note that the above defined objective function (3) involves both the nonparametric and parametric components which should be estimated with different rates of convergence, so it is relatively hard to solve the minimization problem. To this end, we divide the detailed solution method into a three-stage estimation procedure. Specifically, for U in a neighborhood of u, the unknown coefficient functions $\alpha_j(U)$ can be locally linearly approximated as

$$\alpha_j(U) \approx \alpha_j(u) + \alpha_j'(u)(U-u) \triangleq a_j + b_j(U-u)$$

for $j=1,\ldots,p$. Denote by $\{\tilde\beta, \tilde a, \tilde b\}$ the minimizer of the locally weighted quantile loss function

$$\sum_{i=1}^n \rho_\tau\{Y_i - g(Z_i,\beta) - X_i^T[a + b(U_i-u)]\}K_h(U_i-u), \tag{4}$$

where $a=(a_1,\ldots,a_p)^T$, $b=(b_1,\ldots,b_p)^T$, $K_h(\cdot)=K(./h)/h$ with $K(\cdot)$ is a kernel weight function and h is the bandwidth. Then $\tilde a = \tilde\alpha(u)$ and $\tilde b = \tilde\alpha'(u)$.