Statistics and Probability Letters 96 (2015) 21–27

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Polynomial spline approach for variable selection and estimation in varying coefficient models for time series data

Peng Lai [a], Jie Meng [b], Heng Lian [c,*]

[a] School of Mathematics and Statistics, Nanjing University of Information Science & Technology, Nanjing 210044, China

[b] Department of Marketing and Management, Macquarie University, North Ryde, NSW 2109, Australia

[c] Division of Mathematical Sciences, SPMS, Nanyang Technological University, Singapore, 637371, Singapore

ARTICLE INFO

Article history:
Received 20 March 2014
Received in revised form 2 September 2014
Accepted 8 September 2014
Available online 17 September 2014

Keywords:
Autoregressive models
BIC
B-spline basis
SCAD penalty

ABSTRACT

We propose the penalized estimator with the smoothly clipped absolute deviation (SCAD) penalty for varying coefficient time series models, which in autoregressive models actually performs lag order selection. Theoretical properties are established. Some numerical examples are also presented.

1. Introduction

In an effort to address the model misspecification issues associated with parametric models, more flexible nonlinear and non-/semi-parametric models have been used for both independent and time series data. For example, Stone (1986) considered the additive model. Hastie and Tibshirani (1993) studied the varying coefficient model. Li (2000) and Ahmad et al. (2005) considered the partially linear additive model and partially linear varying coefficient model respectively, for independent data. For time series, see Fan and Yao (2003) for an excellent review.

This article adopts a similar context as Cai et al. (2000), which considered varying coefficient models for stationary time series data under α-mixing conditions. Cai et al. (2000) motivated varying coefficient model for time series data by the lynx data which concerns the annual fur returns of lynx at auction in 1821–1934. For this data, Tong (1990) fitted the threshold autoregressive model with two regimes where different autoregressive models are obtained depending on the value of the delay variable at lag 2. Since the threshold model can be regarded as a simplification of the varying behavior into two state, Cai et al. (2000) considered fitting the varying coefficient autoregressive model $x_t = \beta_1(x_{t-2})x_{t-1} + \beta_2(x_{t-2})x_{t-2} + \epsilon_t$ where now β_1 and β_2 are functions that reflect the smooth changing behavior when the population increases or decreases.

Different from Cai et al. (2000) which used kernel estimation methods for estimating the time-varying coefficients, here we use series estimation methods, in particular B-splines, and we also consider penalized variable selection for this model. In the semiparametric time series literature with α-mixing sequences, series estimation methods have so far received relatively little attention in developing its asymptotic properties. One contribution of this article is to fill this gap. Series estimation methods, in particular polynomial splines, for nonparametric models have been used often in statistics/econometrics, for example in Stone (1986); Andrews (1991); Andrews and Whang (1990); Chen (2007); Donald and Newey (1994) and Newey

* Corresponding author.
E-mail address: henglian@ntu.edu.sg (H. Lian).

http://dx.doi.org/10.1016/j.spl.2014.09.008

(1997). Some advantages of series estimation methods were mentioned in Li (2000) including its computational efficiency. A notable disadvantage is that asymptotic normality of the nonparametric components is difficult to establish.

For the special case of varying coefficient autoregressive models for example, an important problem is to determine the lag order (order 2 used in the lynx data), which is a challenging problem in this area. In recent years, variable selection based on regularization methods has attracted a lot of interest (Tibshirani, 1996; Fan and Li, 2001; Zou, 2006; Yuan and Lin, 2006; Zou and Li, 2008). Extensions of the regularization framework to varying coefficient models include Wang et al. (2008); Wang and Xia (2009) and Li and Liang (2008), all for independent data. We will study variable selection problem for varying coefficient times series model. In the autoregressive models, this actually has the effect of lag order selection when an initial sufficiently large order is specified.

The rest of the paper is organized as follows. In Section 2, we formally present the varying coefficient model, the estimation procedure, and the statistical properties. We show the nonparametric oracle property of the penalized estimator. As defined in Storlie et al. (2011), nonparametric oracle property means that we can consistently identify the nonzero coefficients and the convergence rate of the estimator is the same as when the zero coefficients are known a priori. We then discuss the computational approach adopted. In Section 3, we present some numerical results for finite sample performance. Section 4 contains some concluding remarks. All technical proofs are relegated to the Supplementary material (see Appendix A).

2. Estimation method and asymptotic property

2.1. Spline estimation of varying coefficient models

Let $(\mathbf{X}_i, U_i, Y_i)$ be jointly strictly stationary processes. We consider the varying coefficient model given by

$$Y_i = \mathbf{X}_i^T \boldsymbol{\beta}(U_i) + \epsilon_i,$$

where $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})^T$ is p-dimensional, and $E[\epsilon_i|\mathbf{X}_i, U_i] = 0$, and U_i is called the smoothing variable, which might or might not be one component of $\mathbf{X}_i$. We only consider one-dimensional smoothing variable $U_i \in R$. Although multi-dimensional U_i can possibly be accommodated, in practice this is rare due to the worry of curse of dimensionality in high dimensional nonparametric regression. We assume the smoothing variable U_i takes values in a bounded interval $[-T, T]$, which is typical in series estimation methods. Although our approach works for general series estimation methods under mild assumptions, for specificity and ease of presentation, we only consider polynomial splines in the following.

Let $\xi_0 = -T < \xi_1 < \cdots < \xi_{K'} < T = \xi_{K'+1}$ partition $[-T, T]$ into subintervals $[\xi_k, \xi_{k+1}),\ k = 0, \ldots, K'$ with K' internal knots. We only restrict our attention to equally spaced knots although data-driven choice can be considered such as using the quantiles of $\{U_i\}$. A polynomial spline of order d' is a function whose restriction to each subinterval is polynomial of degree $d'-1$ and globally $d'-2$ times continuously differentiable. The collection of splines with a fixed sequence of knots has a normalized B-spline basis $\{B_1(u), \ldots, B_K(u)\}$ with $K = K' + d'$.

Suppose each coefficient $\beta_j(u)$ can be approximated by an expansion $\beta_j(u) \approx \sum_{k=1}^K b_{jk} B_k(u)$. Although in principle different coefficients might be approximated by splines with a different number of knots, in practice it is hard to choose many knots sequences simultaneously and thus for simplicity we assume the same basis is used for different coefficients.

We consider the least squares criterion

$$Q(\mathbf{b}) = \frac{1}{n}\sum_{i=1}^{n}\left[Y_i - \sum_{j=1}^{p}\sum_{k=1}^{K} X_{ij}B_k(U_i)b_{jk}\right]^2.$$

Let $\mathbf{b} = (\mathbf{b}_1^T, \ldots, \mathbf{b}_{p_1}^T)^T = (b_{11}, \ldots, b_{p_1K})^T$, $\mathbf{Z}_i = (X_{i1}B_1(U_i), X_{i1}B_K(U_i), \ldots, X_{ip_1}B_K(U_i))^T$, $\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_n)^T$, $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$, we can rewrite the criterion as

$$Q(\mathbf{b}) = \frac{1}{n}\|\mathbf{Y} - \mathbf{Z}\mathbf{b}\|^2. \quad (1)$$

Denoting the minimizer by $\hat{\mathbf{b}}$, we estimate $\beta_j(u)$ by

$$\hat{\beta}_j(u) = \sum_{k=1}^{K} \hat{b}_{jk} B_k(u).$$

We impose the following assumptions in studying the statistical properties of the estimators, although we do not claim these conditions are the weakest possible.

(A1) The smoothing variable U_i has a bounded density supported on $[-T, T]$.
(A2) The eigenvalues of the matrix $E(\mathbf{X}_i\mathbf{X}_i^T|U_i = u)$ are bounded away from 0 and infinity, uniformly in u.
(A3) The conditional density of (U_1, U_{l+1}) given $(\mathbf{X}_1, \mathbf{X}_{l+1})$ is uniformly bounded on the support of $(\mathbf{X}_1, \mathbf{X}_{l+1})$. The conditional density of U_1 given $\mathbf{X}_1$ is uniformly bounded on the support of $\mathbf{X}_1$.