Computational Statistics and Data Analysis 130 (2019) 94–110

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Two-step estimation of time-varying additive model for locally stationary time series

Lixia Hu [a], Tao Huang [b,*], Jinhong You [b]

[a] School of Statistics and Mathematics, Shanghai Lixin University of Accounting and Finance, Shanghai, China
[b] School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, China

ARTICLE INFO

Article history:
Received 3 April 2018
Received in revised form 25 August 2018
Accepted 25 August 2018
Available online 13 September 2018

Keywords:
Time-varying additive model
Locally stationary process
α-mixing
Local linear estimator
Tensor product

ABSTRACT

In the analysis of locally stationary process, a time-varying additive model (tvAM) can effectively capture the dynamic feature of regression function. In combination with the strengths of tensor product of B-spline smoothing and local linear smoothing method, a two-step estimation method is proposed. It is shown that the proposed estimator is uniformly consistent and asymptotically oracle efficient as if the other component functions were known. Furthermore, a nonparametric bootstrap procedure is proposed to test the time-varying property of regression function. Simulation studies investigate the finite-sample performance of the proposed methods and validate the asymptotic theory. An environmental dataset illustrating the proposed method is also considered.

1. Introduction

In recent decades, parametric and nonparametric regression methods for the analysis of nonlinear time series have made great progress, more details can refer to the monographs (Fan and Yao, 2003; Tong, 1983). Classical time series analysis is built on the assumption of stationarity, which is often deviated in many application fields. A common practice is to eliminate the trend effects and the seasonal effects, and then model the transformed stationary time series. On account of the complexity of involved transformation, it is not easy to make statistical inference on the original data based upon the analysis of the transformed data. In practice, however, we may pay more attention to the statistical properties of the original data, and prefer to model the time series at hand instead of the transformed data. On the other hand, above all, in many application fields including finance, acoustics, and neuroscience, there exists non-stationary time series, which cannot be transformed to stationary process through preconditioning techniques. Therefore, it is necessary to relax the constraint of stationarity to one of non-stationarity, which still derives sensible asymptotic properties such as the consistency, asymptotic normality under proper regularization conditions. A widely accepted generalization of stationarity is the so-called locally stationary process, which is characterized by gradually time-varying second-order moments. Massive research work has been made on the locally stationary time series: parametric regression methods can refer to Dahlhaus (2012) and references therein; quantile regression methods see Draghicescu et al. (2009), Zhou (2010) and Zhou and Wu (2009); and nonparametric regression methods see Vogt (2012); Zhang and Wu (2012, 2015); Zhou and Wu (2010).

The classical additive model (AM) introduced by Breiman and Friedman (1985) is a useful dimension-reduction tool. Meanwhile, it has the merits of flexibility, easy implementation and interpretability. Abundant research results on AM can refer to Hastie and Tibshirani (1990), Huang (1998), Huang and Yang (2004), Stone (1985, 1994) and Wang and Yang (2007). In the analysis of longitudinal data, Zhang et al. (2013) proposed a time-varying additive model (tvAM), which includes AM

* Corresponding author.
E-mail address: huang.tao@mail.shufe.edu.cn (T. Huang).

https://doi.org/10.1016/j.csda.2018.08.023
0167-9473/

as its special case. To characterize the dynamic regression function of the locally stationary process, Vogt (2012) adopted tvAM and proposed a smoothed backfitting estimate (Mammen et al., 1999) in the locally stationary context. Compared with the classical AM, the computation cost greatly increases since the backfitting estimate is done separately at each rescaled time point.

In this paper, we propose a two-step estimation method, which borrows the strengths of spline smoothing and local polynomial smoothing method. We show that the two-step estimators can consistently estimate the true component function at the convergence rate of a bivariate nonparametric function. Furthermore, the proposed estimators are oracle efficient in the sense that it has the same asymptotic property as oracle estimator as if other component functions were known in advance. The asymptotic normality of two-step estimators are considered as well.

Another practical problem alluring us is to test whether the regression function is really time-varying, namely, if the classical AM were sufficient for the real-life data at hand. We design a nonparametric bootstrap procedure to test the time-varying properties of regression function from the motivation of Cai (2007) and Fan et al. (2001). Extensive simulation studies show that the proposed testing procedure is powerful.

The organization of this paper is as follows. Section 2 presents the model setup and proposes a two-step estimation method. The asymptotic results of the newly-proposed estimators are given in Section 3. Section 4 deals with the practical problem on the smoothing parameter selection and the testing of time-varying property of regression function. Two simulation examples presented in Section 5 investigate the finite-sample performance of the proposed methods. We also illustrate our method via an environmental dataset. The brief concluding remarks are given in Section 6. The main proofs are relegated in the Appendix A, and some preliminary lemmas and propositions are shown in the Supplementary Materials.

2. Model and estimation method

2.1. Model assumptions

Let $\{Y_{t,T}, \mathbf{X}_{t,T}, t = 1, \ldots, T\}$ be length-T realization of $d+1$ dimension locally stationary time series, where $\mathbf{X}_{t,T}$ is d-vector of covariates, and the superscript τ denotes the transposition of vector or matrix. It is common to use triangular array instead of sequence in the literature of local stationary process, however, for the sake of convenience, we will suppress T in the triangular array whenever there is no confusion in the given context.

The relationship between the response and the covariates is specified by the following time-varying additive model (tvAM):

$$Y_t = m_0(t/T) + \sum_{k=1}^{d} m_k\left(t/T, X_{t,k}\right) + \sigma\left(t/T, \mathbf{X}_t\right)\varepsilon_t, \tag{1}$$

where $\{\varepsilon_t\}$ are i.i.d, satisfying $\mathbb{E}(\varepsilon_t|\mathbf{X}_t) = 0$ and $\mathbb{E}\left(\varepsilon_t^2|\mathbf{X}_t\right) = 1$, m_0 is the trend function, $m_k(k = 1, \ldots, d)$ are bivariate additive component functions, and σ is a $(d+1)$-dimensional nonparametric function allowing the heteroscedastic case. As the common practice in locally stationary context, we adopt rescaled time $u = t/T$ instead of time t. It should be pointed that the lag variables of response can be taken as candidate covariates, see, for example Huang et al. (2018) and Lei et al. (2016). Our model (1) is a more general regression method, which covers autoregression nonparametric model as a special case.

Model (1) is a flexible regression method, which includes AM as a special case, i.e., all bivariate additive component functions reduce to univariate functions independent of time, and the trend function m_0 becomes a constant. One of advantages of model (1) is that we can handle stationary process, trend stationary and locally stationary process in a unified framework, and thus avoid the trouble of testing stationarity and preconditioning the original data.

We specify the identification condition of bivariate additive component function m_k as follows:

$$\mathbb{E}[m_k(u, X_{t,k}(u))] = 0, \ \forall\, u \in [0, 1],$$

where $X_{t,k}(u)$ is the stationary approximation of locally stationary covariates $X_{t,k}$ at any given rescaled time u. On account of the data sparsity at the tail of the distribution of covariates, we will estimate each component function on a compact subset of the support.

2.2. Two-step estimation method

In the analysis of stationary time series, Wang and Yang (2007) used AM and proposed a spline-backfitted kernel estimate, which combines the strengths of spline smoothing and kernel smoothing. However, the local constant smoothing in the second estimation step still faces the boundary effect problem. In this subsection, we propose a spline-backfitted local linear estimate, which extends the method of Wang and Yang (2007) to the locally stationary context and avoid the boundary effect via local linear fitting in the second estimation step.

Spline method is a popular tool to fit a smooth nonparametric function, and B-spline basis is commonly used because of its computational stability. Let $\{B_1(x), \ldots, B_{N+q}(x)\}$ be q order B-spline basis with N interior knots. In this paper, we adopt linear B-spline basis (with smooth degree $q = 2$), denoted as $B_J(x) = \{B_1(x), \ldots, B_{N+2}(x)\}$. Similar to Wang and Yang (2007),