Journal of Statistical Planning and Inference 142 (2012) 1678–1689

Contents lists available at SciVerse ScienceDirect

Journal of Statistical Planning and Inference

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

Polynomial spline confidence bands for time series trend

Q. Shao [a,*], L. Yang [b,c,1]

[a] Department of Mathematics and Statistics, University of Toledo, Toledo OH 43606, USA
[b] Center for Advanced Statistics and Econometrics Research, School of Mathematical Sciences, Soochow University, Suzhou 215006, China
[c] Department of Statistics and Probability, Michigan State University, East Lansing MI 48824, USA

ARTICLE INFO

Article history:
Received 25 February 2010
Received in revised form
19 December 2011
Accepted 14 February 2012
Available online 22 February 2012

Keywords:
Autoregressive time series
B-splines
Confidence band

ABSTRACT

The paper considers the construction of a confidence band for the trend function of a stationary time series. An explicit formula is derived based on polynomial splines and Sunklodas (1984). The performance of the confidence band is illustrated by simulation studies. The proposed method is applied to the analysis of the annual yields of wheat in the United States.

1. Introduction

Inference on trend functions is one of the classic topics in time series analysis. In this paper, we are interested in constructing a confidence band for a smooth trend function of time series observations $\{y_i, i=1,\ldots,n\}$ as follows:

$$y_i = g(u_i)+x_i, \tag{1.1}$$

where $g(\cdot)$ represents the trend function defined in the interval $[0,1]$ with $u_i = i/n$ and the zero-mean error term x_i is an autoregressive time series of order p (AR(p)) defined by

$$x_t = \sum_{k=1}^{p} \phi_k x_{t-k} + \epsilon_t,$$

where $\{\epsilon_t\}$ is independent and identically distributed (IID) white noise with mean 0 and variance σ^2. To facilitate the discussion, we introduce the vector format of model (1.1) as follows:

$$\mathbf{y} = \mathbf{g} + \mathbf{x},$$

where $\mathbf{y} = (y_1,\ldots,y_n)'$, $\mathbf{g} = (g(u_1),\ldots,g(u_n))'$. Throughout this paper, we will use bold lower-case letters to denote vectors, bold upper-case letters to denote matrices, and lower-case letters to denote both time series and their realizations.

Analysis of the trend function in model (1.1) with an autoregressive error term has received intensive attention due to its wide applications. The classic approach in stationary time series analysis is that the trend is assumed to be a parametric function with known analytical form and unknown parameters. See for example Chapter 9 of Fuller (1996) for details.

* Corresponding author.
E-mail addresses: qin.shao@utoledo.edu (Q. Shao), yanglijian@suda.edu.cn, yang@stt.msu.edu (L. Yang).
[1] His research was supported by NSF awards DMS 0706518 and DMS 1007594, and funding from the "Jiangsu-Specially Appointed Professor Program", Jiangsu Province, China.

0378-3758/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2012.02.028

Although this parametric approach is appropriate for many applications in practice, its major drawback is that the assumption about the trend function is usually artificial. It is desirable to extend the analysis to nonparametric settings which do not need to specify the analytical form of the trend functions. In recent years as a result of advances of computing technology, a great deal of work has been devoted to local kernels and polynomial splines, two of the most commonly used nonparametric methods. It is impossible to cover the vast literature on these two methods here. Fan and Gijbels (1996) provided an overview and some details about local kernels, especially for data from independent and identical distributions. Opsomer et al. (2001), on the other hand, summarized the development of research about local kernels for correlated data. For an overview and basic theoretical results of polynomial splines, interested readers can refer to, for example, Stone (1994) and Huang (2003).

While most research concentrated on estimation, a few authors attacked construction of a confidence band for a smooth unknown function. Bickel and Rosenblatt (1973) is a pioneer work on nonparametric confidence bands for a density curve of independent and identically distributed observations. Since then several authors, such as Hall and Titterington (1988), Xia (1998), Claeskens and Van Keilegom (2003), and Wang and Yang (2009), have investigated this issue for independent observations. Wu and Zhao (2007) proposed a confidence band based on local kernels for the trend with a stationary time series error term. In this paper, we will extend the method of Wang and Yang (2009) to stationary time series trend analysis. The major contribution of this paper is to provide practitioners with the theoretical foundation of a confidence bound and a fast as well as easy to implement algorithm. The band is simultaneous and conservative in the sense that it covers the whole trend function at least with the probability of the given confidence level. It is worth mentioning that compared to the local smoothing obtained by using a kernel, spline smoothing is global, i.e., only a single optimization is needed for the unknown function over an entire range, instead of optimization at every point in the range. As a result, polynomial splines used in this paper can be thousands of times faster than kernel smoothing, which was discussed in detail in, for example, Xue and Yang (2006) and Wang and Yang (2007).

The paper will be organized as follows: in Section 2, we will introduce polynomial splines and confidence bands; in Section 3, we will illustrate by simulation studies the performance of the proposed confidence bands for the trend function in model (1.1) with several AR(1) terms, and analyze the annual yields of wheat in the United States; finally in Section 4, we will provide the details of the proofs of the theoretical results based on which confidence bands are constructed. To facilitate reading, the existing theorems that play important roles in deriving the theoretical results of this paper are provided in the Appendix.

2. Construction of confidence band

2.1. Polynomial splines

Suppose that m is a positive integer. Consider a sequence of equally spaced points or knots $(-m+1)h \le \cdots \le 0 \le h \le 2h \cdots \le Nh \le 1$. Notice that there are $N+m$ knots that divide the interval $[(-m+1)h,1]$ into subintervals $J_j=[jh,(j+1)h), j=-m+1,-m+2,\ldots,N-1$ and $J_N=[Nh,1]$, of width h. For any given $u \in [0,1]$, $j(u)$ is the knot corresponding to the interval that includes u. Let $G_N^{(m-2)}=G_N^{(m-2)}[0,1]$ denote the space of functions that are polynomial of degree $m-1$ on each J_j and have continuous $(m-2)$th derivatives. The B-spline basis of $G_N^{(m-2)}$ is $\mathbf{b}_m(u)=(b_{j,m}(u), j=-m+1,\ldots,N)'$. For any function $\varphi(\cdot)$ in $L^2[0,1]$ define the norm as

$$\|\varphi\|_2^2=\int_0^1 \varphi^2(x)\,dx.$$

The B-spline standardized basis $\mathbf{c}_m(u)=(c_{j,m}(u), j=-m+1,\ldots,N)'$ is defined as

$$c_{j,m}(u)=\frac{b_{j,m}(u)}{\|b_{j,m}\|_2}=\frac{b_{j,m}(u)}{\{\int_0^1 b_{j,m}^2(u)\,du\}^{1/2}}. \tag{2.1}$$

For a realization of time series $\mathbf{y}$, define a vector $\mathbf{c}_{j,m}=(c_{j,m}(u_1),\ldots,c_{j,m}(u_n))'$ with

$$c_{j,m}(u_i)=\frac{b_{j,m}(u_i)}{\|b_{j,m}\|_2},$$

and an $n \times (N+m)$ matrix

$$\mathbf{C}_m=(\mathbf{c}_{-m+1},\ldots,\mathbf{c}_N).$$

We will focus on the cases of $m=1,2$: $G_N^{(-1)}$ is the space of functions that are constant on each J_j, and $G_N^{(0)}$ is the space of functions that are linear on each J_j and continuous on [0,1].

The B-spline basis for $G_N^{(-1)}$ is $\mathbf{b}_1(u)=(b_{j,1}(u), j=0,\ldots,N)'$, where $b_{j,1}(u)$ is the indicator function of J_j, i.e.

$$b_{j,1}(u)=\begin{cases}1, & j=j(u),\\ 0 & \text{otherwise}.\end{cases}$$