Statistical Methodology 21 (2014) 35–48

Contents lists available at ScienceDirect

Statistical Methodology

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

Bayesian inference in nonparametric dynamic state-space models

Anurag Ghosh [a], Soumalya Mukhopadhyay [b], Sandipan Roy [c], Sourabh Bhattacharya [d,*]

[a] *Department of Statistical Science, Duke University, United States*

[b] *Agricultural and Ecological Research Unit, Indian Statistical Institute, India*

[c] *Department of Statistics, University of Michigan, Ann Arbor, United States*

[d] *Bayesian and Interdisciplinary Research Unit, Indian Statistical Institute, 203, B. T. Road, Kolkata 700108, India*

ARTICLE INFO

Article history:
Received 6 September 2013
Received in revised form
16 January 2014
Accepted 19 February 2014

Keywords:
Evolutionary equation
Gaussian process
Look-up table
Markov chain Monte Carlo
Observational equation
State-space model

ABSTRACT

We introduce state-space models where the functionals of the observational and evolutionary equations are unknown, and treated as random functions evolving with time. Thus, our model is nonparametric and generalizes the traditional parametric state-space models. This random function approach also frees us from the restrictive assumption that the functional forms, although time-dependent, are of fixed forms. The traditional approach of assuming known, parametric functional forms is questionable, particularly in state-space models, since the validation of the assumptions require data on both the observed time series and the latent states; however, data on the latter are not available in state-space models.

We specify Gaussian processes as priors of the random functions and exploit the "look-up table approach" of Bhattacharya (2007) to efficiently handle the dynamic structure of the model. We consider both univariate and multivariate situations, using the Markov chain Monte Carlo (MCMC) approach for studying the posterior distributions of interest. We illustrate our methods with simulated data sets, in both univariate and multivariate situations. Moreover, using our Gaussian process approach we analyze a real data set, which has also been analyzed by Shumway & Stoffer (1982) and Carlin, Polson & Stoffer (1992) using the linearity as-

* Corresponding author. Tel.: +91 3325752803.
E-mail address: sourabh@isical.ac.in (S. Bhattacharya).

http://dx.doi.org/10.1016/j.stamet.2014.02.004

sumption. Interestingly, our analyses indicate that towards the end of the time series, the linearity assumption is perhaps questionable.

1. Introduction

The state-space models play important role in dealing with dynamic systems that arise in various disciplines such as finance, engineering, ecology, medicine, and statistics. The time-varying regression structure and the flexibility inherent in the sequential nature of state-space models make them very suitable for analysis and prediction of dynamic data. Indeed, as is well-known, most time series models of interest are expressible as state-space models; see [7,13] for details. However, till date, the state-space models have considered only known forms of the equations, typically linear. But testing the parametric assumptions require data on both the observed time series and the unobserved states; unfortunately, data on the latter are not available in state-space models. Moreover, the regression structures of the state-space models may evolve with time, changing from linear to non-linear, and even the non-linear structure may also evolve with time, yielding further different non-linear structures. We are not aware of any nonparametric state-space approach in the statistical literature that can handle unknown functional forms, which may or may not be evolving with time. Another criticism of the existing state space models is the assumption that the (unobserved) states satisfy the Markov property. Although such Markov models have been useful in many situations where there are natural laws supporting such conditional independence, in general such assumption is not expected to hold. These arguments point towards the need for developing general, nonparametric, approaches to state-space models, and this indeed, is our aim in this article. We adopt the Bayesian paradigm for its inherent flexibility.

In a nutshell, in this work, adopting a nonparametric Bayesian framework, we treat the regression structures as unknown and model these as Gaussian processes, and develop the consequent theory in the Bayesian framework, considering both univariate and multivariate situations. Our Gaussian process approach of viewing the unknown functional forms allows very flexible modeling of the unknown structures, even though they might evolve with time. Also, as we discuss in Section 4.7, as a consequence of our nonparametric approach, the unobserved state variables do not follow any Markov model. Thus our approach provides a realistic dependence structure between the state variables. We also develop efficient MCMC-based methods for simulating from the resulting posterior distributions. We demonstrate our methods in the case of both univariate and multivariate situations using simulated data. Application of our ideas to a real data set which has been analyzed by Shumway and Stoffer [12] and Carlin et al. [3] assuming linearity, provided an interesting insight that, although the linearity assumption may not be unreasonable for most part of the time series, the assumption may be called in question towards the end of the time series. This vindicates that our approach is indeed capable of modeling unknown functions even if the forms are changing with time, without requiring any change point analysis and specification of functional forms before and after change points.

Before introducing our approach, we provide a brief overview of state-space models.

2. Overview of state-space models

Generally, state-space models are of the following form: for $t = 1, 2, \ldots,$

$$y_t = f_t(x_t) + \epsilon_t \tag{1}$$

$$x_t = g_t(x_{t-1}) + \eta_t. \tag{2}$$

In the above, f_t and g_t are assumed to be functions of known forms which may or may not explicitly depend upon t; η_t, ϵ_t are usually assumed to be zero mean *iid* normal variates. The choice $f_t(x_t) = F_t x_t$ and $g_t(x_{t-1}) = G_t x_{t-1}$, assuming known F_t, G_t, have found very wide use in the literature. Obviously,