Computational Statistics and Data Analysis 74 (2014) 64–80

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

A dynamic linear model with extended skew-normal for the initial distribution of the state parameter

Celso Rômulo Barbosa Cabral [a,*], Cibele Queiroz da-Silva [b], Helio S. Migon [c]

[a] Departamento de Estatística, Universidade Federal do Amazonas, Brazil

[b] Departamento de Estatística, Universidade de Brasília, Brazil

[c] Departamento de Estatística, Universidade Federal do Rio de Janeiro, Brazil

ARTICLE INFO

Article history:
Received 14 April 2012
Received in revised form 14 December 2013
Accepted 24 December 2013
Available online 9 January 2014

Keywords:
Dynamic models
Skew-normal distribution
Bayesian analysis
Kalman filter
Gibbs sampler

ABSTRACT

We develop a Bayesian dynamic model for modeling and forecasting multivariate time series relaxing the assumption of normality for the initial distribution of the state space parameter, and replacing it by a more flexible class of distributions, which we call *Generalized Skew-Normal (GSN) Distributions*. We develop a version of the classic Kalman filter, again obtaining GSN predictive and filtering distributions. As we are supposing the random fluctuations covariances to be unknown, a Gibbs-type sampler algorithm is developed in order to perform Bayesian inference. We work with two simulation experiments with scenarios close to real problems in order to show the efficacy of our proposed model. Finally, we apply our technique to a real data set.

1. Introduction

1.1. The normal dynamic linear model

A *Dynamic Linear Model* (hereafter DLM) is a particular case of the *State Space Model*. As a general rule, we can say that, for each $t = 1, \dots, n$, in a state space model we have observations of an r-dimensional time series $\{\mathbf{Y}_t\}$, which is a function of an unobservable p-dimensional random variable $\boldsymbol{\theta}_t$ (called *The State of Nature at Time t*) and of a measurement error $\boldsymbol{\nu}_t$. At the same time, $\boldsymbol{\theta}_t$ follows an evolution process that depends on $\boldsymbol{\theta}_{t-1}$ and a random fluctuation $\mathbf{w}_t$ or, in more formal notation,

$$\mathbf{Y}_t = f_t(\boldsymbol{\theta}_t, \boldsymbol{\nu}_t), \quad \boldsymbol{\theta}_t = g_t(\boldsymbol{\theta}_{t-1}, \mathbf{w}_t),\ t = 1, 2, \dots, n, \tag{1}$$

where $\{f_t\}$ and $\{g_t\}$ are convenient sequences of functions.

There are a variety of phenomena that can be described according to this structure. This model is very general and encompasses several well-known models available in the literature of time series, like ARMA models. For more details about theory and applications in several areas, see the highly recommended texts of West and Harrison (1997) and Petris et al. (2009).

We are interested specifically in the particular class of *Dynamic Linear Models*. In this class, for each $t \in \{1, 2, \dots, n\}$, we observe an r-dimensional random vector $\mathbf{Y}_t$, such that

$$\mathbf{Y}_t = \mathbf{F}_t'\boldsymbol{\theta}_t + \boldsymbol{\nu}_t, \quad \text{with } \boldsymbol{\nu}_t \sim \mathrm{N}_r(0, \mathbf{V}); \tag{2}$$

$$\boldsymbol{\theta}_t = \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \mathbf{w}_t, \quad \text{with } \mathbf{w}_t \sim \mathrm{N}_p(0, \mathbf{W}); \tag{3}$$

$$\boldsymbol{\theta}_0 \sim \mathrm{N}_p(\mathbf{m}_0, \mathbf{C}_0), \tag{4}$$

* Correspondence to: ICE – Departamento de Estatśtica, Av. General Rodrigo Otávio, 3000, Campus Universitário, Coroado I. Manaus, Amazonas, 69077-000, Brazil.

E-mail addresses: celsoromulo@gmail.com, celsoromulo@ufam.edu.br (C.R.B. Cabral), cibeleqs@unb.br (C.Q. da-Silva), migon@ime.ufrj.br (H.S. Migon).

0167-9473/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.csda.2013.12.008

where $\mathbf{Y}_t$ and $\boldsymbol{\theta}_t$ are as before, $\mathbf{F}_t : p \times r$ is a known matrix, $\mathbf{G}_t : p \times p$ is also a known matrix, called *Evolution Matrix*, $\mathbf{V} : r \times r$ and $\mathbf{W} : p \times p$ are (possibly) unknown error covariance matrices, $\mathbf{m}_0 : p \times 1$ and $\mathbf{C}_0 : p \times p$ are, respectively, the mean vector and the (positive definite) covariance matrix of the *Initial Distribution of* $\boldsymbol{\theta}_0$. We call this model the *Normal (or Gaussian) DLM*.

We need some additional assumptions, listed below:

(i) $\mathbf{w}_t$ and $\boldsymbol{\theta}_{t-1}$ are uncorrelated for all t;
(ii) $\boldsymbol{\nu}_t$ and $\boldsymbol{\nu}_s$ are uncorrelated for all $t \neq s$;
(iii) $\mathbf{w}_t$ and $\mathbf{w}_s$ are uncorrelated for all $t \neq s$;
(iv) $\boldsymbol{\nu}_t$ and $\mathbf{w}_s$ are uncorrelated for all t and for all s;
(v) $\boldsymbol{\theta}_0$ and $\boldsymbol{\nu}_t$ are uncorrelated for all t;
(vi) $\boldsymbol{\theta}_0$ and $\mathbf{w}_t$ are uncorrelated for all t.

As observed by Naveau et al. (2005), the assumption of normality can be questionable for a large number of applications see, for example, Grabek et al. (2011) and Higgs (2011). In this work, we discuss the case when the initial distribution of the state space parameter (that is, the distribution of $\boldsymbol{\theta}_0$) is possibly skewed. Our purpose is to replace the normal distribution in (4) by a more flexible one, which is an element of a huge family of distributions that also contains the normal one. Using this simple method, we obtain natural extensions of the classic Kalman filter and of the Gibbs sample procedure used for inferences of some unknown quantities in the normal case. With the help of simulated and real data we show that our method has clear advantage over the normal DLM when dealing with data with possibly skewed nature.

Naveau et al. (2005) obtained a Kalman filter that accommodates skew-normal observational errors in DLMs, but unlike their work we explore the inferential aspects of the problem more deeply. We provide not only the filtered estimates of the states $\boldsymbol{\theta}_t$, as is the case in Naveau et al. (2005), but also expressions for the one-step-ahead predictive distribution of $\mathbf{Y}_t$ and a simple Gibbs sampler scheme to generate draws from the posterior distribution of the proposed model that allow us to estimate some unknown quantities like the observation and evolution covariance matrices. Additionally we develop a method to obtain smoothed estimates of the state vector.

The paper is organized as follows. In Section 2 we present the family of fundamental skew-normal distributions, which is the most important concept to us, because all the predictive and filtering distributions belong to this class. In Section 3 we present our extension of the normal DLM by introducing a skewed prior for the state parameter. In Section 4 we present a Gibbs-type algorithm to perform Bayesian inference. In Section 5 we show the efficiency of our method by analyzing two artificial and one real data set. Finally, in Section 7, we present some discussions and conclusions.

2. Fundamental skew-normal distributions

2.1. Basic concepts

A skew-normal distribution is a distribution that extends the normal one by the introduction of additional parameters that regulate skewness. Some versions, extensions and unifications of the skew-normal distribution are carefully surveyed in works like Azzalini (2005) and Arellano-Valle and Azzalini (2006). In our proposed model, to be presented in the next sections, we work with a subclass of the family of skew distributions presented in Arellano-Valle and Genton (2005), which is called the *Fundamental Skew-Normal Distribution Family*.

In what follows, a vector $\mathbf{x}$ is a $p \times 1$ matrix, $\mathbf{x}'$ is the transpose of $\mathbf{x}$, $\mathbf{0}_p$ is the null $p \times 1$ vector and $\mathbf{I}_p$ is the identity matrix of order p (sometimes we drop the index p when there is no possibility of confusion). In the following definition, $\overset{d}{=}$ means "has the same distribution as", $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the p-variate normal distribution (dropping p when $p = 1$) with mean vector $\boldsymbol{\mu}$ and (always positive definite) covariance matrix $\boldsymbol{\Sigma}$ and, if we consider the random vector $\mathbf{X} = (X_1, \ldots, X_p)'$, we define the Borel set $(\mathbf{X} > \mathbf{0}) = (X_1 > 0, \ldots, X_p > 0)$.

Definition 1. We say that the random vector $\mathbf{Z}$ has a p-variate *Fundamental Skew-Normal (FUSN) Distribution* when $\mathbf{Z} \overset{d}{=} \mathbf{Y} \mid (\boldsymbol{\varphi} > \mathbf{0})$, where $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{\varphi}$ is an m-dimensional random vector.

There is no difficulty in showing that if $P(\boldsymbol{\varphi} > \mathbf{0}) > 0$, then $\mathbf{Z}$ has density

$$g(\mathbf{z}) = [P(\boldsymbol{\varphi} > \mathbf{0})]^{-1} \mathrm{N}_p(\mathbf{z} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) P(\boldsymbol{\varphi} > 0 \mid \mathbf{Y} = \mathbf{y}), \quad \mathbf{z} \in \mathbb{R}^p,$$

where $\mathrm{N}_p(\cdot \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the density of the $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution.

2.2. A particular FUSN distribution

In this work we have particular interest in the case of $m = 1$ and $\boldsymbol{\varphi} \sim \mathrm{N}(\xi, \eta^2)$. Now we construct this version of the FUSN distribution that will be used in our model formulation. Consider, then, the following structure: let $\mathbf{Y}$ be a p-dimensional random vector, $\boldsymbol{\mu} : p \times 1$ and $\boldsymbol{\Delta} : p \times 1$ vectors of constants, $\mathbf{V} : p \times p$ a positive definite matrix, $\xi \in \mathbb{R}$ and $\eta^2 > 0$ such that

$$\mathbf{Y} \mid \varphi \sim \mathrm{N}_p(\boldsymbol{\mu} + \boldsymbol{\Delta}\varphi, \mathbf{V}), \quad \varphi \sim \mathrm{TN}(\xi, \eta^2, (0, \infty)), \tag{5}$$

with $\mathrm{TN}(\xi, \eta^2, (a, b))$ denoting a truncated normal distribution on (a, b) (which is the distribution of $Z \mid (a < Z < b)$ when $Z \sim \mathrm{N}(\xi, \eta^2)$). Note that either $a = -\infty$ or $b = \infty$ is allowed.