



A flexible two-piece normal dynamic linear model

Emanuele Aliverti¹ · Reinaldo B. Arellano-Valle² · Fereshteh Kahrari¹ · Bruno Scarpa^{1,3} 

Received: 9 April 2022 / Accepted: 3 April 2023
© The Author(s) 2023

Abstract

We construct a flexible dynamic linear model for the analysis and prediction of multivariate time series, assuming a two-piece normal initial distribution for the state vector. We derive a novel Kalman filter for this model, obtaining a two components mixture as predictive and filtering distributions. In order to estimate the covariance of the error sequences, we develop a Gibbs-sampling algorithm to perform Bayesian inference. The proposed approach is validated and compared with a Gaussian dynamic linear model in simulations and on a real data set.

Keywords Two-piece normal distribution · Skew-normal distribution · Bayesian inference · Kalman filter · FFBS algorithm

1 Introduction

State-space models have been extensively considered in diverse areas of application for modeling and forecasting time series. An important special case is the class of dynamic linear models (hereafter DLM). This class of models includes the ordinary static linear model as a special case, and assumes that the parameters can change over time, thus incorporating in the observational system variations that can significantly affect the observed behavior of the process of interest. The DLM is defined by

$$\begin{aligned} \mathbf{Y}_t &= \mathbf{F}_t^\top \boldsymbol{\theta}_t + \mathbf{v}_t \quad (\text{observation equation}), \\ \boldsymbol{\theta}_t &= \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \quad (\text{system or state equation}), \end{aligned}$$

for $t = 1, \dots, T$, where \mathbf{Y}_t is a $r \times 1$ response vector, \mathbf{F}_t is a $p \times r$ matrix that links the observed data with $\boldsymbol{\theta}_t$, a $p \times 1$ vector of latent states at time t , \mathbf{G}_t is a $p \times p$ transition

✉ Bruno Scarpa
bruno.scarpa@unipd.it

¹ Dipartimento di Scienze Statistiche, Università di Padova, Padua, Italy

² Departamento de Estadística, Pontificia Universidad Católica de Chile, Santiago, Chile

³ Dipartimento di Matematica “Tullio Levi Civita”, Università di Padova, Padua, Italy

matrix, which describes the evolution of the state parameters. The terms \mathbf{v}_t and $\boldsymbol{\omega}_t$ are mutually independent white-noise vectors of dimension $r \times 1$ and $p \times 1$, respectively, with zero-means and constant variance-covariance matrices \mathbf{V} and \mathbf{W} , respectively. The most popular case is the Gaussian DLM, which assumes that

$$\mathbf{v}_t \stackrel{\text{ind.}}{\sim} N_r(\mathbf{0}, \mathbf{V}), \quad \boldsymbol{\omega}_t \stackrel{\text{ind.}}{\sim} N_p(\mathbf{0}, \mathbf{W}), \quad \boldsymbol{\theta}_0 \sim N_p(\mathbf{m}_0, \mathbf{C}_0),$$

all being mutually independent for each $t = 1, \dots, T$, where $N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the k -variate normal distribution with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. For an extensive introduction to DLMS with a Bayesian perspective, refer to West and Harrison (1997), Petris et al. (2009). In this article, we focus on a setting where \mathbf{F}_t and \mathbf{G}_t are known, consistently with classical Kalman filter setting (Kalman 1960) and recent developments in state-space models (e.g., Fasano et al. 2021).

Leveraging the properties of the multivariate normal distribution and the structure of the Gaussian DLM, it is possible to derive closed form expressions for the predictive and filtering distributions and conduct dynamic inference on the states $\boldsymbol{\theta}_t$ via Kalman filter, conditioning on $\mathbf{F}_t, \mathbf{G}_t, \mathbf{W}$ and \mathbf{V} . However, Naveau et al. (2005) observed that the Gaussian assumption may be questionable for a large number of applications, as many distributions used in a state-space model can be skewed. In order to mitigate this issue, Naveau et al. (2005) assumed that state initial parameter vector $\boldsymbol{\theta}_0$ follows a multivariate closed skew-normal distribution, preserving the typical assumptions of independence and normality for the error sequences \mathbf{v}_t and $\boldsymbol{\omega}_t$. From this work, several authors have proposed different mechanisms to obtain DLM with more skewness. For instance, Kim et al. (2014) extended the results in Naveau et al. (2005) by assuming a scale mixtures of closed skew-normal distributions for the initial state parameter vector $\boldsymbol{\theta}_0$; Cabral et al. (2014) proposed a Bayesian DLM relaxing the assumption of normality and assuming an extended skew-normal (Azzalini and Capitanio 1999) for the initial distribution of the state parameter; Arellano-Valle et al. (2019) proposed a DLM in which the error sequence \mathbf{v}_t in the observational equation are assumed to have multivariate skew-normal distribution. Furthermore, several authors have been dealing with a similar problem; see, for example, Gualtierotti (2005), Pourahmadi (2007), Corns and Satchell (2007), among many others.

In this work we take a similar perspective, and derive a novel DLM that allows one to induce asymmetry by means of a scalar parameter, inducing a skewed initial distribution for the state space parameter $\boldsymbol{\theta}_0$. Our purpose is to replace the normal distribution for $\boldsymbol{\theta}_0$ by a more flexible one, incorporating asymmetry via a two-piece normal (TPN) mixing distribution. Using this simple method, we obtain an extension of the classic Kalman filter, and closed form expressions for one-step-ahead and filtering distributions. These results are further combined into a Markov-Chain Monte Carlo procedure via a forward filtering–backward sampling algorithm to provide inference on the covariances \mathbf{V} and \mathbf{W} of the error terms, providing posterior inference on the unknown quantities.

2 Two-piece normal and skew normal distributions

2.1 Two-piece normal distributions

According to Arellano-Valle et al. (2005), a continuous random variable Y follows a two-piece normal (TPN) distribution with location μ , scale σ and asymmetry parameter γ if its density function for any $y \in \mathbb{R}$ can be written as

$$h(y) = \frac{2}{\sigma(a(\gamma) + b(\gamma))} \left\{ \phi\left(\frac{y - \mu}{\sigma a(\gamma)}\right) I_{[\mu, \infty)}(y) + \phi\left(\frac{y - \mu}{\sigma b(\gamma)}\right) I_{(-\infty, \mu)}(y) \right\}, \quad (1)$$

where $\phi(x)$ denotes the density of a standard Gaussian and $I_A(x)$ denotes the indicator function of the set A ; we write such a distribution compactly as $Y \sim TPN(\mu, \sigma, \gamma)$. Skewness is controlled via two functions $a(\gamma)$ and $b(\gamma)$ satisfying the following properties:

- (i) $a(\gamma)$ and $b(\gamma)$ are positive-valued functions for $\gamma \in (\gamma_L, \gamma_U)$, a possibly infinite interval;
- (ii) one of the functions is (strictly) increasing and the other is (strictly) decreasing;
- (iii) there exists a unique value $\gamma_* \in (\gamma_L, \gamma_U)$ such that $a(\gamma_*) = b(\gamma_*)$, and therefore the TPN density (1) becomes

$$h(y) = \frac{2}{\sigma a(\gamma_*)} \phi\left(\frac{y - \mu}{\sigma a(\gamma_*)}\right).$$

In addition, the TPN distribution has several interesting formal properties in terms of stochastic construction (Arellano-Valle et al. 2005). The following list includes those most relevant for the purposes of this work:

P1. The TPN density (1) can be expressed as a finite mixture of two truncated normal densities f_a and f_b given by

$$f_a(y) = \frac{2}{\sigma a(\gamma)} \phi\left(\frac{y - \mu}{\sigma a(\gamma)}\right) I_{[\mu, \infty)}(y), \quad f_b(y) = \frac{2}{\sigma b(\gamma)} \phi\left(\frac{y - \mu}{\sigma b(\gamma)}\right) I_{(-\infty, \mu)}(y).$$

That is,

$$h(y) = \pi_a f_a(y) + \pi_b f_b(y), \quad y \in \mathbb{R}, \quad (2)$$

where

$$\pi_s = \frac{s(\gamma)}{a(\gamma) + b(\gamma)}, \quad s = a, b. \quad (3)$$

P2. If $Y \sim h$, then $Y \stackrel{d}{=} \mu + \sigma W_\gamma V$, where the notation $\stackrel{d}{=}$ indicates equality in distribution. Specifically, $V \sim TN(0, 1, [0, \infty))$, a truncated normal with location 0, scale 1, truncated over the positive real line, while W_γ is an independent discrete random variable with probability function

$$p(w; \gamma) = \begin{cases} \pi_a & \text{if } w = a(\gamma), \\ \pi_b & \text{if } w = -b(\gamma), \\ 0 & \text{otherwise,} \end{cases}$$

which can be rewritten as

$$p(w; \gamma) = \pi_a^{(s+1)/2} \pi_b^{(s-1)/2} I_{\{-1,1\}}(s),$$

with $s = \text{sign}(w)$ and π_s defined by (3). Equivalently, if $Y = \mu + \sigma W_\gamma |X|$, where $X \sim N(0, 1)$ and is independent of W_γ , then $Y \sim h$. This stochastic representation allows one to obtain the mean and variance of Y leveraging the law of total expectation; refer to Arellano-Valle et al. (2020) for further details.

2.2 Skew-normal distribution

A random vector \mathbf{Y} has a multivariate Skew-Normal SN distribution with location vector $\boldsymbol{\xi}$, positive definite scale matrix $\boldsymbol{\Omega}$ and skewness/shape vector $\boldsymbol{\lambda}$, denoted by $\mathbf{Y} \sim SN_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\lambda})$, if its density function is given by

$$f(\mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\lambda}) = 2\phi_p(\mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\Omega}) \Phi(\boldsymbol{\lambda}^\top \boldsymbol{\Omega}^{-1/2}(\mathbf{y} - \boldsymbol{\xi})), \quad \mathbf{y} \in \mathbb{R}^p.$$

Here, $\phi_p(\cdot; \boldsymbol{\xi}, \boldsymbol{\Omega})$ denotes the density function of the p -variate normal distribution with mean vector $\boldsymbol{\xi}$ and variance-covariance matrix $\boldsymbol{\Omega}$, and $\Phi(\cdot)$ is the cumulative distribution function of a standard normal. The SN random vector $\mathbf{Y} \sim SN_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\lambda})$ can be introduced as the location-scale transformation $\mathbf{Y} = \boldsymbol{\xi} + \boldsymbol{\Omega}^{1/2} \mathbf{X}$, where \mathbf{X} has the following stochastic representation:

$$\mathbf{X} \stackrel{d}{=} \boldsymbol{\delta} |X_0| + \mathbf{X}_1, \quad (4)$$

where $\boldsymbol{\delta} = \boldsymbol{\lambda} / (1 + \boldsymbol{\lambda}^\top \boldsymbol{\lambda})^{1/2}$, $X_0 \sim N(0, 1)$ and $\mathbf{X}_1 \sim N_p(\mathbf{0}, \mathbf{I}_p - \boldsymbol{\delta} \boldsymbol{\delta}^\top)$, which are independent. By (4) we can get that if $\mathbf{Y} \sim SN_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\lambda})$, then there are two independent random quantities Z and \mathbf{U} , with $Z \stackrel{d}{=} |X_0|$ and $\mathbf{U} \stackrel{d}{=} \boldsymbol{\Omega}^{1/2} \mathbf{X}_1$, such that

$$\mathbf{Y} = \boldsymbol{\xi} + \boldsymbol{\Delta} Z + \mathbf{U}, \quad (5)$$

where $\boldsymbol{\Delta} = \boldsymbol{\Omega}^{1/2} \boldsymbol{\delta}$. Note that $Z \sim HN(0, 1)$ and $\mathbf{U} \sim (\mathbf{0}, \boldsymbol{\Omega} - \boldsymbol{\Delta} \boldsymbol{\Delta}^\top)$. Thus, using (5), it can be shown that the mean vector and variance-covariance matrix of \mathbf{Y} are given respectively by

$$E(\mathbf{Y}) = \boldsymbol{\xi} + \sqrt{\frac{2}{\pi}} \boldsymbol{\Delta} \quad \text{and} \quad \text{var}(\mathbf{Y}) = \boldsymbol{\Omega} - \frac{2}{\pi} \boldsymbol{\Delta} \boldsymbol{\Delta}^\top.$$

3 A two-piece normal dynamic linear model

3.1 The initial state distribution

Our proposal in this section is to derive a more flexible DLM that regulates asymmetry through a simple scalar parameter. Specifically, preserving the classical independence assumptions, we consider the DLM defined by

$$\mathbf{Y}_t = \mathbf{F}_t^\top \boldsymbol{\theta}_t + \mathbf{v}_t, \quad \mathbf{v}_t \sim N_r(\mathbf{0}, \mathbf{V}), \tag{6}$$

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N_p(\mathbf{0}, \mathbf{W}), \tag{7}$$

for $t = 1, \dots, T$, replacing the initial state parameter $\boldsymbol{\theta}_0$ distribution with the following hierarchical specification:

$$\boldsymbol{\theta}_0 | \varphi \sim N_p(\mathbf{m}_0 + \varphi \boldsymbol{\beta}_0, \mathbf{C}_0), \tag{8}$$

$$\varphi \sim TPN(\mu, \sigma_0, \gamma_0). \tag{9}$$

The model defined by the Eqs. (6, 7) and (8, 9) will be referred to as two-piece normal dynamic linear model (hereafter TPN-DLM).

As a first important result, we note that the hierarchical specification (8, 9) leads a mixture of two multivariate skew-normals as initial distribution for $\boldsymbol{\theta}_0$. Its proof can be derived as a direct extension of Proposition 2 in Arellano-Valle et al. (2020), and so it is omitted here.

Proposition 3.1 *Under the hierarchical representation defined by Eqs. (8, 9), the initial density of $\boldsymbol{\theta}_0$ is given by*

$$p(\boldsymbol{\theta}_0) = 2\pi_a \phi_p(\boldsymbol{\theta}_0; \boldsymbol{\xi}_0, \boldsymbol{\Omega}_a) \Phi(\boldsymbol{\eta}_a^\top (\boldsymbol{\theta}_0 - \boldsymbol{\xi}_0)) + 2\pi_b \phi_p(\boldsymbol{\theta}_0; \boldsymbol{\xi}_0, \boldsymbol{\Omega}_b) \Phi(\boldsymbol{\eta}_b^\top (\boldsymbol{\theta}_0 - \boldsymbol{\xi}_0)), \tag{10}$$

where, for $s = a, b$, π_s is defined by (3), and

$$\begin{aligned} \boldsymbol{\xi}_0 &= \mathbf{m}_0 + \mu \boldsymbol{\beta}_0, \quad \boldsymbol{\alpha}_s = \sigma_0^s (\gamma_0) \boldsymbol{\beta}_0, \\ \boldsymbol{\Omega}_s &= \mathbf{C}_0 + \boldsymbol{\alpha}_s \boldsymbol{\alpha}_s^\top, \quad \boldsymbol{\eta}_s = (1 - \boldsymbol{\alpha}_s^\top \boldsymbol{\Omega}_s^{-1} \boldsymbol{\alpha}_s)^{-1/2} \boldsymbol{\Omega}_s^{-1} \boldsymbol{\alpha}_s. \end{aligned}$$

Here it should be noted that from the well-known matrix inversion formula

$$\boldsymbol{\Omega}_s^{-1} = (\mathbf{C}_0 + \boldsymbol{\alpha}_s \boldsymbol{\alpha}_s^\top)^{-1} = \mathbf{C}_0^{-1} - \frac{\mathbf{C}_0^{-1} \boldsymbol{\alpha}_s \boldsymbol{\alpha}_s^\top \mathbf{C}_0^{-1}}{1 + \boldsymbol{\alpha}_s^\top \mathbf{C}_0^{-1} \boldsymbol{\alpha}_s},$$

we get, for $s = a, b$, that $1 - \boldsymbol{\alpha}_s^\top \boldsymbol{\Omega}_s^{-1} \boldsymbol{\alpha}_s = (1 + \boldsymbol{\alpha}_s^\top \mathbf{C}_0^{-1} \boldsymbol{\alpha}_s)^{-1} > 0$ and $\boldsymbol{\alpha}_s^\top \boldsymbol{\Omega}_s^{-1} = (1 + \boldsymbol{\alpha}_s^\top \mathbf{C}_0^{-1} \boldsymbol{\alpha}_s)^{-1} \boldsymbol{\alpha}_s^\top \mathbf{C}_0^{-1}$, so that the term $\boldsymbol{\eta}_s$ defined in Proposition 3.1 can be rewritten as

$$\boldsymbol{\eta}_s = (1 + \boldsymbol{\alpha}_s^\top \mathbf{C}_0^{-1} \boldsymbol{\alpha}_s)^{-1/2} \mathbf{C}_0^{-1} \boldsymbol{\alpha}_s.$$

The distribution of the initial random vector $\boldsymbol{\theta}_0$ can be written as

$$p(\boldsymbol{\theta}_0) = 2\pi_a \phi_p(\boldsymbol{\theta}_0; \mathbf{m}_0 + \mu \boldsymbol{\beta}_0, \mathbf{C}_0 + \boldsymbol{\alpha}_a \boldsymbol{\alpha}_a^\top) \Phi \left(\frac{\boldsymbol{\alpha}_a^\top \mathbf{C}_0^{-1} (\boldsymbol{\theta}_0 - \boldsymbol{\xi}_0)}{\sqrt{1 + \boldsymbol{\alpha}_a^\top \mathbf{C}_0^{-1} \boldsymbol{\alpha}_a}} \right) + 2\pi_b \phi_p(\boldsymbol{\theta}_0; \mathbf{m}_0 + \mu \boldsymbol{\beta}_0, \mathbf{C}_0 + \boldsymbol{\alpha}_b \boldsymbol{\alpha}_b^\top) \Phi \left(\frac{\boldsymbol{\alpha}_b^\top \mathbf{C}_0^{-1} (\boldsymbol{\theta}_0 - \boldsymbol{\xi}_0)}{\sqrt{1 + \boldsymbol{\alpha}_b^\top \mathbf{C}_0^{-1} \boldsymbol{\alpha}_b}} \right), \quad (11)$$

which correspond to the density of a two-component mixture of the multivariate skew-normal densities reported in Sect. 2.2. Specifically, from Proposition (3.1), we see that the initial state parameter is distributed as

$$\boldsymbol{\theta}_0 \sim \pi_a SN_p(\boldsymbol{\xi}_a, \boldsymbol{\Omega}_a, \boldsymbol{\lambda}_a) + \pi_b SN_p(\boldsymbol{\xi}_b, \boldsymbol{\Omega}_b, \boldsymbol{\lambda}_b), \quad (12)$$

where

$$\boldsymbol{\xi}_s = \boldsymbol{\xi}_0 \quad \text{and} \quad \boldsymbol{\lambda}_s = (1 - \boldsymbol{\alpha}_s^\top \boldsymbol{\Omega}_s^{-1} \boldsymbol{\alpha}_s)^{-1/2} \boldsymbol{\Omega}_s^{-1/2} \boldsymbol{\alpha}_s, \quad s = a, b.$$

3.2 The Kalman filter

Our next step is to develop a Kalman filter based on the new initial distribution given by (12), and assuming that the conditional distribution of $\boldsymbol{\varphi}$ corresponds to a mixture of two truncated Gaussian densities.

Let $D_t = \{\mathbf{y}_1, \dots, \mathbf{y}_t\}$ denote the available information at time t , where \mathbf{y}_i indicates a realization of the random variable \mathbf{Y}_i . In the proposed TPN-DLM we consider a conditionally normal distribution for $\boldsymbol{\theta}_0$ given $\boldsymbol{\varphi}$, with a TPN initial distribution for $\boldsymbol{\varphi}$ (8). Furthermore, we assume by induction that

$$\boldsymbol{\theta}_{t-1} | \boldsymbol{\varphi}, D_{t-1} \sim N_p(\mathbf{m}_{t-1} + \boldsymbol{\varphi} \boldsymbol{\beta}_{t-1}, \mathbf{C}_{t-1}),$$

$$\boldsymbol{\varphi} | D_{t-1} \sim \pi_{t-1}^a TN(\eta_{t-1}^a, \tau_{t-1}^a, [\mu, \infty)) + \pi_{t-1}^b TN(\eta_{t-1}^b, \tau_{t-1}^b, (-\infty, \mu)). \quad (13)$$

Specifically, the conditional distribution of $\boldsymbol{\varphi}$ correspond to a mixture of two truncated Gaussian with locations η_{t-1}^s , scales τ_{t-1}^s and mixing weights π_{t-1}^s with $s = a, b$ and truncation point μ , defined by the initial distribution given in 9.

Leveraging the conditional independence properties of the TPN-DLM outlined in Eq. 7, the *one-step-ahead predictive distribution* of $\boldsymbol{\theta}_t$ given $(\boldsymbol{\varphi}, D_{t-1})$ is given by

$$\boldsymbol{\theta}_t | (\boldsymbol{\varphi}, D_{t-1}) \stackrel{d}{=} \mathbf{G}_t(\boldsymbol{\theta}_{t-1} | (\boldsymbol{\varphi}, D_{t-1})) + \boldsymbol{\omega}_t \sim N_p(\mathbf{a}_t + \boldsymbol{\varphi} \mathbf{b}_t, \mathbf{R}_t), \quad (14)$$

where

$$\mathbf{a}_t = \mathbf{G}_t \mathbf{m}_{t-1}, \quad \mathbf{b}_t = \mathbf{G}_t \boldsymbol{\beta}_{t-1}, \quad \mathbf{R}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t^\top + \mathbf{W}. \tag{15}$$

Similarly, using (6) we find that the one-step-ahead predictive distribution of \mathbf{Y}_t given (φ, D_{t-1}) becomes

$$\mathbf{Y}_t | (\varphi, D_{t-1}) \stackrel{d}{=} \mathbf{F}_t^\top (\boldsymbol{\theta}_t | (\varphi, D_{t-1})) + \mathbf{v}_t \sim N_r(\mathbf{F}_t^\top \mathbf{a}_t + \varphi \mathbf{F}_t^\top \mathbf{b}_t, \boldsymbol{\Sigma}_t), \tag{16}$$

where

$$\boldsymbol{\Sigma}_t = \mathbf{F}_t^\top \mathbf{R}_t \mathbf{F}_t + \mathbf{V}. \tag{17}$$

In other words, from (14) and (16) we have that

$$\begin{bmatrix} \mathbf{Y}_t \\ \boldsymbol{\theta}_t \end{bmatrix} | (\varphi, D_{t-1}) \stackrel{d}{=} \begin{bmatrix} \mathbf{F}_t^\top \mathbf{G}_t \\ \mathbf{G}_t \end{bmatrix} (\boldsymbol{\theta}_{t-1} | (\varphi, D_{t-1})) + \begin{bmatrix} \mathbf{F}_t^\top & \mathbf{I}_r \\ \mathbf{I}_p & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\omega}_t \\ \mathbf{v}_t \end{bmatrix},$$

and therefore

$$\begin{bmatrix} \mathbf{Y}_t \\ \boldsymbol{\theta}_t \end{bmatrix} | (\varphi, D_{t-1}) \sim N_{p+r} \left(\begin{bmatrix} \mathbf{F}_t^\top \mathbf{a}_t \\ \mathbf{a}_t \end{bmatrix} + \varphi \begin{bmatrix} \mathbf{F}_t^\top \mathbf{b}_t \\ \mathbf{b}_t \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_t & \mathbf{F}_t^\top \mathbf{R}_t \\ \mathbf{R}_t \mathbf{F}_t & \mathbf{R}_t \end{bmatrix} \right).$$

Finally, by applying the properties of the conditional normal distribution, we obtain the following *filtering distribution* of $\boldsymbol{\theta}_t$ given (φ, D_t) :

$$\boldsymbol{\theta}_t | (\varphi, D_t) \sim N_p(\mathbf{m}_t + \varphi \boldsymbol{\beta}_t, \mathbf{C}_t), \tag{18}$$

where

$$\left. \begin{aligned} \mathbf{m}_t &= \mathbf{a}_t + \mathbf{R}_t \mathbf{F}_t \boldsymbol{\Sigma}_t^{-1} (\mathbf{y}_t - \mathbf{F}_t^\top \mathbf{a}_t), \\ \boldsymbol{\beta}_t &= \mathbf{b}_t - \mathbf{R}_t \mathbf{F}_t \boldsymbol{\Sigma}_t^{-1} \mathbf{F}_t^\top \mathbf{b}_t, \\ \mathbf{C}_t &= \mathbf{R}_t - \mathbf{R}_t \mathbf{F}_t \boldsymbol{\Sigma}_t^{-1} \mathbf{F}_t^\top \mathbf{R}_t, \end{aligned} \right\} \tag{19}$$

with \mathbf{a}_t , \mathbf{b}_t , \mathbf{R}_t and $\boldsymbol{\Sigma}_t$ defined as in (15) and (17), respectively.

The above results are formalized below:

Proposition 3.2 Consider the TPN-DLM defined by Eqs. (6)-(7) and (8)-(9), with the induction assumptions (13). Then:

- (i) The one-step-ahead conditional predictive distribution of

$$\boldsymbol{\theta}_t | (\varphi, D_{t-1}) \sim N_p(\mathbf{a}_t + \varphi \mathbf{b}_t, \mathbf{R}_t);$$

- (ii) The one-step-ahead conditional predictive distribution of

$$\mathbf{Y}_t | (\varphi, D_{t-1}) \sim N_r(\mathbf{F}_t^\top \mathbf{a}_t + \varphi \mathbf{F}_t^\top \mathbf{b}_t, \boldsymbol{\Sigma}_t);$$

(iii) The conditional filtering distribution of

$$\theta_t | (\varphi, D_t) \sim N_p(\mathbf{m}_t + \varphi \boldsymbol{\beta}_t, \mathbf{C}_t).$$

The next proposition establishes the conditional distribution of $\varphi | D_t$.

Proposition 3.3 Consider the TPN-DLM defined by Eqs. (6)-(7) and (8)-(9), with the induction assumptions (13). Then the conditional distribution of $\varphi | D_t$ has a finite mixture density of two truncated Gaussian distributions, given by

$$\varphi | D_t \sim \pi_t^a TN(\eta_t^a, \tau_t^a, [\mu, \infty)) + \pi_t^b TN(\eta_t^b, \tau_t^b, (-\infty, \mu)) \quad (20)$$

where, for $s = a, b$,

$$\eta_t^s = \frac{\eta_{t-1}^s + \tau_{t-1}^s \mathbf{b}_t^\top \mathbf{F}_t \boldsymbol{\Sigma}_t^{-1} (\mathbf{y}_t - \mathbf{F}_t^\top \mathbf{a}_t)}{1 + \tau_{t-1}^s \mathbf{b}_t^\top \mathbf{F}_t \boldsymbol{\Sigma}_t^{-1} \mathbf{F}_t^\top \mathbf{b}_t}, \quad \tau_t^s = \frac{\tau_{t-1}^s}{1 + \tau_{t-1}^s \mathbf{b}_t^\top \mathbf{F}_t \boldsymbol{\Sigma}_t^{-1} \mathbf{F}_t^\top \mathbf{b}_t},$$

and

$$\pi_t^a = c_t \pi_{t-1}^a \phi_r(\mathbf{y}_t; \mathbf{F}_t^\top (\mathbf{a}_t + \eta_{t-1}^a \mathbf{b}_t), \boldsymbol{\Sigma}_t + \tau_{t-1}^a \mathbf{F}_t^\top \mathbf{b}_t \mathbf{b}_t^\top \mathbf{F}_t) \Phi\left(-\frac{\mu - \eta_t^a}{\sqrt{\tau_t^a}}\right),$$

$$\pi_t^b = c_t \pi_{t-1}^b \phi_r(\mathbf{y}_t; \mathbf{F}_t^\top (\mathbf{a}_t + \eta_{t-1}^b \mathbf{b}_t), \boldsymbol{\Sigma}_t + \tau_{t-1}^b \mathbf{F}_t^\top \mathbf{b}_t \mathbf{b}_t^\top \mathbf{F}_t) \Phi\left(\frac{\mu - \eta_t^b}{\sqrt{\tau_t^b}}\right),$$

where

$$c_t^{-1} = \pi_{t-1}^a \phi_r(\mathbf{y}_t; \mathbf{F}_t^\top (\mathbf{a}_t + \eta_{t-1}^a \mathbf{b}_t), \boldsymbol{\Sigma}_t + \tau_{t-1}^a \mathbf{F}_t^\top \mathbf{b}_t \mathbf{b}_t^\top \mathbf{F}_t) \Phi\left(-\frac{\mu - \eta_t^a}{\sqrt{\tau_t^a}}\right)$$

$$+ \pi_{t-1}^b \phi_r(\mathbf{y}_t; \mathbf{F}_t^\top (\mathbf{a}_t + \eta_{t-1}^b \mathbf{b}_t), \boldsymbol{\Sigma}_t + \tau_{t-1}^b \mathbf{F}_t^\top \mathbf{b}_t \mathbf{b}_t^\top \mathbf{F}_t) \Phi\left(\frac{\mu - \eta_t^b}{\sqrt{\tau_t^b}}\right)$$

This representation allows to characterize the expected value and variance of $\varphi | D_t$, which can be expressed as

$$E\{\varphi | D_t\} = \pi_t^a \left[\eta_t^a + \sqrt{\tau_t^a} \frac{\phi\left(\frac{\mu - \eta_t^a}{\sqrt{\tau_t^a}}\right)}{\Phi\left(-\frac{\mu - \eta_t^a}{\sqrt{\tau_t^a}}\right)} \right] + \pi_t^b \left[\eta_t^b - \sqrt{\tau_t^b} \frac{\phi\left(\frac{\mu - \eta_t^b}{\sqrt{\tau_t^b}}\right)}{\Phi\left(\frac{\mu - \eta_t^b}{\sqrt{\tau_t^b}}\right)} \right] \quad (21)$$

and

$$\begin{aligned} \text{Var}\{\varphi \mid D_t\} &= \pi_t^a \left[(\eta_t^a)^2 + \tau_t^a + \sqrt{\tau_t^a}(\mu + \eta_t^a) \frac{\phi\left(\frac{\mu - \eta_t^a}{\sqrt{\tau_t^a}}\right)}{\Phi\left(-\frac{\mu - \eta_t^a}{\sqrt{\tau_t^a}}\right)} \right] \\ &+ \pi_t^b \left[(\eta_t^b)^2 + \tau_t^b - \sqrt{\tau_t^b}(\mu + \eta_t^b) \frac{\phi\left(\frac{\mu - \eta_t^b}{\sqrt{\tau_t^b}}\right)}{\Phi\left(\frac{\mu - \eta_t^b}{\sqrt{\tau_t^b}}\right)} \right] \\ &- [E\{\varphi \mid D_t\}]^2. \end{aligned} \tag{22}$$

A simpler expression for $\text{Var}\{\varphi \mid D_t\}$ can be obtained in terms of Chi-square cumulative distribution function, adapting (Barr and Sherrill 1999).

Immediate consequences of these results are given in the following proposition.

Proposition 3.4 Consider the TPN-DLM defined by Eqs. (6)-(7) and (8)-(9), with the induction assumptions (13). Then:

- (i) The one-step-ahead predictive distribution of θ_t given D_{t-1} is

$$\begin{aligned} p(\theta_t \mid D_{t-1}) &= \pi_{t-1}^a \phi_p(\theta_t; \mathbf{a}_t + \eta_{t-1}^a \mathbf{b}_t, \mathbf{R}_t + \tau_{t-1}^a \mathbf{b}_t \mathbf{b}_t^\top) \Phi\left(-\frac{\mu - \chi_t^a}{\sqrt{\vartheta_t^a}}\right) \\ &+ \pi_{t-1}^b \phi_p(\theta_t; \mathbf{a}_t + \eta_{t-1}^b \mathbf{b}_t, \mathbf{R}_t + \tau_{t-1}^b \mathbf{b}_t \mathbf{b}_t^\top) \Phi\left(\frac{\mu - \chi_t^b}{\sqrt{\vartheta_t^b}}\right), \end{aligned}$$

where, for $s = a, b$,

$$\chi_t^s = \frac{\eta_{t-1}^s + \tau_{t-1}^s \mathbf{b}_t^\top \mathbf{R}_t^{-1} (\theta_t - \mathbf{a}_t)}{1 + \tau_{t-1}^s \mathbf{b}_t^\top \mathbf{R}_t^{-1} \mathbf{b}_t}, \quad \vartheta_t^s = \frac{\tau_{t-1}^s}{1 + \tau_{t-1}^s \mathbf{b}_t^\top \mathbf{R}_t^{-1} \mathbf{b}_t}.$$

- (ii) The one-step-ahead predictive distribution of \mathbf{y}_t given D_{t-1} is

$$\begin{aligned} p(\mathbf{y}_t \mid D_{t-1}) &= \pi_{t-1}^a \phi_r(\mathbf{y}_t; \mathbf{F}_t^\top (\mathbf{a}_t + \eta_{t-1}^a \mathbf{b}_t), \Sigma_t + \tau_{t-1}^a \mathbf{F}_t^\top \mathbf{b}_t \mathbf{b}_t^\top \mathbf{F}_t) \Phi\left(-\frac{\mu - \eta_t^a}{\sqrt{\tau_t^a}}\right) \\ &+ \pi_{t-1}^b \phi_r(\mathbf{y}_t; \mathbf{F}_t^\top (\mathbf{a}_t + \eta_{t-1}^b \mathbf{b}_t), \Sigma_t + \tau_{t-1}^b \mathbf{F}_t^\top \mathbf{b}_t \mathbf{b}_t^\top \mathbf{F}_t) \Phi\left(\frac{\mu - \eta_t^b}{\sqrt{\tau_t^b}}\right) \end{aligned}$$

where η_t^s and τ_t^s , for $s = a, b$, are defined above in Proposition 3.3.

(iii) The filtering distribution is

$$\begin{aligned}
 p(\boldsymbol{\theta}_t | D_t) = & \pi_t^a \phi_p(\boldsymbol{\theta}_t; \mathbf{m}_t + \eta_t^a \boldsymbol{\beta}_t, \mathbf{C}_t + \tau_t^a \boldsymbol{\beta}_t \boldsymbol{\beta}_t^\top) \frac{\Phi\left(-\frac{\mu - \delta_t^a}{\sqrt{v_t^a}}\right)}{\Phi\left(-\frac{\mu - \eta_t^a}{\sqrt{\tau_t^a}}\right)} \\
 & + \pi_t^b \phi_p(\boldsymbol{\theta}_t; \mathbf{m}_t + \eta_t^b \boldsymbol{\beta}_t, \mathbf{C}_t + \tau_t^b \boldsymbol{\beta}_t \boldsymbol{\beta}_t^\top) \frac{\Phi\left(\frac{\mu - \delta_t^b}{\sqrt{v_t^b}}\right)}{\Phi\left(\frac{\mu - \eta_t^b}{\sqrt{\tau_t^b}}\right)}, \tag{23}
 \end{aligned}$$

where π_t^s, η_t^s and τ_t^s , for $s = a, b$, are defined in in Proposition 3.3, and

$$\delta_t^s = \frac{\eta_t^s + \tau_t^s \boldsymbol{\beta}_t^\top \mathbf{C}_t^{-1} (\boldsymbol{\theta}_t - \mathbf{m}_t)}{1 + \tau_t^s \boldsymbol{\beta}_t^\top \mathbf{C}_t^{-1} \boldsymbol{\beta}_t}, \quad v_t^s = \frac{\tau_t^s}{1 + \tau_t^s \boldsymbol{\beta}_t^\top \mathbf{C}_t^{-1} \boldsymbol{\beta}_t}.$$

Proposition 3.4 shows the distribution of one-step-ahead predictive distribution of the states is typically skewed, and the same is true for the analogous predictive distribution of the response. Also, Proportion 3.4 shows that the filtering distribution is also typically skewed. This can be seen by comparing the results of our Proposition 3.4 with the results from the usual DLM (see, e.g., Petris et al. (2009)). Finally, since

$$\begin{aligned}
 E(\boldsymbol{\theta}_t | D_{t-1}) &= E\{E(\boldsymbol{\theta}_t | \varphi, D_{t-1})\}, \\
 E(\mathbf{Y}_t | D_{t-1}) &= E\{E(\mathbf{Y}_t | \varphi, D_{t-1})\},
 \end{aligned}$$

and

$$\begin{aligned}
 \text{Var}(\boldsymbol{\theta}_t | D_{t-1}) &= E\{\text{Var}(\boldsymbol{\theta}_t | \varphi, D_{t-1})\} + \text{Var}\{E(\boldsymbol{\theta}_t | \varphi, D_{t-1})\}, \\
 \text{Var}(\mathbf{Y}_t | D_{t-1}) &= E\{\text{Var}(\mathbf{Y}_t | \varphi, D_{t-1})\} + \text{Var}\{E(\mathbf{Y}_t | \varphi, D_{t-1})\},
 \end{aligned}$$

then from equations (14), (16) and the property P3. of the TPN distribution (see Sect. 2.1), we obtain the following results:

Proposition 3.5 *Under the TPN-DLM defined by Eqs. (6)-(7) and (8)-(9), with the induction assumptions (13), the one-step-ahead expected filtering and prediction distributions and their covariance matrices are, respectively, given by*

$$\begin{aligned}
 E(\boldsymbol{\theta}_t | D_{t-1}) &= \mathbf{a}_t + E\{\varphi | D_t\} \mathbf{b}_t, \\
 \text{Var}(\boldsymbol{\theta}_t | D_{t-1}) &= \mathbf{R}_t + \text{Var}\{\varphi | D_t\} \mathbf{b}_t \mathbf{b}_t^\top, \\
 E(\mathbf{Y}_t | D_{t-1}) &= \mathbf{F}_t^\top \mathbf{a}_t + E\{\varphi | D_t\} \mathbf{F}_t^\top \mathbf{b}_t, \\
 \text{Var}(\mathbf{Y}_t | D_{t-1}) &= \boldsymbol{\Sigma}_t + \text{Var}\{\varphi | D_t\} \mathbf{F}_t^\top \mathbf{b}_t \mathbf{b}_t^\top \mathbf{F}_t,
 \end{aligned}$$

where $E\{\varphi | D_t\}$ and $\text{Var}\{\varphi | D_t\}$ are derived in Eqs. 21 and 22, respectively.

4 Outline of Bayesian computation

In this section we combine the results obtained above to derive a *forward filtering backward sampling* (FFBS) to conduct full Bayesian inference on the model’s parameters $\Theta = (\theta_1, \dots, \theta_T)$, V and W via Markov-Chain Monte Carlo (MCMC). In particular, we assign Inverse-Wishart priors on the error covariances V and W as

$$V \sim IW_r(\ell, M) \quad W \sim IW_p(g, Z),$$

where M and Z are positive definite matrix with size $r \times r$ and $p \times p$, respectively, while ℓ and g are scalars such that $\ell > (r - 1)/2$ and $g > (p - 1)/2$. This choice guarantees that the covariance matrices V and W are positive definite.

Conditionally on the latent states, such distributions are conjugate, as the model is conditionally Gaussian. Therefore, the full conditional distributions of V and W are again Inverse-Wishart with

$$V|(D_T, \theta) \sim IW_r\left(\ell + \frac{T}{2}, \frac{1}{2} \sum_{t=1}^T (y_t - F_t^\top \theta_t)(y_t - F_t^\top \theta_t)^\top + M\right), \tag{24}$$

$$W|(D_T, \theta) \sim IW_p\left(g + \frac{T}{2}, \frac{1}{2} \sum_{t=1}^T (\theta_t - G_t \theta_{t-1})(\theta_t - G_t \theta_{t-1})^\top + Z\right). \tag{25}$$

In order to sample from $\Theta|(D_T, \varphi)$ we rely on backward recursions and decompose the filtered distribution for the state parameters following Carter and Kohn (1994), Frühwirth-Schnatter (1994), as

$$p(\Theta|D_T, \varphi) = \prod_{t=0}^T p(\theta_t | \theta_{t+1}, D_t, \varphi) \tag{26}$$

$$= p(\theta_T | D_T, \varphi) \prod_{t=0}^{T-1} p(\theta_t | \theta_{t+1}, D_t, \varphi), \tag{27}$$

where

$$\theta_t | (\theta_{t+1}, D_t, \varphi) \sim N(\mathbf{h}_t, \mathbf{H}_t) \tag{28}$$

and with

$$\begin{aligned} \mathbf{h}_t &= \mathbf{m}_t + \varphi \beta_t + C_t G_{t+1}^\top R_{t+1}^{-1} (\theta_{t+1} - \mathbf{a}_{t+1} - \varphi \mathbf{b}_{t+1}) \\ \mathbf{H}_t &= C_t - C_t G_{t+1}^\top R_{t+1}^{-1} G_{t+1} C_t. \end{aligned}$$

4.1 MCMC algorithm

Posterior sampling can be performed combining the above results in a MCMC algorithm, alternating the Kalman filter with sampling from the conditional distributions. The following pseudo-code illustrates the steps of a single MCMC iteration:

1. Sample Θ using the following modified FFBS algorithm:
 - 1a. For $t = 1, \dots, T$, update the parameters of the distribution $\theta_t|D_t$ using the Kalman filter given in Sect. 3.2 (*forward filtering*)
 - 1b. For $t = 1, \dots, T$, sample $\varphi|D_t$ from the conditional distribution outlined in Eq. 20;
 - 1c. Sample $\theta_T|D_T$ from the filtering distribution reported in Eq. 23;
 - 1d. For $t = T - 1, T - 2, \dots, 1$, sample $\theta_t|(\theta_{t+1}, D_t, \varphi)$ from the distribution outlined in Eq. 28, conditioning on the θ_{t+1} sampled in the previous step (*backward smoothing*)
2. Sample V from its Inverse-Wishart full-conditional distribution, outlined in Equation (24);
3. Sample W from its Inverse-Wishart full-conditional distribution, outlined in Equation (25);

5 Simulation

We propose a simulation study to compare the performance of the proposed approach against a Gaussian DLM, focusing on different settings with varying sample size. We focus on univariate settings, assuming that the matrices $\{G_t\}$ and $\{F_t\}$ are unidimensional and do not depend on time, namely $F = G = 1$. We simulated $T = 50$ observations from the DLM defined by

$$\begin{aligned} Y_t &= \theta_t + v_t, \\ \theta_t &= \theta_{t-1} + \omega_t, \end{aligned}$$

with different specification of the initial distribution θ_0 and the disturbances v_t and ω_t . Specifically, we focus on the following settings:

- (1) Scenario 1: data are generated from a two-piece DLM, with initial distribution $\theta_0|\varphi \sim N(-3 + 2\varphi, 2)$ and $\varphi \sim TPN(3, \sqrt{3}, 0.5)$, letting $a(\gamma) = 1 + \gamma$ and $b(\gamma) = 1 - \gamma$, with Gaussian errors $v_t \sim N(0, 5)$ and $\omega_t \sim N(0, 3)$
- (2) Scenario 2: data are generated from a Gaussian DLM, with $\theta_0 \sim N(-3, 2)$ and Gaussian errors $v_t \sim N(0, 5)$, $\omega_t \sim N(0, 3)$
- (3) Scenario 3: data are generated from a DLM with heavy tails, simulating θ_0 , v_t and ω_t from independent Student's t distribution with 3 degrees of freedom.

We chose diffuse inverse-gamma distributions as priors for V and W , which in this case are scalar, with parameters $\ell = M = g = Z = 0.001$. We compare our approach with a Gaussian DLM with same prior distributions, running both algorithms for 5000 iterations after 500 burn-in samples, and focusing on the one-step-ahead predictions and state parameters. Examination of traceplots of the parameters, auto-correlation function and Rubin’s diagnostics showed no evidence against convergence.

Figure 1, 2, and 3 show the one-step-ahead predictions and filtered estimates in the three scenarios. Current empirical findings indicate that, as expected, the main advantage of the proposed approach is more evident in the initial part of the series, where the impact of the initial distribution is substantial. This result is clearly seen in Fig. 1, where the TPN-DLM is correctly specified, and the Gaussian DLM tend to underestimate the state parameter and the one-step-ahead predictions. When data are generated from a Gaussian DLM, as in Fig. 2, the TPN initial distribution is incorrectly specified. However, its impact vanishes after few steps, and its one-step-ahead predictions are indistinguishable from a Gaussian DLM. Lastly, Fig. 3 focuses on a setting where both models are incorrectly specified, in terms of initial and distribution of the errors. We observe that the proposed TPN-DLM is robust again such misspecification, obtaining one-step-ahead predictions and estimates for the state parameters that are closer to the true level.

These findings are further explored replicating the simulations scenarios for different sample sizes, ranging $T \in \{10, 50, 100\}$, with $T = 50$ corresponding to the results from Figs. 1,2 and 3. Results are reported in Table 1, comparing the Mean Squared Error (MSE) of the expected value of the one-step-ahead distributions under both approaches. Empirical results are consistent with the previous

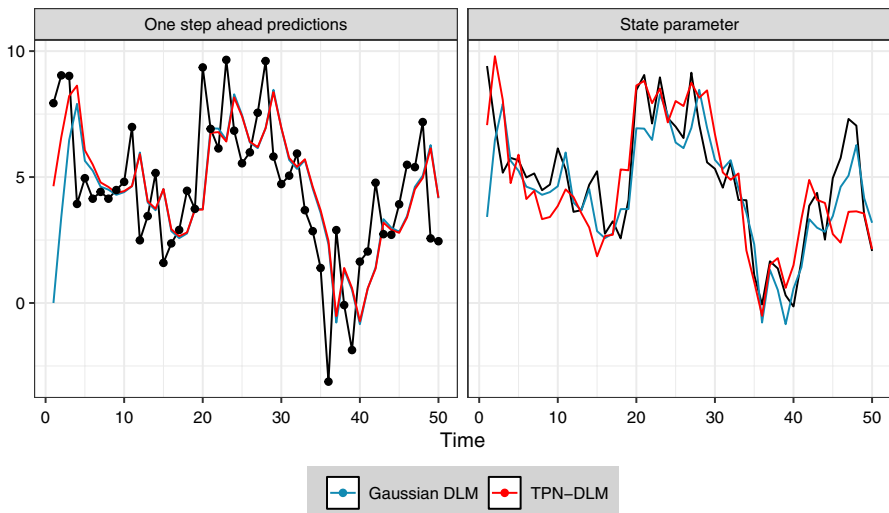


Fig. 1 One-step-ahead predictions and filtered estimates for Scenario 1. Black lines denote the observed time series and the true state parameters

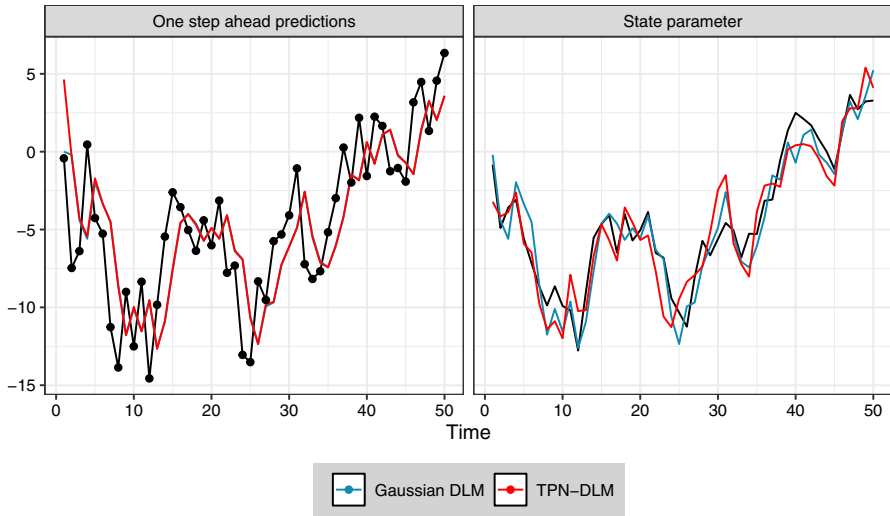


Fig. 2 One-step-ahead predictions and filtered estimates for Scenario 2. Black lines denote the observed time series and the true state parameters

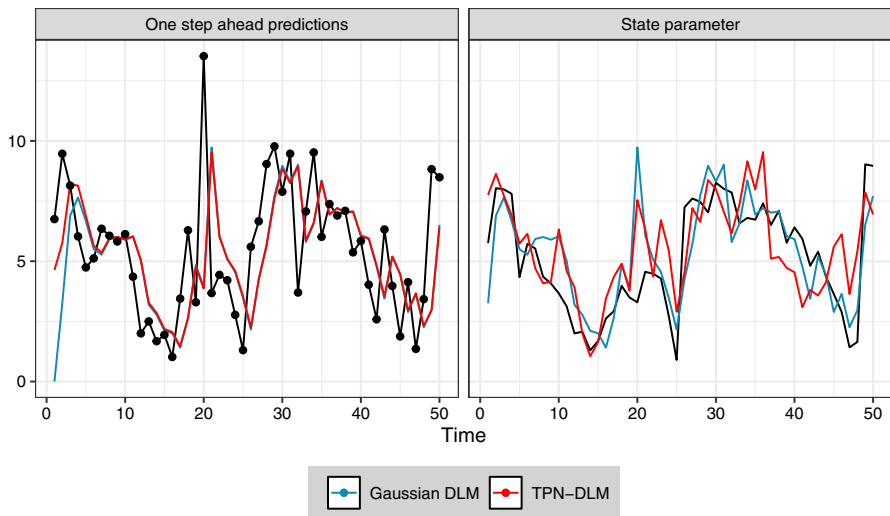


Fig. 3 One-step-ahead predictions and filtered estimates for Scenario 3. Black lines denote the observed time series and the true state parameters

discussion, with the TPN-DLM performing particularly well with small sample sizes, under correct specification and with heavy-tails processes.

Table 1 Mean squared error for the expected value of the one-step-ahead distribution

	$T = 10$		$T = 50$		$T = 100$	
	Gaussian	TPN	Gaussian	TPN	Gaussian	TPN
Scenario 1	118.74	43.01	327.60	251.34	599.97	524.32
Scenario 2	183.86	206.06	560.51	580.75	996.08	1016.96
Scenario 3	93.45	28.62	418.34	351.09	806.81	742.56

6 Analysis of real data

Finally, we illustrate the TPN-DLM by analyzing the quarterly earnings in dollars per Johnson and Johnson share from 1960 to 1980 (Shumway et al. 2000, Example 1.1).

Data are characterized by a seasonality larger in the starting and ending years, almost missing for the central years. Trend is increasing and regular. Following Shumway et al. (2000), the time series will be modelled with the trend and seasonality components added to a white noise

$$Y_t = T_t + S_t + v_t;$$

trend will be modelled as follow

$$T_t = \phi T_{t-1} + \omega_{t1}$$

and we assume that the seasonal component is expected to sum to zero over a complete period of four quarters

$$S_t + S_{t-1} + S_{t-2} + S_{t-3} = \omega_{t2}.$$

We may express the model in state-space form, by choosing $[T_t, S_t, S_{t-1}, S_{t-2}]^T$ as state vector:

$$Y_t = [1 \ 1 \ 0 \ 0] \begin{bmatrix} T_t \\ S_t \\ S_{t-1} \\ S_{t-2} \end{bmatrix} + v_t \quad \text{where} \quad v_t \sim N(0, V)$$

$$\begin{bmatrix} T_t \\ S_t \\ S_{t-1} \\ S_{t-2} \end{bmatrix} = \begin{bmatrix} \phi & 0 & 0 & 0 \\ 0 & -1 & -1 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} T_{t-1} \\ S_{t-1} \\ S_{t-2} \\ S_{t-3} \end{bmatrix} + \begin{bmatrix} \omega_{t1} \\ \omega_{t2} \\ 0 \\ 0 \end{bmatrix}$$

$$\text{where} \quad \begin{bmatrix} \omega_{t1} \\ \omega_{t2} \\ 0 \\ 0 \end{bmatrix} \sim N_4 \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} W_{11} & 0 & 0 & 0 \\ 0 & W_{22} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \right)$$

The parameters to be estimated are the observation noise variance, V , and the state noise variances associated with the trend, W_{11} , and the seasonal components, W_{22} .

In addition, we need to estimate the transition parameter associated with the growth rate, ϕ . Following Shumway et al. (2000), Example 6.27, we write $\phi = 1 + \zeta$, where $0 < \zeta \leq 1$, and we rewrite the trend component as

$$T_t - T_{t-1} = \zeta T_{t-1} + \omega_{t1},$$

so that, conditionally on the states, ζ is the slope of the linear regression of $(T_t - T_{t-1})$ on T_{t-1} and ω_{t1} is the error. We choose a reference uninformative prior on (ζ, ω_{t1}) and weakly informative priors for the remaining parameters by letting $\ell = M = 0.001$ and $g = 0.05$ and $\mathbf{Z} = \text{diag}\{0.05, 0.05, 0, 0\}$. We ran the algorithm for 5000 iterations collected after 5000 burn-in samples. Examination of traceplots of the parameters, auto-correlation function and Rubin's diagnostics showed no evidence against convergence. Fig. 4 displays the comparison of the trends (T_t) and season ($T_t + S_t$) along with 99% credible intervals for Gaussian DLM and TPN-DLM. Figure 5 displays the data and the one-step-ahead predictions for the time series Y_t , again along with 99% credible intervals for Gaussian DLM and two-piece DLM.

Figures 4 and 5 show that the 99% credibility intervals of state and response are different between of the two-piece DLM and DLM: as a consequence the entire distributions of those quantities are different. In addition we note that skewness of predictive distributions is maintained with the increasing of time, showing the usefulness of the two-piece DLM.

Mean squared error was 0.2131 for the two-piece DLM and 0.3512 for the DLM, showing an advantage towards the two-piece DLM. We also considered a BIC criteria for competing alternative models k , $k = 1, 2, \dots, K$, the smaller-is-better criterion BIC is $\text{BIC}_k = n \log \text{MSE}_k + m_k \log(n)$, where MSE_k is the predicting mean squared error and m_k is the number of independent parameters used to fit model k . We obtained -70.18 for the DLM, where $m_k = 4$ and -107.69 for the two-piece DLM, where $m_k = 5$, confirming that also including complexity of the model, two-piece DLM is preferable.

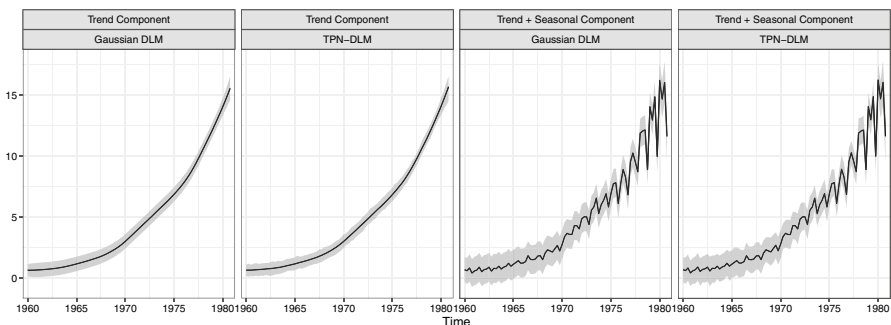


Fig. 4 Posterior estimate of trend (T_t) and trend plus season ($T_t + S_t$) along with corresponding 99% credible intervals for the Johnson & Johnson data

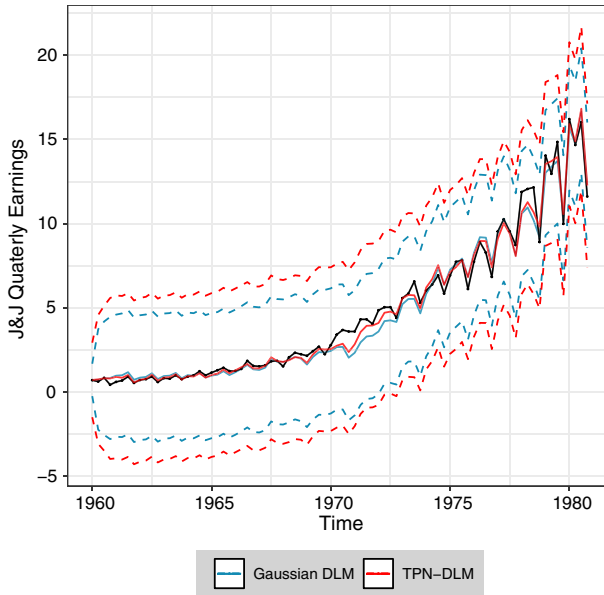


Fig. 5 One-step-ahead predictions for the Johnson & Johnson quarterly earnings series. Dotted lines refer to 99% credible intervals

7 Conclusion

In this article we proposed a flexible dynamic linear model (DLM) for modeling and forecasting multivariate time series relaxing the assumption of normality for the initial distribution of the state space parameter, replacing it by a more flexible class of distributions, which is called two-piece normal distributions. This model allows the initial distribution of the state space parameter to be skewed, and the asymmetry can be controlled by a scalar parameter. We derive a Kalman filter for this model, obtaining a two component mixture as predictive and filtering distributions that maintain skeweness.

In our opinion, the main contribution of this article is to present a simple and effective tool to model time series with possibly skewed distribution, like the Example 1.1 in Shumway et al. (2000) here. Also since we obtained a two component mixture for predictive and filtering distributions so this new model can simultaneously deal with some issues related to departures from normality like skewed, heavy-tailed data and also, multi-modality.

Appendix: Proofs

Proof of Proposition 3.1 Using the hierarchical representation defined by Eqs. (8)–(9), and the fact that φ has a TPN density (2), we obtain

$$\begin{aligned}
p(\theta_0) &= \int_{-\infty}^{\infty} \phi_p(\theta_0; \mathbf{m}_0 + \varphi \boldsymbol{\beta}_0, \mathbf{C}_0) h(\varphi; \mu, \sigma_0, \gamma_0) d\varphi \\
&= \frac{2\pi_a}{\sigma_0 a(\gamma_0)} \int_{\mu}^{\infty} \phi_p(\theta_0; \mathbf{m}_0 + \varphi \boldsymbol{\beta}_0, \mathbf{C}_0) \phi\left(\frac{\varphi - \mu}{\sigma_0 a(\gamma_0)}\right) d\varphi \\
&\quad + \frac{2\pi_b}{\sigma_0 b(\gamma_0)} \int_{-\infty}^{\mu} \phi_p(\theta_0; \mathbf{m}_0 + \varphi \boldsymbol{\beta}_0, \mathbf{C}_0) \phi\left(\frac{\varphi - \mu}{\sigma_0 b(\gamma_0)}\right) d\varphi
\end{aligned}$$

By making the change of variables $\varphi_a = \frac{\varphi - \mu}{\sigma_0 a(\gamma_0)}$ and $\varphi_b = \frac{-(\varphi - \mu)}{\sigma_0 b(\gamma_0)}$ we have

$$\begin{aligned}
p(\theta_0) &= 2\pi_a \int_0^{\infty} \phi_p(\theta_0; \xi_0 + \alpha_a \varphi_a, \mathbf{C}_0) \phi(\varphi_a; 0, 1) d\varphi_a \\
&\quad + 2\pi_b \int_0^{\infty} \phi_p(\theta_0; \xi_0 + \alpha_b \varphi_b, \mathbf{C}_0) \phi(\varphi_b; 0, 1) d\varphi_b \\
&= 2\pi_a \phi_p(\theta_0; \xi_0, \boldsymbol{\Omega}_a) \int_0^{\infty} \phi(\varphi_a; \boldsymbol{\alpha}_a^{\top} \boldsymbol{\Omega}_a^{-1} (\theta_0 - \xi_0), 1 - \boldsymbol{\alpha}_a^{\top} \boldsymbol{\Omega}_a^{-1} \boldsymbol{\alpha}_a) d\varphi_a \\
&\quad + 2\pi_b \phi_p(\theta_0; \xi_0, \boldsymbol{\Omega}_b) \int_0^{\infty} \phi(\varphi_b; \boldsymbol{\alpha}_b^{\top} \boldsymbol{\Omega}_b^{-1} (\theta_0 - \xi_0), 1 - \boldsymbol{\alpha}_b^{\top} \boldsymbol{\Omega}_b^{-1} \boldsymbol{\alpha}_b) d\varphi_b \\
&= 2\pi_a \phi_p(\theta_0; \xi_0, \boldsymbol{\Omega}_a) \Phi\left(\frac{\boldsymbol{\alpha}_a^{\top} \boldsymbol{\Omega}_a^{-1} (\theta_0 - \xi_0)}{\sqrt{1 - \boldsymbol{\alpha}_a^{\top} \boldsymbol{\Omega}_a^{-1} \boldsymbol{\alpha}_a}}\right) \\
&\quad + 2\pi_b \phi_p(\theta_0; \xi_0, \boldsymbol{\Omega}_b) \Phi\left(\frac{\boldsymbol{\alpha}_b^{\top} \boldsymbol{\Omega}_b^{-1} (\theta_0 - \xi_0)}{\sqrt{1 - \boldsymbol{\alpha}_b^{\top} \boldsymbol{\Omega}_b^{-1} \boldsymbol{\alpha}_b}}\right).
\end{aligned}$$

□

Proof of Proposition 3.3 By the induction hypotheses, the conditional distribution of φ given D_t is

$$p(\varphi|D_t) \propto p(\mathbf{y}_t|\varphi, D_{t-1})p(\varphi|D_{t-1}).$$

Using (6) and (16), we get

$$\begin{aligned}
p(\varphi|D_t) &\propto 2\pi_{t-1}^a \phi_r(\mathbf{y}_t; \mathbf{F}_t^{\top}(\mathbf{a}_t + \varphi \mathbf{b}_t), \boldsymbol{\Sigma}_t) \phi(\varphi; \eta_{t-1}^a, \tau_{t-1}^a) I_{[\mu, \infty)}(\varphi) \\
&\quad + 2\pi_{t-1}^b \phi_r(\mathbf{y}_t; \mathbf{F}_t^{\top}(\mathbf{a}_t + \varphi \mathbf{b}_t), \boldsymbol{\Sigma}_t) \phi(\varphi; \eta_{t-1}^b, \tau_{t-1}^b) I_{(-\infty, \mu)}(\varphi).
\end{aligned}$$

From the marginal/conditional representation of the multivariate normal distribution, we find, for $s = a, b$, the following identity:

$$\begin{aligned}
&\phi_r(\mathbf{y}_t; \mathbf{F}_t^{\top}(\mathbf{a}_t + \varphi \mathbf{b}_t), \boldsymbol{\Sigma}_t) \phi(\varphi; \eta_{t-1}^s, \tau_{t-1}^s) \\
&= \phi_r(\mathbf{y}_t; \mathbf{F}_t^{\top}(\mathbf{a}_t + \eta_{t-1}^s \mathbf{b}_t), \boldsymbol{\Sigma}_t + \tau_{t-1}^s \mathbf{F}_t^{\top} \mathbf{b}_t \mathbf{b}_t^{\top} \mathbf{F}_t) \phi(\varphi; \eta_t^s, \tau_t^s),
\end{aligned}$$

where, for $s = a, b$,

$$\begin{aligned} \eta_t^s &= \eta_{t-1}^s + \tau_{t-1}^s \mathbf{b}_t^\top \mathbf{F}_t (\boldsymbol{\Sigma}_t + \tau_{t-1}^s \mathbf{F}_t^\top \mathbf{b}_t \mathbf{b}_t^\top \mathbf{F}_t)^{-1} (\mathbf{y}_t - \mathbf{F}_t^\top (\mathbf{a}_t - \eta_{t-1}^s \mathbf{b}_t)) \\ &= \frac{\eta_{t-1}^s + \tau_{t-1}^s \mathbf{b}_t^\top \mathbf{F}_t \boldsymbol{\Sigma}_t^{-1} (\mathbf{y}_t - \mathbf{F}_t^\top \mathbf{a}_t)}{1 + \tau_{t-1}^s \mathbf{b}_t^\top \mathbf{F}_t \boldsymbol{\Sigma}_t^{-1} \mathbf{F}_t^\top \mathbf{b}_t}, \\ \tau_t^s &= \tau_{t-1}^s - [\tau_{t-1}^s]^2 \mathbf{b}_t^\top \mathbf{F}_t (\boldsymbol{\Sigma}_t + \tau_{t-1}^s \mathbf{F}_t^\top \mathbf{b}_t \mathbf{b}_t^\top \mathbf{F}_t)^{-1} \mathbf{F}_t^\top \mathbf{b}_t \\ &= \frac{\tau_{t-1}^s}{1 + \tau_{t-1}^s \mathbf{b}_t^\top \mathbf{F}_t \boldsymbol{\Sigma}_t^{-1} \mathbf{F}_t^\top \mathbf{b}_t}. \end{aligned}$$

Hence, we have

$$\begin{aligned} p(\varphi | D_t) &\propto \pi_{t-1}^a \phi_r(\mathbf{y}_t; \mathbf{F}_t^\top (\mathbf{a}_t + \eta_{t-1}^a \mathbf{b}_t), \boldsymbol{\Sigma}_t + \tau_{t-1}^a \mathbf{F}_t^\top \mathbf{b}_t \mathbf{b}_t^\top \mathbf{F}_t) \phi(\varphi; \eta_t^a, \tau_t^a) I_{[\mu, \infty)}(\varphi) \\ &\quad + \pi_{t-1}^b \phi_r(\mathbf{y}_t; \mathbf{F}_t^\top (\mathbf{a}_t + \eta_{t-1}^b \mathbf{b}_t), \boldsymbol{\Sigma}_t + \tau_{t-1}^b \mathbf{F}_t^\top \mathbf{b}_t \mathbf{b}_t^\top \mathbf{F}_t) \phi(\varphi; \eta_t^b, \tau_t^b) I_{(-\infty, \mu)}(\varphi), \end{aligned}$$

from where the inverse of the proportionality/normalization constant becomes

$$\begin{aligned} c_t^{-1} &= \pi_{t-1}^a \phi_r(\mathbf{y}_t; \mathbf{F}_t^\top (\mathbf{a}_t + \eta_{t-1}^a \mathbf{b}_t), \boldsymbol{\Sigma}_t + \tau_{t-1}^a \mathbf{F}_t^\top \mathbf{b}_t \mathbf{b}_t^\top \mathbf{F}_t) \int_{-\infty}^{\infty} \phi(\varphi; \eta_t^a, \tau_t^a) I_{[\mu, \infty)}(\varphi) d\varphi \\ &\quad + \pi_{t-1}^b \phi_r(\mathbf{y}_t; \mathbf{F}_t^\top (\mathbf{a}_t + \eta_{t-1}^b \mathbf{b}_t), \boldsymbol{\Sigma}_t + \tau_{t-1}^b \mathbf{F}_t^\top \mathbf{b}_t \mathbf{b}_t^\top \mathbf{F}_t) \int_{-\infty}^{\infty} \phi(\varphi; \eta_t^b, \tau_t^b) I_{(-\infty, \mu)}(\varphi) d\varphi \\ &= \pi_{t-1}^a \phi_r(\mathbf{y}_t; \mathbf{F}_t^\top (\mathbf{a}_t + \eta_{t-1}^a \mathbf{b}_t), \boldsymbol{\Sigma}_t + \tau_{t-1}^a \mathbf{F}_t^\top \mathbf{b}_t \mathbf{b}_t^\top \mathbf{F}_t) \Phi\left(-\frac{\mu - \eta_t^a}{\sqrt{\tau_t^a}}\right) \\ &\quad + \pi_{t-1}^b \phi_r(\mathbf{y}_t; \mathbf{F}_t^\top (\mathbf{a}_t + \eta_{t-1}^b \mathbf{b}_t), \boldsymbol{\Sigma}_t + \tau_{t-1}^b \mathbf{F}_t^\top \mathbf{b}_t \mathbf{b}_t^\top \mathbf{F}_t) \Phi\left(\frac{\mu - \eta_t^b}{\sqrt{\tau_t^b}}\right). \end{aligned} \tag{29}$$

Therefore, the conditional distribution of $\varphi | D_t$ can be written as

$$p(\varphi | D_t) = \pi_t^a \frac{\phi(\varphi; \eta_t^a, \tau_t^a)}{\Phi\left(-\frac{\mu - \eta_t^a}{\sqrt{\tau_t^a}}\right)} I_{[\mu, \infty)}(\varphi) + \pi_t^b \frac{\phi(\varphi; \eta_t^b, \tau_t^b)}{\Phi\left(\frac{\mu - \eta_t^b}{\sqrt{\tau_t^b}}\right)} I_{(-\infty, \mu)}(\varphi),$$

with

$$\begin{aligned} \pi_t^a &= c_t \pi_{t-1}^a \phi_r(\mathbf{y}_t; \mathbf{F}_t^\top (\mathbf{a}_t + \eta_{t-1}^a \mathbf{b}_t), \boldsymbol{\Sigma}_t + \tau_{t-1}^a \mathbf{F}_t^\top \mathbf{b}_t \mathbf{b}_t^\top \mathbf{F}_t) \Phi\left(-\frac{\mu - \eta_t^a}{\sqrt{\tau_t^a}}\right), \\ \pi_t^b &= c_t \pi_{t-1}^b \phi_r(\mathbf{y}_t; \mathbf{F}_t^\top (\mathbf{a}_t + \eta_{t-1}^b \mathbf{b}_t), \boldsymbol{\Sigma}_t + \tau_{t-1}^b \mathbf{F}_t^\top \mathbf{b}_t \mathbf{b}_t^\top \mathbf{F}_t) \Phi\left(\frac{\mu - \eta_t^b}{\sqrt{\tau_t^b}}\right), \end{aligned}$$

which correspond to the density of

$$\varphi|D_t \sim \pi_t^a TN(\eta_t^a, \tau_t^a, [\mu, \infty)) + \pi_t^b TN(\eta_t^b, \tau_t^b, (-\infty, \mu))$$

and the result is proved. \square

Proof of Proposition 3.4 Part (i): by using (14) and (6) we have

$$\begin{aligned} p(\theta_t|D_{t-1}) &= \int_{\varphi} p(\theta_t|\varphi, D_{t-1})p(\varphi|D_{t-1})d\varphi \\ &\propto \pi_{t-1}^a \int_{-\infty}^{\infty} \phi_p(\theta_t; \mathbf{a}_t + \varphi \mathbf{b}_t, \mathbf{R}_t) \phi(\varphi; \eta_{t-1}^a, \tau_{t-1}^a) d\varphi \\ &\quad + \pi_{t-1}^b \int_{-\infty}^{\infty} \phi_p(\theta_t; \mathbf{a}_t + \varphi \mathbf{b}_t, \mathbf{R}_t) \phi(\varphi; \eta_{t-1}^b, \tau_{t-1}^b) d\varphi \\ &= \pi_{t-1}^a \phi_p(\theta_t; \mathbf{a}_t + \eta_{t-1}^a \mathbf{b}_t, \mathbf{R}_t + \tau_{t-1}^a \mathbf{b}_t \mathbf{b}_t^\top) \int_{\mu}^{\infty} \phi(\varphi; \chi_t^a, \vartheta_t^a) d\varphi \\ &\quad + \pi_{t-1}^b \phi_p(\theta_t; \mathbf{a}_t + \eta_{t-1}^b \mathbf{b}_t, \mathbf{R}_t + \tau_{t-1}^b \mathbf{b}_t \mathbf{b}_t^\top) \int_{-\infty}^{\mu} \phi(\varphi; \chi_t^b, \vartheta_t^b) d\varphi \\ &= \pi_{t-1}^a \phi_p(\theta_t; \mathbf{a}_t + \eta_{t-1}^a \mathbf{b}_t, \mathbf{R}_t + \tau_{t-1}^a \mathbf{b}_t \mathbf{b}_t^\top) \Phi\left(-\frac{\mu - \chi_t^a}{\sqrt{\vartheta_t^a}}\right) \\ &\quad + \pi_{t-1}^b \phi_p(\theta_t; \mathbf{a}_t + \eta_{t-1}^b \mathbf{b}_t, \mathbf{R}_t + \tau_{t-1}^b \mathbf{b}_t \mathbf{b}_t^\top) \Phi\left(\frac{\mu - \chi_t^b}{\sqrt{\vartheta_t^b}}\right), \end{aligned}$$

where, for $s = a, b$, we use that

$$\phi_p(\theta_t; \mathbf{a}_t + \varphi \mathbf{b}_t, \mathbf{R}_t) \phi(\varphi; \eta_{t-1}^s, \tau_{t-1}^s) = \phi_p(\theta_t; \mathbf{a}_t + \eta_{t-1}^s \mathbf{b}_t, \mathbf{R}_t + \tau_{t-1}^s \mathbf{b}_t \mathbf{b}_t^\top) \phi(\varphi; \chi_t^s, \vartheta_t^s),$$

by considering again the marginal/conditional factorization of the multivariate normal density, and

$$\begin{aligned} \chi_t^s &= \eta_{t-1}^s + \tau_{t-1}^s \mathbf{b}_t^\top (\mathbf{R}_t + \tau_{t-1}^s \mathbf{b}_t \mathbf{b}_t^\top)^{-1} (\theta_t - \mathbf{a}_t - \eta_{t-1}^s \mathbf{b}_t) \\ &= \frac{\eta_{t-1}^s + \tau_{t-1}^s \mathbf{b}_t^\top \mathbf{R}_t^{-1} (\theta_t - \mathbf{a}_t)}{1 + \tau_{t-1}^s \mathbf{b}_t^\top \mathbf{R}_t^{-1} \mathbf{b}_t}, \\ \vartheta_t^s &= \tau_{t-1}^s - [\tau_{t-1}^s]^2 \mathbf{b}_t^\top (\mathbf{R}_t + \tau_{t-1}^s \mathbf{b}_t \mathbf{b}_t^\top)^{-1} \mathbf{b}_t \\ &= \frac{\tau_{t-1}^s}{1 + \tau_{t-1}^s \mathbf{b}_t^\top \mathbf{R}_t^{-1} \mathbf{b}_t}, \end{aligned}$$

which leads to the proof of part (i).

Part (ii): The proof of this part is direct from (29) since $c_t^{-1} = p(\mathbf{y}_t|D_{t-1})$.

Part (iii): Proceeding similarly to part (i), but now by using part (iii) of Proposition 3.2 and Proposition 3.3, we get

$$\begin{aligned}
 p(\theta_t|D_t) &= \int_{-\infty}^{\infty} p(\theta_t|\varphi, D_t)p(\varphi|D_t)d\varphi \\
 &= \pi_t^a \int_{\mu}^{\infty} \phi_p(\theta_t; \mathbf{m}_t + \varphi \boldsymbol{\beta}_t, \mathbf{C}_t) \frac{\phi(\varphi; \eta_t^a, \tau_t^a)}{\Phi\left(-\frac{\mu - \eta_t^a}{\sqrt{\tau_t^a}}\right)} d\varphi \\
 &\quad + \pi_t^b \int_{-\infty}^{\mu} \phi_p(\theta_t; \mathbf{m}_t + \varphi \boldsymbol{\beta}_t, \mathbf{C}_t) \frac{\phi(\varphi; \eta_t^b, \tau_t^b)}{\Phi\left(\frac{\mu - \eta_t^b}{\sqrt{\tau_t^b}}\right)} d\varphi \\
 &= \pi_t^a \phi_p(\theta_t; \mathbf{m}_t + \eta_t^a \boldsymbol{\beta}_t, \mathbf{C}_t + \tau_t^a \boldsymbol{\beta}_t \boldsymbol{\beta}_t^{\top}) \int_{\mu}^{\infty} \frac{\phi(\varphi; \delta_t^a, v_t^a)}{\Phi\left(-\frac{\mu - \eta_t^a}{\sqrt{\tau_t^a}}\right)} d\varphi \\
 &\quad + \pi_t^b \phi_p(\theta_t; \mathbf{m}_t + \eta_t^b \boldsymbol{\beta}_t, \mathbf{C}_t + \tau_t^b \boldsymbol{\beta}_t \boldsymbol{\beta}_t^{\top}) \int_{-\infty}^{\mu} \frac{\phi(\varphi; \delta_t^b, v_t^b)}{\Phi\left(\frac{\mu - \eta_t^b}{\sqrt{\tau_t^b}}\right)} d\varphi \\
 &= \pi_t^a \phi_p(\theta_t; \mathbf{m}_t + \eta_t^a \boldsymbol{\beta}_t, \mathbf{C}_t + \tau_t^a \boldsymbol{\beta}_t \boldsymbol{\beta}_t^{\top}) \frac{\Phi\left(-\frac{\mu - \delta_t^a}{\sqrt{v_t^a}}\right)}{\Phi\left(-\frac{\mu - \eta_t^a}{\sqrt{\tau_t^a}}\right)} \\
 &\quad + \pi_t^b \phi_p(\theta_t; \mathbf{m}_t + \eta_t^b \boldsymbol{\beta}_t, \mathbf{C}_t + \tau_t^b \boldsymbol{\beta}_t \boldsymbol{\beta}_t^{\top}) \frac{\Phi\left(\frac{\mu - \delta_t^b}{\sqrt{v_t^b}}\right)}{\Phi\left(\frac{\mu - \eta_t^b}{\sqrt{\tau_t^b}}\right)},
 \end{aligned}$$

where, for

$$\begin{aligned}
 \delta_t^s &= \eta_t^s + \tau_t^s \boldsymbol{\beta}_t^{\top} (\mathbf{C}_t + \tau_t^s \boldsymbol{\beta}_t \boldsymbol{\beta}_t^{\top})^{-1} (\theta_t - \mathbf{m}_t - \eta_t^s \boldsymbol{\beta}_t) = \frac{\eta_t^s + \tau_t^s \boldsymbol{\beta}_t^{\top} \mathbf{C}_t^{-1} (\theta_t - \mathbf{m}_t)}{1 + \tau_t^s \boldsymbol{\beta}_t^{\top} \mathbf{C}_t^{-1} \boldsymbol{\beta}_t}, \\
 v_t^s &= \tau_t^s - (\tau_t^s)^2 \boldsymbol{\beta}_t^{\top} (\mathbf{C}_t + \tau_t^s \boldsymbol{\beta}_t \boldsymbol{\beta}_t^{\top})^{-1} \boldsymbol{\beta}_t = \frac{\tau_t^s}{1 + \tau_t^s \boldsymbol{\beta}_t^{\top} \mathbf{C}_t^{-1} \boldsymbol{\beta}_t}.
 \end{aligned}$$

This proves part (iii). □

Funding Open access funding provided by Università degli Studi di Padova within the CRUI-CARE Agreement.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article

are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Arellano-Valle RB, Azzalini A, Ferreira CS, Santoro K (2020) A two-piece normal measurement error model. *Comput Stat Data Anal* 144:106863
- Arellano-Valle RB, Contreras-Reyes JE, Quintero FOL, Valdebenito A (2019) A skew-normal dynamic linear model and Bayesian forecasting. *Comput Stat* 34(3):1055–1085
- Arellano-Valle RB, Gómez HW, Quintana FA (2005) Statistical inference for a general class of asymmetric distributions. *J Stat Plan Inference* 128(2):427–443
- Azzalini A, Capitanio A (1999) Statistical applications of the multivariate skew normal distribution. *J R Stat Soc Ser B Stat Methodol* 61:579–602
- Barr DR, Sherrill ET (1999) Mean and variance of truncated normal distributions. *Am Stat* 53(4):357–361
- Cabral CRB, Da-Silva CQ, Migon HS (2014) A dynamic linear model with extended skew-normal for the initial distribution of the state parameter. *Comput Stat Data Anal* 74:64–80
- Carter CK, Kohn R (1994) On Gibbs sampling for state space models. *Biometrika* 81(3):541–553
- Corns T, Satchell S (2007) Skew Brownian motion and pricing European options. *Eur J Financ* 13(6):523–544
- Fasano A, Rebaudo G, Durante D, Petrone S (2021) A closed-form filter for binary time series. *Stat Comput* 31(4):1–20
- Frühwirth-Schnatter S (1994) Data augmentation and dynamic linear models. *J Time Ser Anal* 15(2):183–202
- Gualtierotti A (2005) Skew-normal processes as models for random signals corrupted by Gaussian noise. *Int J Pure Appl Math* 20:109–142
- Kalman RE (1960) A new approach to linear filtering and prediction problems. *J Basic Eng* 82:35–45
- Kim H-M, Ryu D, Mallick BK, Genton MG (2014) Mixtures of skewed Kalman filters. *J Multivar Anal* 123:228–251
- Naveau P, Genton MG, Shen X (2005) A skewed Kalman filter. *J Multivar Anal* 94(2):382–400
- Petris G, Petrone S, Campagnoli P (2009) *Dynamic linear models with R*. Springer, Berlin
- Pourahmadi M (2007) Skew-normal ARMA models with nonlinear heteroscedastic predictors. *Commun Stat Theory Methods* 36(9):1803–1819
- Shumway RH, Stoffer DS, Stoffer DS (2000) *Time series analysis and its applications, vol 3*. Springer, Berlin
- West M, Harrison J (1997) *Bayesian forecasting and dynamic models*. Springer Science & Business Media, Berlin

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.