A non-Gaussian family of state-space models with exact marginal likelihood

Dani Gamerman[a] Thiago Rezende dos Santos[b] Glaura C. Franco[b*]

[a] *Department of Statistical Methods, Universidade Federal do Rio de Janeiro, RJ, Brazil.*
[b] *Department of Statistics, Universidade Federal de Minas Gerais, MG, Brazil.*

Abstract

The Gaussian assumption generally employed in many state space models is usually not satisfied for real time series. Thus, in this work a broad family of non-Gaussian models is defined by integrating and expanding previous work in the literature. The expansion is obtained at two levels: at the observational level it allows for many distributions not previously considered and at the latent state level it involves an expanded specification for the system evolution. The class retains analytical availability of the marginal likelihood function, uncommon outside Gaussianity. This expansion considerably increases the applicability of the models and solves many previously existing problems such as long-term prediction, missing values and irregular temporal spacing. Inference about the state components can be performed due to the introduction of a new and exact smoothing procedure, in addition to filtered distributions. Inference for the hyperparameters is presented from the classical and Bayesian perspectives. Performance of the model can be assessed through diagnostic tools and predictive and fit measures. The results seem to indicate competitive results of the models when compared to other non-Gaussian state-space models available. The methodology is applied to Gaussian and non-Gaussian dynamic linear models with time-varying means and variances and provides a computationally simple solution to inference in these models. The methodology is illustrated in a number of examples, including time series with unknown data support. The paper is concluded with directions for further work.

Keywords: classical inference; Bayesian; forecasting; non-linear system evolution; smoothing.

1 Introduction

Several models are built in the literature based on the normality, homoscedasticity and independence assumptions of the errors. However, in many situations, these assumptions are not satisfied. Error independence is rarely attained in the time series context, while the normality assumption is often discarded in a wide variety of situations.

*Correspondence author: Glaura C. Franco. Address: Universidade Federal de Minas Gerais, Belo Horizonte, MG, 31270-901, Brazil. E-mail:glauraf@ufmg.br

Many state space models (SSM) use results based on normality (Harvey, 1989; West & Harrison , 1997, see) largely due to their analytical tractability. This nice and computationally convenient feature is lost once the normality assumption is removed. The goal of this paper is to present possibilities for analyzing time series under the SSM framework beyond Gaussianity while still retaining tractability.

The origin of this work can be traced back to Nelder & Wedderburn (1972), who proposed the *generalized linear models* (GLM). The basic idea of these models consists in opening the range of options for the response-variable distribution, by allowing it to belong to the exponential family of distributions. According to Nelder & Wedderburn (1972), a link function is used to relate the mean function of the data to the linear predictor. The paper of Nelder & Wedderburn (1972) was concerned with independent data, while in the time series context, the observations generally present a correlation structure.

In this direction, a more general structure, called *Dynamic Generalized Linear Models* (DGLM), proposed by West, Harrison & Migon (1985), attracted a great deal of interest due to the applicability of GLM in time series. Papers in this context include Grunwald, Raftery & Guttorp (1993), Fahrmeir & Kaufmann (1987), Lindsey & Lambert (1995), Gamerman (1991, 1998), Chiogna & Gaetan (2002), Shephard & Pitt (1997) and Godolphin & Triantafyllopoulos (2006). The books by Durbin & Koopman (2001) and Fahrmeir & Tutz (2001) also present and discuss these models in addition to providing alternatives for analyzing non-Gaussian time series. The problem with this class of models is that the analytical form is easily lost, even using simple model components. Thus, the predictive distribution, that is essential for the inference process, can only be obtained approximately.

Many researchers have worked in the last decades in auxiliary procedures in order to make inference for non-Gaussian state space models under the Bayesian approach. Gamerman (1998) proposed a GLM based MCMC algorithm for inference, Frühwirth-Schnatter & Wagner (2006) used auxiliary mixture sampling and Pitt & Walker (2005) utilized auxiliary variables for constructing stationary time series. Andrieu & Doucet (2002) and Carvalho et al. (2010) are just a sample of a large community devoted to inference via sequential Monte Carlo or particle filtering in the state space models context. However, it is important to emphasize that these methods are approximated and the computing time can be large.

The objective of this paper is to present a family of models that allows for exact, analytical computation of the marginal likelihood and thus the predictive distribution. This family is obtained by generalizations of results from Smith & Miller (1986). They considered an exponential observational model and an exact evolution equation that enables the analytical integration of the state and the attainment of the 1-step-ahead predictive distributions. Harvey & Fernandes (1989) and Shephard (1994) showed that the same tractability was obtained with Poisson and normal scale observational models, respectively. The family introduced in this paper unifies these and several other models that were treated separately. Additionally, an expanded evolution equation is proposed, thus allowing for exact expression for long-term prediction. It will also be seen that these exact computations are very easy to code.

Thus, the main contributions of this paper are to consider and to characterize this family of models, by presenting special cases of observational time series that belong

to this family. In addition, a smoothing procedure, similar to the forward filtering backward sampling (FFBS) algorithm used in Gaussian SSM, is also introduced. General expressions for k-step-ahead predictions are provided and treatment of time series containing missing data and/or unequally spaced observation times is considered. The ideas of these models are also applied to Gaussian and non-Gaussian dynamic linear models and are shown to provide a computationally simple solution to the problem of estimation when both means and variances are subject to temporal variation.

The paper is organized as follows. The models are presented in Section 2, where the main theoretical results for filtered or on-line inference for the latent variables and prediction of future observations are provided. Section 3 introduces the smoothing procedure, vital for inference on the latent variables based on the complete time series. Section 4 presents the inferential procedure for the model hyperparameter from classical and Bayesian perspectives. Point and interval estimation procedures are described in addition to model checking procedures. Section 5 revisits dynamic linear models and shows how the methods of this paper can be adapted to cope with time-varying observational means and variances. Section 6 presents illustrative analysis of real data series in a number of areas of application. Finally, Section 7 presents the main conclusions and final remarks.

2 Model definition and properties

In this section, the models are introduced. They are a generalization and integration of the non-Gaussian dynamic models of Smith & Miller (1986) and Harvey & Fernandes (1989). The main advantage of these models compared to the DGLM is that exact inference can be performed. Due to the form of the model equations, some common components, such as seasonality and influence of other predictor series, are inserted as fixed effects.

A time series $\{y_t\}$ is in this class of models if it satisfies the following assumptions:

A0 its probability (density) function can be written in the form:

$$\begin{aligned} p(y_t|\mu_t, \mu_{t-1}, \ldots, \mu_1, \boldsymbol{Y}_{t-1}, \boldsymbol{\varphi}) &= p(y_t|\mu_t, \boldsymbol{Y}_{t-1}, \boldsymbol{\varphi}) \\ &= a(y_t, \boldsymbol{\varphi}) \mu_t^{b(y_t, \boldsymbol{\varphi})} \exp\left(-\mu_t c(y_t, \boldsymbol{\varphi})\right), \end{aligned} \quad (1)$$

for $y_t \in S(\boldsymbol{\varphi}) \subset \Re$ and $p(y_t|\mu_t, \boldsymbol{\varphi}) = 0$, otherwise. Functions $a(\cdot)$, $b(\cdot)$, $c(\cdot)$ and $S(\cdot)$ are such that $p(y_t|\mu_t, \boldsymbol{\varphi}) \geq 0$ and therefore $\mu_t > 0$, for all $t > 0$. It is also assumed that $\boldsymbol{\varphi}$ varies in the p-dimensional parameter space Φ.

A1 If x_t is a covariate vector, the link function g relates the predictor to the parameter μ_t through the relation $\mu_t = \lambda_t g(x_t, \boldsymbol{\beta})$, where $\boldsymbol{\beta}$ are the regression coefficients (one of the components of $\boldsymbol{\varphi}$) and λ_t is the latent state variable related to the description of the dynamic level. If the predictor is linear, then $g(x_t, \boldsymbol{\beta}) = g(x_t^{'}\boldsymbol{\beta})$.

A2 The dynamic level λ_t evolves according to the system equation $\lambda_{t+1} = w^{-1}\lambda_t \varsigma_{t+1}$ where $\varsigma_{t+1}|\boldsymbol{Y}_t, \boldsymbol{\varphi} \sim Beta\left(wa_t, (1-w)a_t\right)$, that is,

$$w\frac{\lambda_{t+1}}{\lambda_t} \mid \lambda_t, \boldsymbol{Y}_t, \boldsymbol{\varphi} \sim Beta\left(wa_t, (1-w)a_t\right),$$

where a_t are known constants to be specified later in the text, $0 < w \leq 1$, $\boldsymbol{Y}_t = \{Y_0, y_1, \ldots, y_t\}$, for $t = 1, 2, ...$ and Y_0 represents previously available information.

A3 The dynamic level λ_t is initialized with prior distribution $\lambda_0 | Y_0 \sim Gamma(a_0, b_0)$.

The data-dependent evolution **A2** is not usually encountered in most SSM but does not invalidate model specification. Appendix A details the calculations that prove that the state-space model above is well-defined even with this evolution.

There is a wide range of distributions that belong to this class of models. It includes many commonly known discrete and continuous distributions such as Poisson, Gamma and Normal (with static mean) but also includes many other distributions that are not so common.

Table 1 provides the form of functions a, b, c and S for some distributions in this family. The more common cases such as Poisson and Exponential were previously singled out in the literature. Several other cases of this family are introduced here and they include continuous and discrete distributions. Some of them are well known such as Normal and Pareto but the family includes also the Borel-Tanner (Haight & Breuer , 1960) and the Rayleigh distributions, for example. The picture that emerges is a collection that is capable of representing a variety of features that are present in time series applications and thus are of practical importance.

Table 1: Special cases of the non-Gaussian SSM.

Models	$\boldsymbol{\varphi}$	$a(y_t, \boldsymbol{\varphi})$	$b(y_t, \boldsymbol{\varphi})$	$c(y_t, \boldsymbol{\varphi})$	$S(\boldsymbol{\varphi})$
Poisson	$(w, \beta)'$	$(y_t!)^{-1}$	y_t	1	$\{0, 1, \ldots\}$
Borel-Tanner	(w, ρ, β)	$\frac{\rho}{(y_t-\rho)!} y_t^{y_t-\rho-1}$	$y_t - \rho$	y_t	$\{\rho, \rho+1, \ldots\}$
Gamma	(w, χ, β)	$y_t^{\chi-1}/\Gamma(\chi)$	χ	y_t	$(0, \infty)$
Weibull	(w, ν, β)	$\nu (y_t)^{\nu-1}$	1	$(y_t)^{\nu}$	$(0, \infty)$
Pareto	(w, ρ, β)	y_t^{-1}	1	$\ln y_t - \ln \rho$	(ρ, ∞)
Normal	(w, θ)	$(2\pi)^{-1/2}$	$1/2$	$(y_t - \theta)^2/2$	$(-\infty, \infty)$
Laplace	(w, θ)	$\frac{1}{\sqrt{2}}$	1	$\sqrt{2}\,\lvert y_t - \theta \rvert$	$(-\infty, \infty)$
Inverse Gaussian	(w, θ)	$\frac{1}{\sqrt{2\pi y_t^3}}$	$1/2$	$\frac{(y_t-\theta)^2}{2 y_t \theta^2}$	$(0, \infty)$
Rayleigh	(w, θ)	y_t	1	$\frac{1}{2}(y_t - \theta)^2$	$(0, \infty)$
Power Exponential (GED)	(w, ν, κ, θ)	$\frac{\nu}{\kappa 2^{\frac{\nu+1}{\nu}} \Gamma(1/\nu)}$	$1/\nu$	$\frac{(y_t-\theta)^{\nu}}{2\kappa^{\nu}}$	$(-\infty, \infty)$
Generalized Gamma	(w, ν, χ)	$\nu y_t^{\nu\chi-1}/\Gamma(\chi)$	χ	y_t^{ν}	$(0, \infty)$

The usual specification for the link function g is the logarithmic function given the positive nature of μ_t, but other link functions dictated by the application may also be used. It is interesting to note that the evolution equation can be rewritten as $\ln(\lambda_t) = \ln(\lambda_{t-1}) + \varsigma_t^*$, where $\varsigma_t^* = \ln(\varsigma_t / w) \in \Re$, similar to the random walk evolution of the non-Gaussian local level model. The main difference with respect to other common non-Gaussian SSM is the use of the scaled log-Beta disturbances instead of normal disturbances. This change proves vital for obtaining exact results.

The parameter w varies between 0 and 1 and also belongs to $\boldsymbol{\varphi}$. As it will be seen in what follows, w is responsible for increasing the variance over time. Thus, it plays

a similar role to that of discount factors, used in the Bayesian approach to state space models.

The family in **A0** is quite general and functions a, b and c must satisfy constraints to ensure a proper distribution. Many properties can also be derived as, for example, $E[b(y_t, \boldsymbol{\varphi})] = \mu_t E[c(y_t, \boldsymbol{\varphi})]$. However, these derivations are not required in the sequel and are not pursued here. If $b(y_t, \boldsymbol{\varphi}) = b(y)$ or $c(y_t, \boldsymbol{\varphi}) = c(y)$ and $S(\boldsymbol{\varphi})$ does not depend on $\boldsymbol{\varphi}$, the observational model in **A0** becomes a special case of the exponential family of distributions.

Theorem 1 below provides basic results of these models for sequential or on-line inference for the level λ_t (filtering results) and the predictive distribution.

Theorem 1. If the model is defined in **A0**-**A3**, the following results can be obtained for t=1, 2, ...

1. the prior distribution $\lambda_t|\boldsymbol{Y}_{t-1}, \boldsymbol{\varphi}$ follows a Gamma($a_{t|t-1}$, $b_{t|t-1}$) distribution such that

$$a_{t|t-1} = w a_{t-1}, \tag{2}$$
$$b_{t|t-1} = w b_{t-1}. \tag{3}$$

2. The on-line or updated distribution of $\lambda_t|\boldsymbol{Y}_t, \boldsymbol{\varphi}$ is Gamma (a_t,b_t), where

$$a_t = a_{t|t-1} + b(y_t, \boldsymbol{\varphi}), \tag{4}$$
$$b_t = b_{t|t-1} + c(y_t, \boldsymbol{\varphi}). \tag{5}$$

3. The one step ahead predictive density function is given by

$$p(y_t|\boldsymbol{Y}_{t-1}, \boldsymbol{\varphi}) = \frac{\Gamma(b(y_t, \boldsymbol{\varphi}) + a_{t|t-1}) a(y_t, \boldsymbol{\varphi}) (b_{t|t-1})^{a_{t|t-1}}}{\Gamma(a_{t|t-1}) [c(y_t, \boldsymbol{\varphi}) + b_{t|t-1}]^{b(y_t, \boldsymbol{\varphi}) + a_{t|t-1}}}, \quad y_t \in S(\boldsymbol{\varphi}), \tag{6}$$

$\forall t \in N$ and $\Gamma(\cdot)$ is the gamma function.

The proof of Theorem 1 is in Appendix B.

The corresponding distribution of $\mu_t = \lambda_t g(x_t, \boldsymbol{\beta})$ is easily obtained from Items 1 and 2 of Theorem 1, using the scale property of the Gamma distribution. For example, $\mu_t|\mathbf{Y}_t \sim$ Gamma $\left[a_t, b_t (g(x_t, \boldsymbol{\beta}))^{-1}\right]$.

Note from (2)-(3) that $E(\lambda_t|\boldsymbol{Y}_{t-1}) = E(\lambda_{t-1}|\boldsymbol{Y}_{t-1})$ and $Var(\lambda_t|\boldsymbol{Y}_{t-1}, \boldsymbol{\varphi}) = w^{-1} Var(\lambda_{t-1}|\ \boldsymbol{Y}_{t-1}, \boldsymbol{\varphi})$. The passage of time from $t-1$ to t implies that means are preserved but only $100w\%$ of the information (in terms of precision) is retained. This is exactly the role of the discount factors used in West & Harrison (1997). Theorem 1 enables exact inference about the state parameters when $\boldsymbol{\varphi}$ is known. This is a distinctive feature of the models **A0**-**A3** and it is rarely seen outside linear normal SSM.

The model can be trivially extended for time-varying w. One would simply replace w by w_t in **A2** and the only changes implied are replacement of w by w_t in item 1 of

Theorem 1. This issue allows extra flexibility that will prove important for handling irregularities such as missing points or temporally unequal-spaced observations.

Longer term predictions can be obtained by applying the evolution repeatedly over time. The predictive distributions are obtained by

$$\begin{aligned} f(y_{t+h}|\boldsymbol{Y}_t, \boldsymbol{\varphi}) &= \int f(y_{t+1:t+h}|\boldsymbol{Y}_t, \boldsymbol{\varphi})\, dy_{t+1:t+h-1} \\ &= \int \prod_{j=1}^{h} f(y_{t+j}|\boldsymbol{Y}_{t+j-1}, \boldsymbol{\varphi})\, dy_{t+1:t+h-1} \end{aligned} \tag{7}$$

where the notation of Apendix A is used for indexing vectors. This integration can not be performed analytically. Nevertheless, (7) provides the way to sample from $(y_{t+h}|\boldsymbol{Y}_t, \boldsymbol{\varphi})$. This is described in the algorithm below:

1. set $j = 1$;
2. draw y^v_{t+j} from $f(y_{t+j}|\boldsymbol{Y}_{t+j-1}, \boldsymbol{\varphi})$ and set $\boldsymbol{Y}_{t+j} = \{\boldsymbol{Y}_{t+j-1}, y^v_{t+j}\}$;
3. set $j \to j + 1$ and return to 1, if $j \leq h$; otherwise, stop.

In many cases, draws y^v_{t+j} from $f(y_{t+j}|\boldsymbol{Y}_{t+j-1}, \boldsymbol{\varphi})$ are easier to obtain with the intermediate step of a draw λ^v_{t+j} from $f(\lambda_{t+j}|\boldsymbol{Y}_{t+j-1}, \boldsymbol{\varphi})$ followed by drawing y^v_{t+j} from $f(y_{t+j}|\lambda^v_{t+j}, \boldsymbol{Y}_{t+j-1}, \boldsymbol{\varphi})$. Note also that $y^v_{t+1:t+h} = (y^v_{t+1}, ..., y^v_{t+h})$ are a joint sample from the predictive trajectory $f(y_{t+1:t+h}|\boldsymbol{Y}_t, \boldsymbol{\varphi})$.

The distribution in (7) describes the uncertainty associated with the time series forecasts. Summary measures, like mean, median and percentiles can be easily extracted from them. These results are based on knowledge of $\boldsymbol{\varphi}$. The procedures required when $\boldsymbol{\varphi}$ is unknown are described in Section 4.

Unfortunately, analytic expressions for these predictive distributions are not available. The multiplicative nature of the evolution may be used as a basis to obtain analytic approximations. It suggests that the h-steps ahead evolution could be approximated by

$$\lambda_{t+h}|\boldsymbol{Y}_t, \boldsymbol{\varphi} \quad \sim \quad \text{Gamma}(a_{t+h|t}, b_{t+h|t}) \tag{8}$$

where $a_{t+h|t} = w^h a_t$ and $b_{t+h|t} = w^h b_t$. This would lead to the predictive density function of the observations h steps ahead

$$p(y_{t+h}|\boldsymbol{Y}_t, \boldsymbol{\varphi}) = \frac{\Gamma(b(y_{t+h}, \boldsymbol{\varphi}) + a_{t+h|t}) a(y_{t+h}, \boldsymbol{\varphi}) (b_{t+h|t})^{a_{t+h|t}}}{\Gamma(a_{t+h|t}) [c(y_{t+h}, \boldsymbol{\varphi}) + b_{t+h|t}]^{b(y_{t+h}, \boldsymbol{\varphi}) + a_{t+h|t}}}, \quad y_{t+h} \in S(\boldsymbol{\varphi}). \tag{9}$$

Appendix B shows that predictive distributions (8) and (9) are obtained if **A2** were generalised to

$$w \frac{\lambda_{t+h+1}}{\lambda_{t+h}} \mid \lambda_{t+h}, \boldsymbol{Y}_t, , \boldsymbol{\varphi} \sim Beta\left(w^{h+1} a_t, (1-w) w^h a_t\right), \text{ for } h = 0, 1, 2, ..., \tag{10}$$

Simulations not reported here show that assumption (10) and hence density (9) provide reasonable approximations for single digit horizons and could be used as an alternative.

The above results could be used in conjunction with a time-varying w for handling irregularly spaced series and missing observations. If a collection of r observations are missing after time s then w could be replaced by w^r for the evolution at time s. Similarly, if observations are collected at times $t_1, t_2, ...$ then this data irregularity could be reflected by respective replacements of w by $w^{t_i - t_{i-1}}$, for all i.

3 Smoothing

The main interest in many situations is to estimate the level component $\lambda = (\lambda_1, \ldots, \lambda_n)'$ based on all available information $\boldsymbol{Y}_n$ instead of the sequence of marginal on-line distributions $\lambda_t|\boldsymbol{Y}_t$, $\forall t$. Smoothing techniques should be used in these cases. Harvey & Fernandes (1989) present an estimate of the level component in an application to a real series, obtained by using an approximated smoothing algorithm of fixed interval (Harvey, 1989). They acknowledge the approximating nature of their scheme by naming it a *quasi-smoothing* procedure.

In the normal linear context, the joint distribution of all state parameters given all available information is multivariate normal and its expression is given in Migon et al. (2005). The precision or inverse covariance matrix is in tridiagonal block form, as a reflection of the Markovian structure of the model. This opens up the possibility for devising better schemes based on the sparcity of this matrix. This is specially relevant when the time series length n is large. One such scheme is the FFBS algorithm proposed by Frühwirth-Schnatter (1994) and Carter & Kohn (1994). This algorithm shows how the multivariate normal joint distribution of the states can be broken down into smaller but still normal components.

The FFBS decomposition is always possible in SSM but the linear normal SSM is the only known situation where it leads to a tractable solution. Theorem 2 below provides a non-Gaussian version of FFBS for the models of this paper, showing that they also lead to a tractable solution.

Theorem 2. The joint distribution of $(\lambda|\boldsymbol{Y}_n, \boldsymbol{\varphi})$ has density

$$p(\lambda|\boldsymbol{\varphi}, \boldsymbol{Y}_n) = p(\lambda_n|\boldsymbol{\varphi}, \boldsymbol{Y}_n) \prod_{t=1}^{n-1} p(\lambda_t|\lambda_{t+1}, \boldsymbol{\varphi}, \boldsymbol{Y}_t) p(\boldsymbol{\varphi}|\boldsymbol{Y}_n),$$

where the distribution of $(\lambda_t|\lambda_{t+1}, \boldsymbol{\varphi}, \boldsymbol{Y}_t)$, is given by

$$\lambda_t - w\lambda_{t+1}|\lambda_{t+1}, \boldsymbol{\varphi}, \boldsymbol{Y}_t \sim \text{Gamma}\left((1-w)a_t, b_t\right), \forall t \geq 0. \tag{11}$$

Proof of Theorem 2 can be found in Appendix B.

Based on Theorem 2, an exact sample of the joint distribution of $(\lambda|\boldsymbol{\varphi}, \boldsymbol{Y}_n)$ can be obtained following the algorithm below:

1. set $t = n$ and sample $p(\lambda_n|\boldsymbol{\varphi}, \boldsymbol{Y}_n)$, using Theorem 1 with $t = n$;
2. set $t = t - 1$ and sample $p(\lambda_t|\lambda_{t+1}, \boldsymbol{\varphi}, \boldsymbol{Y}_t)$, using (11);

3. if $t > 1$, go back to step 2; otherwise, the sample of $(\lambda_1, \ldots, \lambda_n|\boldsymbol{\varphi}, \boldsymbol{Y}_n)$ is complete.

The result in (11) allows the implementation of step 2 of the algorithm above and thus enables an exact sample from the smoothed distribution of the states once the hyperparameter is known. This result will prove to be crucial for inference about the states even when the hyperparameter is not known.

4 Inference for hyperparameters

The model parameters were divided into the latent states $\{\lambda_t\}$ and fixed parameters $\boldsymbol{\varphi}$, usually called hyperparameters. The *on-line* and smoothed inference for the state parameters were presented in Sections 2 and 3, respectively. Knowledge of the hyperparameter was assumed in both cases. In this section, inference for the hyperparameters and the latent states is discussed.

4.1 Classical Inference

One way of making classical inference about the parameter vector $\boldsymbol{\varphi}$ is through the marginal likelihood function, whose form is given by

$$L(\boldsymbol{\varphi}; \boldsymbol{Y}_n) = \prod_{t=1}^{n} p(y_t|\boldsymbol{Y}_{t-1}, \boldsymbol{\varphi}) = \prod_{t=1}^{n} \frac{\Gamma(b(y_t,\boldsymbol{\varphi})+a_{t|t-1})a(y_t,\boldsymbol{\varphi})(b_{t|t-1})^{a_{t|t-1}}}{\Gamma(a_{t|t-1})[c(y_t,\boldsymbol{\varphi})+b_{t|t-1}]^{b(y_t,\boldsymbol{\varphi})+a_{t|t-1}}}, \quad (12)$$

where $y_t \in S(\boldsymbol{\varphi})$ and $\boldsymbol{\varphi}$ is composed by w, β and by parameters of the specific model. Maximization of the marginal likelihood function (12) is typically performed numerically.

The Gamma prior distribution used as the initial distribution for μ_t tends to become non-informative when $a_0, b_0 \to 0$, and is improper when $a_0 = b_0 = 0$. Note that if $a_0, b_0 \to 0$, the on-line distribution of $(\lambda_1|\boldsymbol{Y}_1)$ can be improper, so that the predictive density function will not be defined. Thus, from now on it will be assumed that $a_0 > 0$ and $b_0 > 0$. If $a_0 = b_0 = 0$, the index t in the product above can be initialized at instant $t = \tau$ instead of $t = 1$, where τ is the smallest value of t for which the distribution of $[\lambda_t|\boldsymbol{Y}_t]$ is proper.

Under some regularity conditions, the asymptotic properties of the maximum likelihood estimator (MLE) (Harvey, 1989, page 128) lead to

$$I_n^{1/2}(\hat{\boldsymbol{\varphi}})(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}) \xrightarrow{\mathrm{D}} \mathrm{N}[\boldsymbol{0}, I_p], \quad (13)$$

where $\hat{\varphi}$ is the MLE of φ, $I_n(\boldsymbol{\varphi})$ is the Fisher information matrix and I_p is the identity matrix of dimension p. The conditions are satisfied for the non-Gaussian SSM, subjected to:

1. $\boldsymbol{\varphi}$ is an interior point of the parametric space. This happens if $w < 1$ and all other components of $\boldsymbol{\varphi}$ defined in the positive line are not equal to zero.

2. the derivatives, at an arbitrary instant t, for $t = 1, \ldots, n$, exist and can be computed by derivation of the predictive density function in (6) with respect to $\boldsymbol{\varphi}$ and its continuity is the result of the continuity of $a_{t|t-1}$ and $b_{t|t-1}$. When one of the model parameters depends on the support, the derivatives of the respective parameter may not exist.

3. two different points $\boldsymbol{\varphi}^1$ and $\boldsymbol{\varphi}^2$ will produce different models and, as a result, different values for the likelihood function.

The asymptotic confidence interval for $\boldsymbol{\varphi}$ is built based on a numerical approximation for $I_n(\boldsymbol{\varphi})$, using $I_n(\boldsymbol{\varphi}) \cong -H(\boldsymbol{\varphi})$, where $-H(\boldsymbol{\varphi})$ is the matrix of second derivatives of the log-likelihood function with respect to the parameters.

Let φ_i, $i = 1, \ldots, p$, be any component of $\boldsymbol{\varphi}$. Then, an asymptotic confidence interval of $100(1-\kappa)\%$ for φ_i is given by

$$\hat{\varphi}_i \pm z_{\kappa/2}\sqrt{\widehat{Var}(\hat{\varphi}_i)},$$

where $z_{\kappa/2}$ is the $\kappa/2$ percentile of the standard normal distribution and $\widehat{Var}(\hat{\varphi}_i)$ is obtained from the diagonal elements of the Fisher information matrix.

The observed information matrix is asymptotically equivalent to the expected information matrix (Migon & Gamerman, 1999). This result seems to be corroborated for SSM through simulation in Cavanaugh & Shumway (1996). Approximation of the expected information matrix by the observed information matrix is relatively common and suggested in many texts like Cavanaugh & Shumway (1996) and Sallas & Harville (1988), mainly for large samples.

Asymptotic confidence intervals can present border problems, that is, the interval limits can overtake the borders of the parametric space. In these cases, the Delta method (Casella & Berger, 2002) could be used to solve the problem by appropriately transforming the components of $\boldsymbol{\varphi}$ to the real line.

Inference for the latent variables can be made in a number of forms. The simplest is to replace $\boldsymbol{\varphi}$ by its MLE estimator in the calculations of Sections 2 and 3. The main advantage of this plug-in procedure is its simplicity. It also provides adequate mean estimates but has the drawback of underestimating the uncertainty associated with the estimation of $\boldsymbol{\varphi}$. Alternatives taking into account this uncertainty may be available via bootstrap procedures. Similar comments are valid for the predictive distributions.

4.2 Bayesian Inference

Bayesian inference for $\boldsymbol{\varphi}$ can be performed using MCMC algorithms (Gamerman & Lopes, 2006), since the posterior distribution of the hyperparameter is not analytically tractable. The marginal posterior distribution of parameter vector $\boldsymbol{\varphi}$ is given by

$$\pi(\boldsymbol{\varphi}|\boldsymbol{Y}_n) \propto L(\boldsymbol{\varphi}; \boldsymbol{Y}_n)\pi(\boldsymbol{\varphi}), \tag{14}$$

where $L(\boldsymbol{\varphi}; \boldsymbol{Y}_n)$ is the likelihood function defined in (12) and $\pi(\boldsymbol{\varphi})$ is the prior distribution of $\boldsymbol{\varphi}$. In this work, proper uniform priors are used for $\boldsymbol{\varphi}$.

Credibility intervals for φ_i, $i = 1, ..., p$ are built as follows. Given a value $0 < \kappa < 1$, the interval $[c_1, c_2]$ satisfying

$$\int_{c_1}^{c_2} \pi(\varphi_i \mid \boldsymbol{Y}_n)\, d\varphi_i = 1 - \kappa$$

is a credibility interval for φ_i with level $100(1-\kappa)\%$.

Inference for the latent variables can be made with the output from the MCMC algorithm. Once a sample $\boldsymbol{\varphi}^{(1)}, ..., \boldsymbol{\varphi}^{(M)}$ is available, posterior samples $\lambda^{(1)}, ..., \lambda^{(M)}$ from the latent variables are obtained as follows:

1. set $j = 1$;
2. sample the hyperparameter $\boldsymbol{\varphi}^{(j)}$ from the MCMC algorithm;
3. sample the set $\lambda^{(j)}$ of latent variables from $p(\lambda|\boldsymbol{\varphi}^{(j)}, \boldsymbol{Y}_n)$ using Theorem 2;
4. set $j \to j + 1$ and return to 1, if $j \leq M$; otherwise, stop.

Again, similar comments are valid for the predictive distributions. Note that

$$p(y_{t+h}|\boldsymbol{Y}_t) = \int p(y_{t+h}|\boldsymbol{Y}_t, \boldsymbol{\varphi})\, \pi(\boldsymbol{\varphi}|\boldsymbol{Y}_n)\, d\boldsymbol{\varphi}. \tag{15}$$

Thus, h-step-ahead predictive distributions can be approximated by

$$\frac{1}{M} \sum_{j=1}^{M} p(y_{t+h}|\boldsymbol{Y}_t, \boldsymbol{\varphi}^{(j)})$$

from which summaries such as means, variances and credibility intervals can be obtained. Since $p(y_{t+h}|\boldsymbol{Y}_t, \boldsymbol{\varphi})$ is not available analytically, a draw $y_{t+h}^{(s)}$ from $p(y_{t+h}|\boldsymbol{Y}_t)$ can be obtained from (15) by sampling $\boldsymbol{\varphi}^{(s)}$ from $\pi(\boldsymbol{\varphi}|\boldsymbol{Y}_n)$ and then sampling $y_{t+h}^{(s)}$ from $p(y_{t+h}|\boldsymbol{Y}_t, \boldsymbol{\varphi}^{(s)})$.

4.3 Model adequacy

Model adequacy is an important topic of the modeling process. Some of the diagnostic methods suggested in the literature are described below.

Harvey & Fernandes (1989) suggest diagnostic methods based on the (standardized) Pearson residuals, given by

$$\nu_t^p = \frac{y_t - E(y_t|\boldsymbol{Y}_{t-1}, \boldsymbol{\varphi})}{DP(y_t|\boldsymbol{Y}_{t-1}, \boldsymbol{\varphi})}, \tag{16}$$

where $DP(y_t|\boldsymbol{Y}_{t-1}, \boldsymbol{\varphi})$ is the standard deviation of the distribution of $y_t|\boldsymbol{Y}_{t-1}, \boldsymbol{\varphi}$.

The diagnostic methods referred by them are:

1. examination of residual graphs versus time and versus an estimate of the level component.

2. check whether the sample variance of the standardized residuals is close to 1. A value greater (smaller) than 1 indicates overdispersion (underdispersion) of the model.

Another alternative is to use the deviance residuals (McCullagh & Nelder, 1989), which are defined by:

$$\nu_t^d = \left\{ 2 \ln \left[\frac{p(y_t|y_t, \boldsymbol{\varphi})}{p(y_t|\hat{\gamma}_t, \boldsymbol{\varphi})} \right] \right\}^{1/2}, \tag{17}$$

where $\hat{\gamma}_t = E(y_t|\boldsymbol{Y}_{t-1}, \boldsymbol{\varphi})$.

For example, in the Poisson model, the deviance residual is

$$\begin{aligned} \nu_t^d &= \left\{ 2 \ln \left[\frac{y_t^{y_t} \exp(-y_t)/y_t!}{\hat{\gamma}_t{}^{y_t} \exp(-\hat{\gamma}_t)/y_t!} \right] \right\}^{1/2} \\ &= \left\{ 2 \left[y_t \ln(y_t/\hat{\gamma}_t) - (y_t - \hat{\gamma}_t) \right] \right\}^{1/2}, \end{aligned}$$

where $\hat{\gamma}_t = a_{t+1|t}/b_{t+1|t}$. Diagnostic plots, like residuals versus order and fitted values, may also be used.

Sometimes, it is possible to have more than one model for modeling the data, but it is required to choose one of them. Therefore, some fit criteria are needed for deciding which model to choose.

According to Harvey (1989, page 80), the AIC and BIC criterion are the most used in practice. They are given by

$$AIC = -2\frac{\log\ L(\hat{\boldsymbol{\varphi}}; \boldsymbol{Y}_n)}{n} + \frac{2p}{n} \text{ and } BIC = -2\frac{\log\ L(\hat{\boldsymbol{\varphi}}; \boldsymbol{Y}_n)}{n} + \frac{p \ln(n)}{n},$$

where p is the number of parameters and n the number of observations.

The DIC criterion (*Deviance Information Criterion*) (Spiegelhalter *et al.* , 2002) can also be used for comparing models under the Bayesian approach. It is also based on the log likelihood with a penalization term to account for model complexity and can be approximated with an MCMC sample from $p(\boldsymbol{\varphi}|\boldsymbol{Y}_n)$.

5 Dynamic linear models revisited

The results of this paper are also useful to a closely related and very frequently used time series model. One of the most important classes of state-space models is formed by the dynamic linear models proposed by West & Harrison (1997) for Gaussian observations with time varying means and variances. Their model is easily extended by consideration of non-Gaussian distributions for system and observation disturbances

via scale mixtures of normal distributions as

$$
\begin{aligned}
y_t &= F_t x_t + v_t, \text{ where } v_t|\gamma_t \sim N(0, \gamma_t \lambda_t^{-1}) && (18)\\
x_{t+1} &= G_t x_t + w_{t+1}, \text{ where } w_{t+1}|\delta_t, \boldsymbol{\varphi} \sim N(0, \delta_t W) && (19)\\
\lambda_{t+1} &= w^{-1}\lambda_t \varsigma_{t+1}, \text{ where } \varsigma_{t+1}|\boldsymbol{Y}_t, \boldsymbol{\varphi} \sim Beta\,(wa_t, (1-w)a_t) && (20)\\
x_0|Y_0 &\sim N(m_0, C_0) \text{ independent of } \lambda_0|Y_0 \sim Gamma(a_0, b_0). && (21)
\end{aligned}
$$

The model is completed with mixing densities f_γ and f_δ for independent scales γ_t's and δ_t's, respectively. The hyperparameter in the above model is $\boldsymbol{\varphi} = (w, W)$. Results in Appendix A show that the model is well defined.

This formulation gives rise to a large class of non-Gaussian distributions for the disturbances v_t and w_t, studied by West (1987). For example, marginal double exponential distributions are obtained for v_t and w_t by specification of exponential distributions for γ_t's and δ_t's. Similarly, t-Student, logistic, exponential power and stable distributions are obtained by suitable choices of mixing distributions. Details can be found in West (1987).

The above formulation encompasses many models previously proposed in the literature. The dynamic linear models of West & Harrison (1997) are obtained by fixing mixing scales to 1. The non-Gaussian linear models of Carlin, Polson and Stoffer (1992) are obtained by setting $\varsigma_{t+1} = w = 1$, for all t, ie, the observational variances are fixed over time and are not subject to stochastic variation.

The different temporal dependencies in the stochastic specifications of the γ_t's and λ_t's ensure their identification, except for an arbitrary constant c as $\gamma_t\lambda_t^{-1} = (c\gamma_t)(c\lambda_t)^{-1}$. This causes no concern for the identification of the temporal variation of λ_t's, which is their most relevant feature. Also, the magnitudes of the observational variances are always identified and the magnitudes of the λ_t's may be identified if required with the prior in (21).

The importance of these models was not overlooked in the literature; there is no reason to believe that only the means vary stochastically in time series. West & Harrison (1997) discuss this model in detail and set it as a default model in their BATS software (West, Harrison & Pole, 1987), after fixing w at a suitably large value close to 1. However, they had to resort to different sets of approximations to obtain filtering and smoothing procedures.

An alternative formulation is provided by the use of stochastic volatility models as in Jacquier, Polson and Rossi (1994). Virtually exact results are obtained via MCMC procedures. These procedures are based on the use of Metropolis-Hastings algorithms. These are either not straightforward to tune, specially due to the strong correlation exhibited by the λ_t's, or require yet another approximation of the resulting Gamma densities by a finite mixture of normals (Kim, Shephard and Chib, 1998) to run Gibbs steps.

None of these problems arise here since all full conditional distributions are available for sampling in closed form with the results above. The highly correlated mean components $x_{0:n}$ are jointly sampled with standard FFBS and the highly correlated variance components $\lambda_{0:n}$ are jointly sampled with the smoothing procedures of Section 3, with $a_t = a_{t|t-1} + 1/2$ and $b_t = b_{t|t-1} + \gamma_t^{-1}(y_t - F_t x_t)^2/2$. Thus, convergence problems are mitigated by the good mixing properties of the MCMC chains.

The mixing components γ_t's and δ_t's are jointly sampled from their componentwise independent full conditionals. Carlin, Polson and Stoffer (1992) provide the details for a few mixing distributions. For example, assume t-Student distributions are required for observational and/or system disturbances and denote by $IG(a, b)$ the inverse Gamma distribution with density proportional to $z^{-(a+1)}e^{-b/z}$, for $z > 0$. If the conditionally normal distributions for v_t's and w_t's in (18) and (19) are respectively mixed with $IG(\nu_\gamma/2, \nu_\gamma S_\gamma/2)$ and $IG(\nu_\delta/2, \nu_\delta S_\delta/2)$ distributions for the γ_t's and δ_t's then $t_{\nu_\gamma}(0, S_\gamma \lambda_t^{-1})$ and $t_{\nu_\delta}(0, S_\delta W)$ are obtained marginally for v_t's and w_t's. The full conditional distributions for γ_t and δ_t are $IG(\nu_\gamma^*/2, \nu_\gamma^* S_\gamma^*/2)$ and $IG(\nu_\delta^*/2, \nu_\delta^* S_\delta^*/2)$ distributions where $\nu_\gamma^* = \nu_\gamma + 1$, $\nu_\delta^* = \nu_\delta^* + p$, $\nu_\gamma^* S_\gamma^* = \nu_\gamma S_\gamma + \lambda_t(y_t - F_t x_t)^2$ and $\nu_\delta^* S_\delta^* = \nu_\delta S_\delta + (x_t - G_t x_{t-1})^T W^{-1}(x_t - G_t x_{t-1})$, where p is the dimension of x_t. These distributions are easily sampled from.

When W is unknown with inverse Wishart $IW(\nu_W/2, \nu_W S_W/2)$ prior, its full conditional distribution is also inverse Wishart due to conjugacy and thus easily sampled from with results of Gamerman & Lopes (2006, p. 67). The availability of these easily coded procedures makes this model a potential candidate for automated yet fairly flexible implementation of non-Gaussian time series analysis with time-varying means and variances.

6 Application to real time series

In this subsection, the models are applied to real time series, using both Bayesian and classical inference. In general, vague but proper priors are assumed for the hyperparameters, either as uniform with suitably large limits or Gamma with small parameter values. For implementation of the Bayesian inference, the Metropolis-Hastings algorithm is used with generation of two chains. The Coda package in R is used for diagnostic methods, checking the chains convergence through graphic methods such as the autocorrelogram, time series and trace plots. The MLE is calculated through the well known Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm (Nocedal & Wright, 2006), and the confidence intervals are built using numerical derivatives to obtain the Fisher information matrix, as in Franco *et al.* (2008).

6.1 Respiratory disease data

The non-Gaussian state space model (NGSSM) with Poisson observations is used to fit the daily data of the number of patients with respiratory disease (RD) in São Paulo, Brazil, from 12/02/1999 to 12/31/2000 (396 observations). The last 10 observations are excluded from the fit with the purpose of comparing the forecasts, and thus $n = 386$. Time series of sulfur dioxide SO_2 (x_{1t}) and carbon monoxide CO (x_{2t}), $t = 1, \ldots, n$, were considered as covariates since exposure to these air pollutants can increase the risk of respiratory illnesses ((Ferris *et al.*, 1983); (Alves, Gamerman & Ferreira, 2010)). These series were chosen because, usually, high concentrations of CO, produced by vehicles, are found in urban areas and the fuels based on the sulfur (for instance, oils) in combustion produce SO_2. The RD series and pollutants SO_2 and CO were obtained, respectively, by the Brazilian Health Ministry site (http://www.datasus.gov.br) and the

Technology and Environmental Company from São Paulo (CETESB). This application illustrates the use of the smoothing procedure introduced in this work, exemplifies the residual analysis and exhibits results of h steps ahead forecasts, for $h \geq 1$, using Equation (7).

For comparison purposes, the data was also fitted with the normal disturbances Poisson SSM (NDPSSM) of Durbin & Koopman (2001) using the package `sspir` (Dethlefsen & Lundbye-Christensen, 2006) in `R`. It uses the importance sampling technique for making approximate inference. Their model only differs by assuming that the system disturbances ς_t^* follow a Gaussian rather than a scaled log-Beta distribution. The forecasting is done according to Durbin & Koopman (2001, p. 214-215) and the smoothed estimates of the state components are obtained using the *iterated extended Kalman filter* (Durbin & Koopman, 2001).

Table 2 presents a summary of the estimation results. The estimates of w are around 0.78 and the estimated regression coefficients of the NDPSSM and the NGSSM are very similar. The regression coefficients of the pollutants are positive, as expected.

Table 2: Point and interval estimation for the Poisson model under the NDPSSM and NGSSM approaches fitted to the RD series.

	NGSSM					NDPSSM	
φ	**MLE**	**Conf. Int.**	**BE-Median**	**BE-Mean**	**Cred. Int.**	**MLE**	**Conf. Int.**
w	0.784	[0.735; 0.833]	0.778	0.775	[0.721; 0.821]	-	-
β_1	0.005	[0.001; 0.008]	0.005	0.005	[0.001; 0.008]	0.006	[0.002; 0.010]
β_2	0.017	[-0.003; 0.038]	0.017	0.017	[-0.003; 0.038]	0.010	[-0.014; 0.032]

The values of the Gelman and Rubin's criterion (Gelman, 1996) for the two chains of parameters w, β_1 and β_2 are around 1.00 and the traces overlap over the entire trajectory, indicating convergence of the chains. The model fit does not seem to present any inadequacies, as can be seen from the Pearson and deviance residual analysis of Figure 1.

Figure 2 shows the RD series with its estimated smoothed means. The estimates of the mean seem to follow better the behavior of the series with the NGSSM. Summary measures based on fit and forecast errors are comparable between the two models, with a slight preference for the NGSSM. This improvement is probably due to better adaptation to changes by the log-scaled beta disturbances. The figure also depicts the forecasts for future observations, based on a future scenario for the covariates and on the results of Theorem 2. The forecasts seem to follow the behaviour of the smoothed mean values estimated at the end of the series, and the interval widths present an increase with the forecast horizon, as expected.

6.2 CEMIG return data

The second example refers to the daily data of CEMIG stock returns in the period from 01/03/2005 to 06/08/2011 (1590 observations) and presents several new models written in the NGSSM form, which do not belong to exponential family, like the Generalized

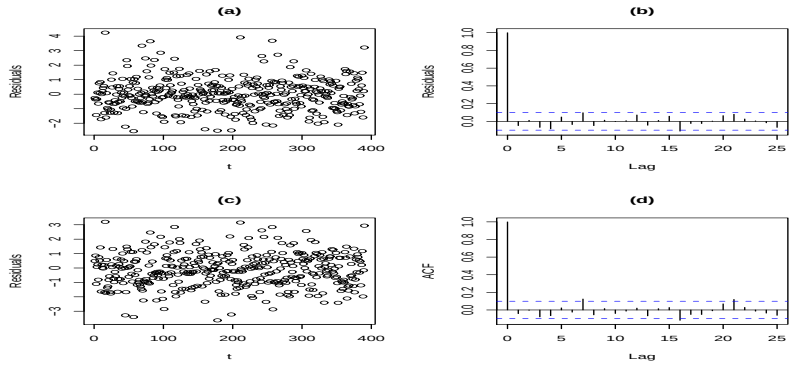

Figure 1: Pearson and deviance residual analysis for the Poisson model. (a): time series of the Pearson residuals; (b): autocorrelation plot of the Pearson residuals; (c): time series of the deviance residuals; (d): autocorrelation plot of the deviance residuals.

gamma, Pareto with unknown scale parameter and Laplace with unknown mean models. Here the return at time t is defined as $y_t = \ln\left(\frac{P_t}{P_{t-1}}\right)$, where P_t is the daily closing spot price. The data irregularity due to holidays and weekends will be ignored. Figure 3 presents the time series plot of the CEMIG returns. A distinctive feature of financial series is that they usually present non-constant variance or volatility. This application illustrates the comparison of the fit of several observational distributions. For the analysis of the CEMIG series, a temporal correlation structure is assumed for the variance and two approaches are considered. In the first one, the Normal and Laplace models are fitted to the CEMIG returns y_t. In the second, the square transformation y_t^2 is used, so that Generalized Gamma, Gamma, Pareto and Inverse Gaussian models (with support as the positive line) listed in Table 1 can be fitted to the series.

The Inverse Gaussian model was suggested by Barndorff-Nielsen & Shephard (2001) for modeling volatility series. Lopes & Migon (2002) used the Generalized gamma model for modelling volatility series and Achar & Bolfarine (1986) utilized this log-linear model in survival analysis. The pareto model is well-known in the literature for modelling extreme events.

Table 3 shows the log-likelihood, AIC and BIC for the classical fits, and the DIC criterion for the Bayesian fits. For a fair comparison between the models, the log-likelihood in the second approach is corrected by the Jacobian of the transformation. The results show that the fit criteria discard the Inverse Gaussian model. The remaining models are somewhat comparable with a slight preference for the Generalized gamma model.

Table 4 shows the point and interval estimates for the parameters of the Generalized gamma model. It can be seen that classical and Bayesian estimates are very similar. The residual analysis did not present any indication of model inadequacy (see Figure 4).

For illustration, Figure 5 shows the graph of the volatility obtained by the General-

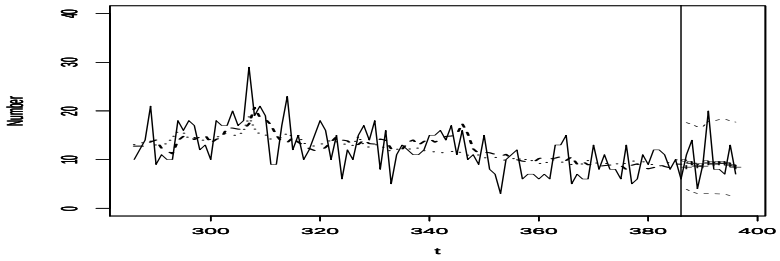

Figure 2: The full, dashed and dotted lines represent, respectively, the last 100 observations of RD time series, the smoothed mean of the Bayesian estimation obtained through the NGSSM approach and the smoothed mean obtained through the DGLM approach. The vertical line separates the fitted data points from the forecast horizon. NGSSM mean: o; NDPSSM mean: +; NGSSM 95% interval limits: $- - -$.

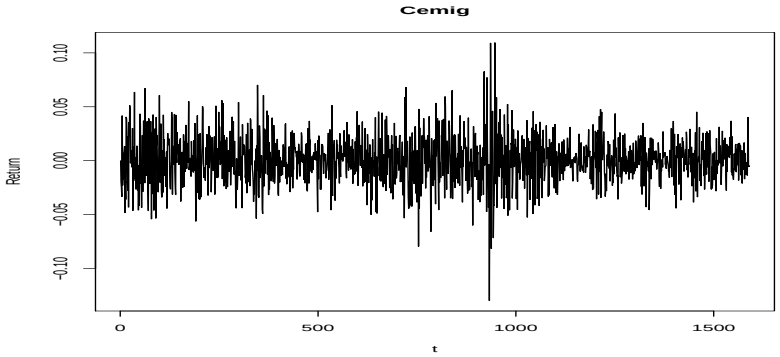

Figure 3: CEMIG return series.

ized gamma model under the Bayesian approach. The peaks, corresponding to periods of crisis known in the literature, can be observed in the figure, and are pointed out by the model.

6.3 The Candy data

This section illustrates the application of dynamic linear models with non-Gaussian disturbances and time-varying means and variances. Posterior samples of the static and dynamic parameters are easily obtained for the model below using the approach of Section 5. The Candy data, presented in Figure 6, is a well-known time series introduced in West & Harrison (1997, p. 331) and used for illustration of many features in that book. They consist on the monetary values of monthly total sales, on a standardized, deflated scale, of a widely consumed and established food product in the UK market. The series is composed of 72 monthly observations in the period from

Table 3: Values of log-likelihood, AIC, BIC and DIC for the models fitted to the CEMIG return.

Models	**log-likelihood**	**AIC**$^{(c)}$	**BIC**$^{(c)}$	**DIC**$^{(b)}$
Generalized Gamma	3929.03	-4.94	-4.93	-7852.10
Gamma	3928.00	-4.94	-4.93	-7852.00
Normal	3926.00	-4.93	-4.93	-7850.00
Laplace	3863.74	-4.86	-4.85	-7725.80
Pareto	3639.53	-4.86	-4.85	-6517.00
Inverse Gaussian	2629.25	-3.31	-3.30	-5256.25

Note: (c) classical fit. (b) Bayesian fit.

Table 4: Point and interval (level 95%) estimation for the Generalized gamma model fitted to the CEMIG series.

φ	**MLE**	**Conf. Int.**	**Posterior mean**	**Cred. Int.**
w	0.941	[0.930; 0.960]	0.940	[0.920; 0.957]
χ	0.397	[0.310; 0.484]	0.407	[0.332; 0.506]
ν	1.135	[0.948; 1.321]	1.124	[0.952; 1.300]

01/1976 to 12/1981. The Index covariate is included in the analysis for explaining partial movements of the series. This index is built based on market prices, distribution and production costs.

As illustration, the model defined in Section 5 is fitted to the series, assuming $F_t = \begin{pmatrix} 1 & x_t \end{pmatrix}$, $G_t = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $x_t = \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix}$, $w_t = \begin{pmatrix} \eta_t \\ \xi_t \end{pmatrix}$ and $W = \begin{pmatrix} \sigma_\eta^2 & 0 \\ 0 & \sigma_\xi^2 \end{pmatrix}$. This model is a local level model for time-varying intercept and regression coefficient.

G(1.5,1.5) distributions are specified for γ_t's and δ_t's in order to obtain the t_3-Student errors for the observation and system disturbances. The number of degrees of freedom ν is known, but can also be estimated (Fonseca, Ferreira & Migon, 2008). The parameter w is fixed at 0.98 and initial values are $m_0 = \begin{pmatrix} 9.5 \\ -0.7 \end{pmatrix}$, $P_0 = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix}$, $a_0 = 0.1$ and $b_0 = 0.1$, as suggested in West, Harrison & Pole (1987, p. 333).

Figure 6 presents the results of the estimation, where the smoothed mean responses follow well the behavior of the series.

Figure 7 shows smoothed estimates of the non-observable components fitted to the series. The level component μ_t oscillates in time in the early part of the series to compensate seasonal effects while β_t decreases over time and λ_t^{-1} shows a discrete decline.

Figure 8 summarizes estimation of hyperparameters and the time-varying scales. The posterior mean of the variance parameters are 1.11 and 1.10, respectively and their distributions are slightly asymmetric, as expected.

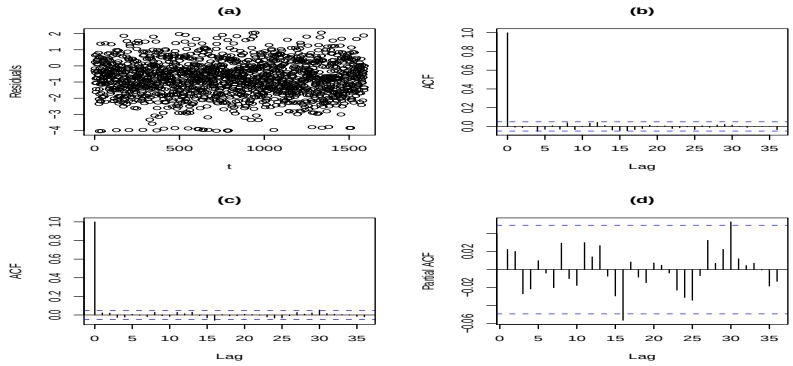

Figure 4: Deviance residual analysis for the Gamma model. (a): time series of the deviance residuals; (b): autocorrelation plot of the deviance residuals; (c): autocorrelation plot of the square of the deviance residuals; (d): partial autocorrelation plot of the square of the deviance residuals.

7 Discussion

Several possibilities of related future work can be thought of. Studies exploring properties of the family of observational models in **A0** and of special cases in this class can be made. For example, the piecewise exponential distribution (Gamerman , 1994), which has a wide field of applications in reliability and survival analysis, may be adapted to these non-Gaussian models. More general evolution equations with stationary and non-stationary autoregressive structures may be proposed. Hypothesis tests can be explored, as well as the use and application of the bootstrap technique for making inference about the parameters. Extension towards non-Gaussian multivariate time series along the lines set by Uhlig (1994) could also be entertained.

The existence of easily available procedures for inference is an attractive feature of these models. This is particularly relevant for use in time-varying components of larger, more complex models. These components are typically latent and precise quantitative assessments for them are usually not available. Qualitative assessments are the basis for their probabilistic specifications. Availability of the mathematically tractable options may play an important role in the decision, when facing options that are equally acceptable. This point was illustrated in Section 5.

The models are applicable to situations with fixed effects of the covariates. An important research development is the extension towards time-varying regression coefficients as considered in dynamic generalized linear models, but without losing analytic tractability.

Acknowledgements

D. Gamerman was supported by CNPq-Brazil and *Fundação de Amparo à Pesquisa no Estado do Rio de Janeiro* (FAPERJ foundation). T. R. Santos was supported by CNPq-

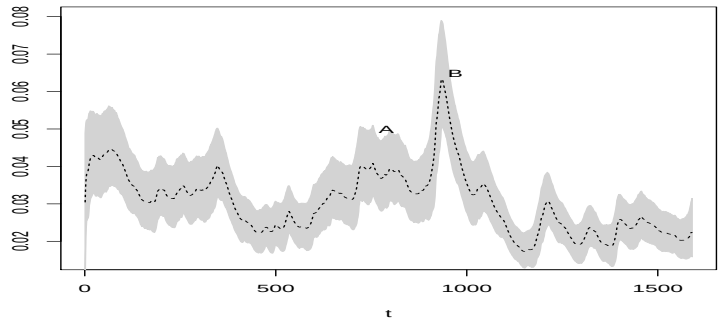

Figure 5: The CEMIG data: The dashed line represents the smoothed estimate of the stochastic volatility, obtained by the fit of the Generalized gamma model under the Bayesian approach. The grey area indicates the 95% credibility intervals. A represents USA's real estate crisis in August 2007, B represents L-B's crisis in September 2008.

Brazil and CAPES-Brazil. G. C. Franco was supported by CNPq-Brazil and *Fundação de Amparo à Pesquisa no Estado de Minas Gerais* (FAPEMIG foundation).

Appendix A

This appendix is concerned with asserting that models of this paper are well defined. In what follows, $z_{l:m} = (z_l, z_{l+1}, \ldots, z_{m-1}, z_m)$, for integers $l < m$, and hyperparameters $\boldsymbol{\varphi}$ are removed from the notation for clarity.

Model A0-A3:

The joint distribution of all model quantities is given by

$$\begin{aligned} p(y_{1:n}, \lambda_{0:n}) &= \prod_{t=1}^{n} p(y_t, \lambda_t | \lambda_{0:t-1}, \boldsymbol{Y}_{t-1}) p(\lambda_0 | Y_0) \\ &= \prod_{t=1}^{n} p(y_t | \lambda_{0:t}, \boldsymbol{Y}_{t-1}) p(\lambda_t | \lambda_{t-1}, \boldsymbol{Y}_{t-1}) p(\lambda_0 | Y_0). \end{aligned}$$

The three densities of the product in the RHS are uniquely defined by **A0**, **A2** and **A3**, respectively. Thus, the basic model **A0**-**A3** is uniquely defined.

Model (18)-(21):

The joint distribution of all model quantities is given by

$$f(y_{1:n}, x_{0:n}, \lambda_{0:n}, \gamma_{1:n}, \delta_{1:n} | Y_0) = f(y_{1:n}, x_{0:n}, \lambda_{0:n} | \gamma_{1:n}, \delta_{1:n}, Y_0) \prod_{t=1}^{n} f_\gamma(\gamma_t) f_\delta(\delta_t). \quad (22)$$

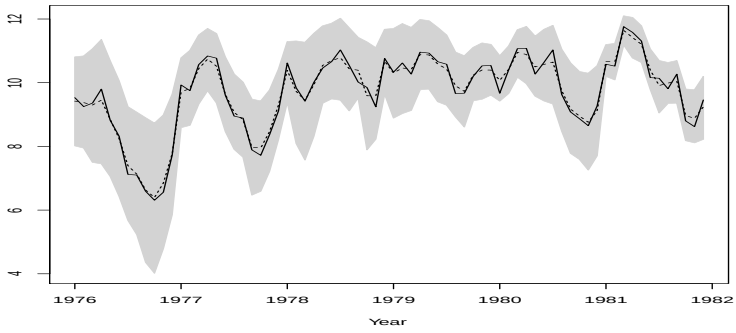

Figure 6: The full and dashed lines represent the series and the smoothed mean estimate, obtained by the fit under the Bayesian approach. The grey area indicates the 95% credibility intervals.

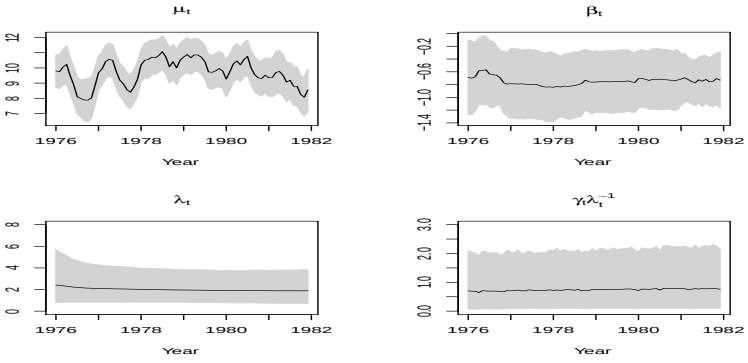

Figure 7: The smoothed estimates of dynamic components. Top left: smoothed estimate of the level (μ_t); Top right: smoothed estimate of the regression coefficient (β_t); Bottom left: smoothed estimate of the precision (λ_t); Bottom right: smoothed estimate of the variance ($\gamma_t \lambda_t^{-1}$).

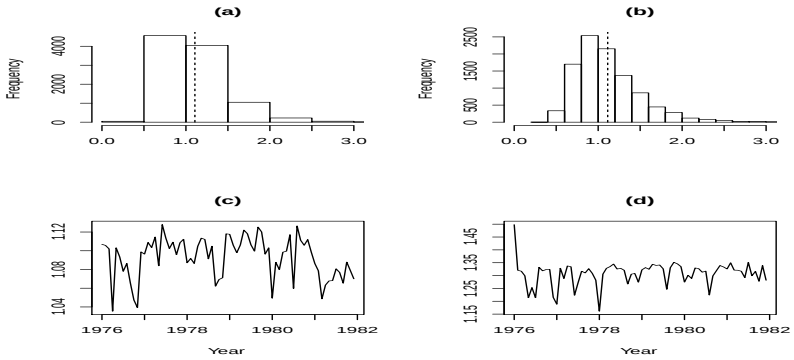

Figure 8: Posterior estimates: (a) - posterior histogram of σ_η^2; (b) - posterior histogram of σ_ξ^2; (c) - smoothed means of γ_t; (d) - smoothed means of δ_t. Vertical dashed lines indicate the posterior means.

But $f(y_{1:n}, x_{0:n}, \lambda_{0:n}|\gamma_{1:n}, \delta_{1:n}, Y_0) =$

$$\prod_{t=1}^{n} f(y_t, x_t, \lambda_t|\boldsymbol{Y}_{t-1}, x_{0:t-1}, \lambda_{0:t-1}, \gamma_{1:n}, \delta_{1:n}) f(x_0|Y_0) f(\lambda_0|Y_0) \tag{23}$$

and the joint densities $f(y_t, x_t, \lambda_t|\boldsymbol{Y}_{t-1}, x_{0:t-1}, \lambda_{0:t-1}, \gamma_{1:n}, \delta_{1:n})$ can be decomposed as

$$\begin{aligned} f(y_t|\boldsymbol{Y}_{t-1}, x_{0:t}, \lambda_{0:t}) f(\lambda_t|\lambda_{0:t-1}, \boldsymbol{Y}_{t-1}, x_{0:t}) f(x_t|x_{0:t-1}, \boldsymbol{Y}_{t-1}) = \\ f(y_t|\boldsymbol{Y}_{t-1}, x_t, \lambda_t) f(\lambda_t|\lambda_{t-1}, \boldsymbol{Y}_{t-1}, x_{0:t}) f(x_t|x_{t-1}, \boldsymbol{Y}_{t-1}), \end{aligned} \tag{24}$$

where the dependence on $(\gamma_{1:n}, \delta_{1:n})$ was removed from all terms in (24) for conciseness. The densities in the RHS of (24) correspond to model specifications (18), (19) and (20), respectively, whereas the prior distributions (21) for the latent state components are provided in (23) and the mixing distributions are provided in (22). So, each model component in the distribution of $(y_{1:n}, x_{0:n}, \lambda_{0:n}, \gamma_{1:n}, \delta_{1:n}|Y_0)$ is uniquely defined.

Appendix B

This appendix presents the proof of the theorems provided in the text. For ease of notation, hyperparameter vector $\boldsymbol{\varphi}$ will be omitted from the proofs.

Proof of Theorem 1.

In what follows, proofs of Results 1 to 3 from Theorem 1 are provided. In order to prove the theorem, the following lemma is needed.

Lemma. $\lambda_t|\lambda_{t-1}, \mathrm{Y}_{t-1}$ has density

$$p(\lambda_t|\lambda_{t-1}, \mathrm{Y}_{t-1}) = \begin{cases} k\ \frac{w}{\lambda_{t-1}} \left(\frac{w\lambda_t}{\lambda_{t-1}}\right)^{wa_{t-1}-1} \left(1 - \frac{w\lambda_t}{\lambda_{t-1}}\right)^{(1-w)a_{t-1}-1} \\ \quad \text{if} \quad 0 < \lambda_t < w^{-1}\lambda_{t-1}, \\ 0, \quad \text{otherwise.} \end{cases}, \quad (25)$$

where $k = \dfrac{\Gamma(wa_{t-1})\Gamma((1-w)a_{t-1})}{\Gamma(a_{t-1})}$.

Proof: The proof is easily obtained by using the Jacobian method and assumption **A2** in Section 2 for $h = 0$.

Proof of Result (1) in Theorem 1:
Assume by the induction hypothesis that $\lambda_{t-1}|\boldsymbol{Y}_{t-1} \sim Gamma\,(a_{t-1}, b_{t-1})$ and the distribution of $\lambda_t = w^{-1}\lambda_{t-1}\varsigma_t$ is given in Equation (25). Then,

$$\begin{aligned} p(\lambda_t|\boldsymbol{Y}_{t-1}) &= \int p(\lambda_{t-1}|\boldsymbol{Y}_{t-1}) p(\lambda_t|\lambda_{t-1}, \boldsymbol{Y}_{t-1}) d\lambda_{t-1} \\ &= \int_{w\lambda_t}^{\infty} \left[\frac{\lambda_{t-1}^{a_{t-1}-1} \exp(-b_{t-1}\lambda_{t-1})}{\Gamma(a_{t-1}) b_{t-1}^{-a_{t-1}}}\right] \left[\frac{w\lambda_{t-1}^{-1}(\frac{w\lambda_t}{\lambda_{t-1}})^{wa_{t-1}-1}(1 - \frac{w\lambda_t}{\lambda_{t-1}})^{(1-w)a_{t-1}-1}}{\frac{\Gamma(wa_{t-1})\Gamma((1-w)a_{t-1})}{\Gamma(a_{t-1})}}\right] d_{\lambda_{t-1}} \\ &\propto \int_{w\lambda_t}^{\infty} \left[\lambda_{t-1}^{a_{t-1}-1-wa_{t-1}+1-1} \exp(-b_{t-1}\lambda_{t-1})\right] \left[(1 - \frac{w\lambda_t}{\lambda_{t-1}})^{(1-w)a_{t-1}-1}\right] d_{\lambda_{t-1}} \\ &\propto \int_{w\lambda_t}^{\infty} \left[\lambda_{t-1}^{(1-w)a_{t-1}-1} \exp(-b_{t-1}\lambda_{t-1})\right] \left[(1 - \frac{w\lambda_t}{\lambda_{t-1}})^{(1-w)a_{t-1}-1}\right] d_{\lambda_{t-1}} \\ &\propto \int_{w\lambda_t}^{\infty} \exp(-b_{t-1}\lambda_{t-1})(\lambda_{t-1} - w\lambda_t)^{(1-w)a_{t-1}-1} d_{\lambda_{t-1}}. \\ &= \frac{\lambda_t^{wa_{t-1}-1} \exp(-wb_{t-1}\lambda_t)}{(wb_{t-1})^{-wa_{t-1}}\Gamma(wa_{t-1})}, \text{ for } \lambda_t > 0. \end{aligned}$$

This is the density of a $Gamma(wa_{t-1}, wb_{t-1})$ distribution, completing the proof of Result 1.

Proof of Result (2) in Theorem 1:
By Bayes theorem,
$p(\lambda_t|\boldsymbol{Y}_t, \boldsymbol{\varphi}) \propto p(y_t|\lambda_t, \boldsymbol{\varphi}) p(\lambda_t|\boldsymbol{Y}_{t-1}, \boldsymbol{\varphi}) \propto \lambda_t^{(a_{t|t-1}+b(y_t,\boldsymbol{\varphi}))-1} \exp[-\lambda_t(b_{t|t-1}+c(y_t, \boldsymbol{\varphi}))]$.

Then, it follows that $\lambda_t|\boldsymbol{Y}_t, \boldsymbol{\varphi} \sim Gamma\,(a_t, b_t)$, where $a_t = a_{t|t-1} + b(y_t, \boldsymbol{\varphi})$ and $b_t = b_{t|t-1}+c(y_t, \boldsymbol{\varphi})$, completing the induction. Since $\lambda_0|\boldsymbol{Y}_0, \boldsymbol{\varphi} \sim Gamma\,(a_0, b_0)$,

the proof is complete, for all t.

Proof of Result (3) in Theorem 1:

$$\begin{aligned} p(y_t|\boldsymbol{Y}_{t-1},\boldsymbol{\varphi}) &= \int_0^\infty p(y_t|\lambda_t,\boldsymbol{\varphi})p(\lambda_t|\boldsymbol{Y}_{t-1},\boldsymbol{\varphi})d\lambda_t \\ &= \frac{a(y_t)}{\Gamma(a_{t|t-1})(b_{t|t-1})^{-a_{t|t-1}}}\int_0^\infty \left[\lambda_t^{b(y_t)+a_{t|t-1}-1}\exp\left(-\lambda_t(c(y_t)+b_{t|t-1})\right)\right]d\lambda_t \\ &= \frac{\Gamma\left(b(y_t,\boldsymbol{\varphi})+a_{t|t-1}\right)a(y_t,\boldsymbol{\varphi})(b_{t|t-1})^{a_{t|t-1}}}{\Gamma(a_{t|t-1})\left(c(y_t,\boldsymbol{\varphi})+b_{t|t-1}\right)^{a_{t|t-1}+b(y_t,\boldsymbol{\varphi})}}, \end{aligned}$$

where $a_{t|t-1}=wa_{t-1}$, $b_{t|t-1}=wb_{t-1}$ and $y_t\in S(\boldsymbol{\varphi})$.

Proof that (10) implies (9)

Let $h=1$. Then, by Result 1 of Theorem 1, $\lambda_{t+1}|\boldsymbol{Y}_t\sim Gamma\left(a_{t+1|t},b_{t+1|t}\right)$. Assume by induction that $\lambda_{t+h}|\boldsymbol{Y}_t\sim Gamma\left(w^ha_t,w^hb_t\right)$. Combining the above evolution equation (10) with the induction hypothesis above leads, after integrating out λ_{t+h}, to $\lambda_{t+h+1}|\boldsymbol{Y}_t\sim Gamma\left(w(w^ha_t),w(w^hb_t)\right)=Gamma\left(w^{h+1}a_t,w^{h+1}b_t\right)$. Using the observation equation at time $t+h$ and the result above, the h-step-ahead predictive distribution can be found exactly as in Result 3 from Theorem 1 for $t+h$ with merely notational changes of $(y_{t+1},a_{t+1|t},b_{t+1|t})$ by $(y_{t+h},a_{t+h|t},b_{t+h|t})$.

Proof of Theorem 2:

The joint distribution of $(\lambda|\boldsymbol{Y}_n,\boldsymbol{\varphi})$ can always be written as

$$\begin{aligned} p(\lambda|\boldsymbol{Y}_n) &= p(\lambda_n|\boldsymbol{Y}_n)\prod_{t=1}^{n-1}p(\lambda_t|\lambda_{t+1},\ldots,\lambda_n,\boldsymbol{Y}_n) \\ &= p(\lambda_n|\boldsymbol{Y}_n)\prod_{t=1}^{n-1}p(\lambda_t|\lambda_{t+1},\boldsymbol{Y}_t), \end{aligned}$$

where the last equality follows from the Markovian structure of the model.

The required density of $[\lambda_t|\lambda_{t+1}, \boldsymbol{Y}_t]$ is given by

$$\begin{aligned} p(\lambda_{t-1}|\lambda_t, \boldsymbol{Y}_{t-1}) &= \frac{p(\lambda_t|\lambda_{t-1}, \boldsymbol{Y}_{t-1})p(\lambda_{t-1}|\boldsymbol{Y}_{t-1})}{p(\lambda_t|\boldsymbol{Y}_{t-1})} \\ &= \frac{\Gamma(wa_{t-1})\Gamma((1-w)a_{t-1})}{\Gamma(a_{t-1})}\frac{w}{\lambda_{t-1}}\left(\frac{w\lambda_t}{\lambda_{t-1}}\right)^{wa_{t-1}-1}\left(1-\frac{w\lambda_t}{\lambda_{t-1}}\right)^{(1-w)a_{t-1}-1} \times \\ &\quad \frac{b_{t-1}^{a_{t-1}}}{\Gamma(a_{t-1})}\lambda_{t-1}^{a_{t-1}-1}\exp(-\lambda_{t-1}b_{t-1})\frac{\frac{\Gamma(wa_{t-1})}{(wb_{t-1})^{wa_{t-1}}}}{\lambda_t^{wa_{t-1}-1}\exp(-\lambda_t w b_{t-1})} \\ &\propto (\lambda_{t-1}-w\lambda_t)^{(1-w)a_{t-1}-1}\exp\left(-b_{t-1}(\lambda_{t-1}-w\lambda_t)\right). \end{aligned}$$

Now, define $\eta_{t-1} = \lambda_{t-1} - w\lambda_t$ by removal of the constant $w\lambda_t$ from λ_{t-1}. Then, $p(\eta_{t-1}|\lambda_t, \boldsymbol{Y}_{t-1}) \propto \eta_{t-1}^{(1-w)a_{t-1}-1}\exp\left(-b_{t-1}\eta_{t-1}\right)$. It is clear that $\eta_{t-1} = \lambda_{t-1} - w\lambda_t|\lambda_t, \boldsymbol{Y}_{t-1} \sim$ Gama $((1-w)a_{t-1}, b_{t-1})$, completing the proof.

References

ACHAR, J. & BOLFARINE, H. (1986). The log-linear model with a generalized gamma distribution for the error: a Bayesian apporach. *Stat. Prob. Lett.* **4**, 325–332.

ALVES, M. B., GAMERMAN, D. & FERREIRA, M. A. R. (2010). Transfer functions in dynamic generalized linear models. *Stat. Model.* **10**, 3–40.

ARNOLD, B. C. & PRESS, S. J. (1989). Bayesian estimation and prediction for pareto data. *J. Amer. Statist. Assoc.* **84**, 1079–1084.

ANDRIEU, C. & DOUCET, A. (2002). Particle filtering for partially observed Gaussian state space models. *J. R. Statist. Soc.* B **64**, 827–836.

BARNDORFF-NIELSEN, O. E. & SHEPHARD, N. (2001). Non-Gaussian Ornstein-Uhlenbeck based models and some of their uses in financial economics (with discussion). *J. R. Statist. Soc.* B **63**, 167–241.

CARLIN, B.P., POLSON, N.G. & STOFFER, D.S. (1992). A Monte Carlo approach to nonnormal and nonlinear state-space modeling. *J. Amer. Statist. Assoc.* **87**, 493-500.

CARTER, C. K. & KOHN, R. (1994). On Gibbs sampling for state space models. *Biometrika* **81**, 541–553.

CARVALHO, C. M., JOHANNES, M. S., LOPES, H. F. & POLSON, N. G. (2010). Particle Learning and Smoothing. *Statist. Sci.* **25**, 88–106.

CASELLA, G. & BERGER, R. L. (2002). *Statistical Inference*. (2nd edition). Michigan: Thomson Learning.

CAVANAUGH, J. E. & SHUMWAY, R. H. (1996). On computing the expected Fisher information matrix for state-space model parameters. *Statist. Probab. Lett.* **26**, 347–355.

CHIOGNA, M. & GAETAN, C. (2002). Dynamic generalized linear models with application to environmental epidemiology. *Appl. Statist.* **51**, 453–468.

DETHLEFSEN, C. & LUNDBYE-CHRISTENSEN, S. (2006). Formulating state space models in R with focus on longitudinal regression models. *J. Statist. Software* **16**, 1–15.

DURBIN, J. & KOOPMAN, S. J. (2001). *Time Series Analysis by State Space Methods*. Oxford:Oxford University Press.

FAHRMEIR, L. & KAUFMANN, H. (1987). Regression models for non-stationary categorical time series. *J. Time Series Anal.* **8**, 147–160.

FAHRMEIR, L. & TUTZ, G. (2001). *Multivariate statistical modelling based on generalized linear models*. Springer: New York.

FERRIS, B. G., DOCKERY, D. W., WARE, J. H., SPEIZER, F. E. & SPIRO, R. (1983). The six-city study: examples of problems in analysis of the data. *Environmental Health Perspectives* **52**, 115–123.

FONSECA, T. C. O., FERREIRA, M. A. R. AND MIGON, H. S. (2008). Objective Bayesian analysis for the Student-t regression model. *Biometrika* **95**, 325–333.

FRANCO, G. C., SANTOS, T. R., RIBEIRO, J. A. & CRUZ, F. R. B. (2008). Confidence intervals for hyperparameters in structural models. *Comm. Statist. Simulation Comput.* **37**, 486–497.

FRÜHWIRTH-SCHNATTER, S. (1994). Applied state space modelling of non-Gaussian time series using integration-based Kalman filtering. *Stat. Comput.* **4**, 259–269.

FRÜHWIRTH-SCHNATTER, S. & WAGNER, H. (2006). Auxiliary mixture sampling for parameter-driven models of time series of counts with applications to state space modelling. *Biometrika* **93**, 827–841.

GAMERMAN, D. (1991). Dynamic Bayesian models for survival data. *Appl. Statist.* **40**, 63–79.

GAMERMAN, D. (1994). Bayes estimation of the piece-wise exponential distribution. *IEEE Trans. Reliab.* **43**, 128–131.

GAMERMAN, D. (1998). Markov chain Monte Carlo for dynamic generalised linear models. *Biometrika* **85**, 215–227.

GAMERMAN, D. & LOPES, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. (2nd edition). London: Chapman and Hall.

GELMAN, A. (1996). *Inference and monitoring convergence.* In *Markov Chain Monte Carlo in Practice*. W.R. Gilks, Richardson, S. Spiegelhalter, D. J. (eds). London: Chapman and Hall, pp. 131-143.

GODOLPHIN, E. J. & TRIANTAFYLLOPOULOS, K. (2006). Decomposition of time series models in state-space form. *Comput. Statist. Data Anal.* **50**, 2232–2246.

GRUNWALD, G. K., RAFTERY, A. E. & GUTTORP, P. (1993). Time series of continuous proportions. *J. R. Statist. Soc.* B **55**, 103–116.

HAIGHT, F. A. & BREUER, M. A. (1960). The Borel-Tanner Distribution. *Biometrika* **47**, 143–150.

HARVEY, A.C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

HARVEY, A. C. & FERNANDES, C. (1989). The time series models for count or qualitative observations. *J. Bus. Econom. Statist.* **7**, 407–417.

JACQUIER, E., POLSON, N. G. & ROSSI, P. E. (1994). Bayesian analysis of stochastic volatility models. *J. Bus. Econom. Statist.*, **12**, 371-389.

KIM, S., SHEPHARD, N. & CHIB, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Rev. Econom. Stud.*, **65**, 361-93.

LINDSEY, J. K. & LAMBERT, P. (1995). Dynamic generalized linear-models and repeated measurements. *J. Statist. Plann. Inference* **47**, 129–139.

LOPES, H. F. & MIGON, H. S. (2002). *Comovements and contagion in emergent markets: stock indexes volatilities.* In: *C. Gatsonis, R. Kass, A. Carriquiry, A. Gelman and D. Higdon (Eds.), Case Studies in Bayesian Statistics, Volume VI (p. 285-300). New York: Springer-Verlag.*

MIGON, H. S. & GAMERMAN, D. (1999). *Statistical Inference: An Integrated Approach*. London: Arnold.

MIGON, H. S., GAMERMAN, D., LOPES, H. F. & FERREIRA, M. A. R. (2005). *Dynamic models.* In: *Handbook of Statistics - Bayesian Thinking, Modeling and Computation*. Dipak K. Dey and C. R. Rao (Eds). Amsterdam: Elsevier, pp. 553-588.

MCCULLAGH, P. & NELDER. J. A. (1989). *Generalized Linear Models*. New York: Chapman and Hall.

NELDER. J. A. & WEDDERBURN, R. W. M. (1972). Generalized linear models. *J. R. Statist. Soc.* A **135**, 370–384.

NOCEDAL, J. & WRIGHT, S. J. (2006). *Numerical Optimization (2nd ed.)*. New York: Springer-Verlag.

PITT, M. K. & WALKER, S. G. (2005). Constructing stationary time series models using auxiliary variables with applications. *J. Amer. Statist. Assoc.* **100**, 554–564.

SALLAS, W. M. & HARVILLE, D. A. (1988). Noninformative priors and restricted maximum likelihood estimation in the Kalman filter. *Bayesian Analysis of Time Series and Dynamic Models*, J. C. Spall (Ed). New York:Marcel Dekker Inc.

SHEPHARD, N. (1994). Local scale models: state space alternative to integrated GARCH processes. *J. Econometrics* **60**, 181–202.

SHEPHARD, N. & PITT, M. K. (1997). Likelihood analysis of non-Gaussian measurement time series *Biometrika* **84**, 653–667.

SILVA, R. S., LOPES, H. F. & MIGON, H. S. (2006). The extended generalized inverse Gaussian distribution for log-linear and volatility models. *Braz. J. Prob. Stat.* B **20**, 67–91.

SMITH, R. L. & MILLER, J. E. (1986). A non-Gaussian state space model and application to prediction of records. *J. R. Statist. Soc.* B **48**, 79–88.

SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P. & VAN DER LINDE, A. (2002). Bayesian measures of model complexity and fit (with discussion). *J. R. Statist. Soc.* B **64**, 583–639.

UHLIG, H. (1994). On singular Wishart and singular multivariate beta distributions. *Ann. Statist.* **22**, 395-405.

WEST, M. (1987). On scale mixtures of normal distributions. *Biometrika*, **74**, 646-648.

WEST, M. & HARRISON, P. J. (1997). *Bayesian Forecasting and Dynamic Models*. New York: Springer.

WEST, M., HARRISON, P. J. & MIGON, H. S. (1985). Dynamic generalized linear models and Bayesian forecasting (with discussion) *J. Amer. Statist. Assoc.* **80**, 73–97.

WEST, M., HARRISON, P.J. & POLE, A. (1987). BATS: Bayesian analysis of time series. *The Professional Statistician*, **6**, 43-46.