Dynamic quantile linear models: a Bayesian approach

Kelly C. M. Gonçalves
IM-UFRJ

Helio S. Migon
IM-UFRJ

Leonardo S. Bastos
PROCC-Fiocruz

Abstract

A new class of models, named dynamic quantile linear models, is presented. It combines dynamic linear models with distribution free quantile regression producing a robust statistical method. Bayesian inference for dynamic quantile linear models can be performed using an efficient Markov chain Monte Carlo algorithm. A fast sequential procedure suited for high-dimensional predictive modeling applications with massive data, in which the generating process is itself changing overtime, is also proposed. The proposed model is evaluated using synthetic and well-known time series data. The model is also applied to predict annual incidence of tuberculosis in Rio de Janeiro state for future years and compared with global strategy targets set by the World Health Organization.

This paper aims to combine two innovative areas developed during the last quarter of the twentieth century building a useful broad new class of models, namely dynamic linear models and quantile regression. In the 1970s some new ideas to model time series data were put forward by Jeff Harrison and co-authors (Harrison and Stevens 1976). This class of models can be ingeniously viewed as regression models with parameters varying throughout time. At almost the same time, Roger Koenker introduced the quantile regression models, generalizing the L_1 regression, a robust procedure that has since been successfully applied to a range of statistical models (Koenker and Bassett 1978). It provides richer information on the effects of the predictors than does the traditional mean regression and it is very insensitive to heteroscedasticity and outliers, accommodating the non-normal errors often encountered in practical applications.

The inferential approaches of dynamic linear models and quantile regression are, nevertheless, completely distinct. While the first contribution follows the Bayesian paradigm, the other resorts to optimization techniques to solve the stated minimization problem, and its inference is theoretically founded on large sample theory. In the next paragraphs, we state the main novelties in both consolidated areas.

A simple minimization problem yielding the ordinary sample quantiles in the location model was shown to generalize naturally to the linear model, generating a broad new class of models introduced in the 1970s by Koenker and Bassett (1978) named "quantile regression models". Bayesian inference for quantile regression proceeds by forming the likelihood function based on the asymmetric Laplace distribution (Yu and Moyeed 2001). The recent literature includes a large number of papers: Yue and Rue (2011) add random effects to account for over-dispersion caused by unobserved heterogeneity or for correlation in longitudinal data; Luo et al. (2012) discuss inference using MCMC methods for longitudinal data models with random effects; Alhamzawi et al. (2011) discuss prior elicitation; Peng et al. (2014) explore varying covariate effects, including a new perspective on variable selection; Yang et al. (2016) evaluate the asymptotic validity of posterior inference; Yu (2015) presents quantile regression method for hierarchical linear models.

Quantile regression is also applied to temporal and spatially referenced data as a flexible and interpretable method of simultaneously detecting changes in several features of the distribution of some variables. A method based on estimating the conditional distribution was given by Cai (2002) and quantile autoregression was introduced by Koenker and Xiao (2006). Moreover, Reich et al. (2011) and Reich (2012) developed, respectively, a spatial and a spatiotemporal quantile regression models and, Reich and Smith (2013) proposed a semi-parametric quantile regression model for censored survival data. Applications of quantile regression have become widespread in recent years: in finance, actuarial modelign (Taylor 1999; Bassett Jr and Chen 2002; Kudryavtsev 2009; Sriram et al. 2016); in public policies evaluation (Neelon et al. 2015; Rahman 2016); in residual lifetime (Li et al. 2016).

Dynamic linear models (DLM) were introduced by Harrison and Stevens (1976) and extended to generalized linear models by West et al. (1985). The DLM are part of a broad class of models with time varying parameters, useful for modeling and forecasting time series and regression data (West and Harrison 1997; Durbin and Koopman 2002; Migon et al. 2005; Petris et al. 2009; Prado and West 2010). The advent of stochastic simulation techniques stimulated applications of the state space methodology to model complex stochastic structures, like dynamic spatiotemporal models (Gamerman et al.

2003), dynamic survival models (Bastos and Gamerman 2006), dynamic latent factor models (Lopes et al. 2008), and multiscale modeling (Ferreira et al. 2011; Fonseca and Ferreira 2017). Real applications of the methodology has been also appeared in hydrology (Ravines et al. 2008; Fernandes et al. 2009), intraday electricity load (Migon and Alves 2013), finance (Zhao et al. 2016), and many other areas.

We extend the dynamic linear models to a new class, named dynamic quantile linear models, where a linear function of the state parameters is set equal to a quantile of the response variable at time t, y_t, similar to the quantile regression of Koenker (2005). This method is suited for high-dimensional predictive modeling applications with massive data in which the generating process itself changes over time. Our proposal keeps the most relevant characteristics of DLM such as: (i) all relevant information sources are used: history, factual or subjective experiences, including knowledge of forthcoming events; (ii) everyday forecasting is generated by a statistical model and exceptions can be taken into account as an anticipative or retrospective base; (iii) *what happened* and *what if* analysis are easily accommodated; and (iv) the model can be decomposed into independent components describing particular features of the process under study.

The relative ease with which Markov chain Monte Carlo (MCMC) methods can be used to obtain the posterior distributions, even in complex situations, has made Bayesian inference very useful and attractive, however, at the expense of losing the sequential analysis of both states and parameters. Therefore, some fast computing alternatives exploring analytical approximations are welcome, as in Da-Silva et al. (2011) and Souza et al. (2018). In this paper, we introduce the inference via MCMC methods, and also via an alternative approach based on normal approximations and Bayes linear estimation (West et al. 1985), which besides being computationally faster than MCMC, recovers the sequential analysis of the data. Our approximation procedure provides the marginal likelihood sequentially as new data arrive, which is essential to perform sequential model monitoring and model selection (West 1981).

The remainder of the paper is organized as follows. Section 2 explores in more details the dynamic quantile linear model. Section 3 presents our efficient MCMC algorithm and the sequential approach for the dynamic quantile linear modeling. Section 4 illustrates the proposed method with synthetic data, and also presents some results of the well-known time series of the annual flow of the Nile River at Aswan from 1871 to 1970. The model is also applied to time series data on tuberculosis in Rio de Janeiro state, Brazil, from 2001 to 2015. We also predict the incidence for future years and compare the results with those of the global tuberculosis strategy targets

established. Finally, Section 5 concludes with discussion of the paper and some possible extensions.

1 Dynamic quantile linear model

The τ-th quantile of a random variable y_t at time t can be represented as a linear combination of explanatory variables,

$$\mathbb{Q}_\tau(y_t) = \mathbf{F}_t'\boldsymbol{\theta}_t^{(\tau)},$$

where $\mathbb{Q}_\tau(y_t)$ is the τ-quantile of y_t, formally defined as $\mathbb{Q}_\tau(y_t) = \inf\{y^* : P(y_t < y^*) \geq \tau\}$, for $0 < \tau < 1$. $\mathbf{F}_t$ is a $p \times 1$ vector of explanatory variables at time t, and $\boldsymbol{\theta}_t^{(\tau)}$ is a $p \times 1$ vector of coefficients depending on τ and t. From now on, the superscript τ will be omitted in order to keep the notation as simple as possible.

For a given time t, Koenker and Bassett (1978) defined the τ-th quantile regression estimator of $\boldsymbol{\theta}_t$ as any solution of the quantile minimization problem

$$\min_{\boldsymbol{\theta}_t} \rho_\tau \left(y_t - \mathbf{F}_t'\boldsymbol{\theta}_t\right),$$

where $\rho_\tau(.)$ is the loss (or check) function defined by $\rho_\tau(u) = u(\tau - I(u < 0))$, with $I(\cdot)$ denoting the indicator function. Minimizing the loss function $\rho_\tau(\cdot)$ is equivalent to maximizing the likelihood function of an asymmetric Laplace ($\mathcal{AL}$) distribution, as pointed out by Yu and Moyeed (2001). However, instead of maximizing the likelihood, as Yu and Moyeed (2001), we derive the posterior distribution of the τ-th quantile regression coefficients at time t using the $\mathcal{AL}$. Therefore, regardless the distribution of y_t, it is enough to assume that:

$$y_t \mid \mu_t, \phi, \tau \sim \mathcal{AL}\left(\mu_t, \phi^{-1/2}, \tau\right), \quad t = 1, 2, \ldots, T, \tag{1}$$

where $\mu_t = \mathbf{F}_t'\boldsymbol{\theta}_t \in \mathbb{R}$ is a location parameter, $\phi^{-1/2} > 0$ is a scale parameter, and $\tau \in (0, 1)$ is a skewness parameter representing the quantile of interest. In dynamic modeling, one goal is also to obtain the predictive distribution. This can be done in a robust fashion using a grid of values of $\tau \in (0, 1)$ to describe the full predictive distribution of y_t. Nevertheless, in this paper, we focus on providing precise inference about the linear predictor μ_t for each τ-th quantile.

In a dynamic linear model, the states at time t depend on the states at time $t-1$ according to an evolution equation $\boldsymbol{\theta}_t = \mathbf{G}_t\boldsymbol{\theta}_{t-1} + w_t$, where $\mathbf{G}_t$ is a $(p \times p)$ matrix describing the evolution parameters and w_t is a zero mean

Gaussian error with variance matrix $\mathbf{W}_t$. Therefore the proposed dynamic quantile linear model (DQLM) is defined as

$$\begin{array}{rcl} y_t|\boldsymbol{\theta}_t, \phi, \tau & \sim & \mathcal{AL}\left(\mathbf{F}_t'\boldsymbol{\theta}_t, \phi^{-1/2}, \tau\right), \\ \boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, \mathbf{W}_t & \sim & N_p(\mathbf{G}_t\boldsymbol{\theta}_{t-1}, \mathbf{W}_t), \quad t = 1, 2, \ldots, T, \end{array} \tag{2}$$

where $N_p(\mu, \Sigma)$ denotes de multivariate Gaussian distribution with p-dimensional mean vector μ and $p \times p$ covariance matrix Σ.

In the next section, we present two methods of inference for the proposed model. In the first one, we extend the algorithm proposed by Kozumi and Kobayashi (2011) for Bayesian (static) quantile regression and propose an efficient MCMC algorithm to sample from the posterior distribution of the unknown quantities of the dynamic quantile linear model (2). The second one is a computationally cheaper alternative that explores analytical approximations and Bayes linear optimality. An advantage of the last one is that it provides the marginal likelihood sequentially as new data arrive.

2 Posterior inference for the DQLM

The main tasks involved in state space models inference are estimation of the states and prediction of future values based on the current information. The conditional densities $\pi(\boldsymbol{\theta}_s \mid \mathcal{D}_t)$, where $\mathcal{D}_t = \{\mathcal{D}_{t-1} \cup I_t \cup y_t\}$ represents all information until time t, are calculated for different values of s and t. The filtering corresponds to the case $s = t$, state prediction to $s > t$ and smoothing to $s < t$, for $t = 1, 2, \ldots, T$.

The inference in DLM, assumes the data comes sequentially in time. The expressions for updating from the filtering density $\pi(\boldsymbol{\theta}_{t-1} \mid \mathcal{D}_{t-1})$ to $\pi(\boldsymbol{\theta}_t \mid \mathcal{D}_t)$ are easily obtained following the Bayesian argument. Given the state posterior distribution at time $t-1$, denoted by $\pi(\boldsymbol{\theta}_{t-1}|\mathcal{D}_{t-1})$ we obtain the prior distribution $\pi(\boldsymbol{\theta}_t|\mathcal{D}_{t-1})$ using the state parameters time evolution equation. As soon as we get a new information we can obtain by Bayes' theorem the posterior distribution at time t. The one step ahead predictive distribution is obtained by $\pi(y_t \mid \mathcal{D}_{t-1}) = \int \pi(y_t, \boldsymbol{\theta}_t|\mathcal{D}_{t-1}) d\boldsymbol{\theta}_t$.

The retrospective analysis consists instead in estimating the state sequence at times $1, \ldots, t$, given $\mathcal{D}_t$. It is solved by computing the conditional distribution of $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_t)$ given $\mathcal{D}_t$. As for filtering, smoothing can be implemented as a recursive algorithm.

Moreover, for a DLM including a possibly multidimensional unknown parameter in its specification, the joint posterior distribution of the parameter and unobservable states, in general, is not available in closed form. The inclusion of the states in the posterior distribution usually

simplifies the design of an efficient sampler. In fact, drawing the posterior distribution of the parameter given the states is almost invariably easier than drawing it from the marginal; in addition, efficient algorithms to generate the states conditionally on the data and the unknown parameter are available, such as the forward filtering backward sampling (FFBS) in the normal case (Frühwirth-Schnatter 1994).

However, if one needs to update the posterior distribution after one or more new observations become available, then one has to run MCMC all over again, and this can be extremely inefficient. On-line analysis and simulation-based sequential updating of the posterior distribution of states and unknown parameters are best dealt with employing sequential techniques, such as Sequential Monte Carlo (SMC) methods (Doucet et al. 2001).

Based on those ideias, we present two different algorithms to do inference in the proposed dynamic quantile linear model (2). The first method consists of an efficient MCMC algorithm to sample from the posterior distribution of the unknown quantities of the model . The second one is a faster alternative which provides a sequential analysis as new data arrive.

2.1 Efficient MCMC algorithm

Kotz et al. (2001) presented a location-scale mixture representation of the $\mathcal{AL}$ that allows finding analytical expressions for the conditional posterior densities of the model. In this way, if a random variable follows an asymmetric Laplace distribution, i.e. $y_t \mid \mu_t, \phi, \tau \sim \mathcal{AL}\left(\mu_t, \phi^{-1/2}, \tau\right)$, then we can write y_t using the following mixture representation:

$$\begin{aligned} y_t \mid \mu_t, U_t, \phi, \tau &\sim N(\mu_t + a_\tau U_t, b_\tau \phi^{-1/2} U_t), \\ U_t \mid \phi &\sim Ga(1, \phi^{1/2}), \end{aligned} \tag{3}$$

where $a_\tau = \frac{1-2\tau}{\tau(1-\tau)}$ and $b_\tau = \frac{2}{\tau(1-\tau)}$ are constants that depend only on τ, with $\mathrm{Ga}(\alpha, \beta)$ denoting the gamma distribution with mean α/β and variance α/β^2.

Therefore, the dynamic quantile linear model (2) can be rewritten as the following hierarchical model:

$$\begin{aligned} y_t|\boldsymbol{\theta}_t, U_t, \phi, \tau &\sim N\left(\mathbf{F}_t'\boldsymbol{\theta}_t + a_\tau U_t, b_\tau \phi^{-1/2} U_t\right), \\ \boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, \mathbf{W}_t &\sim N_p\left(\mathbf{G}_t\boldsymbol{\theta}_{t-1}, \mathbf{W}_t\right), \\ U_t|\phi &\sim Ga(1, \phi^{1/2}), \end{aligned} \tag{4}$$

for $t = 1, \ldots, T$. To allow for some flexibility in the model (4), we can even assume that $\phi^{-1/2}$ changes with time, which can be done in logarithmic scale according to a random walk or using the discounted variance learning

through the multiplicative gamma-beta-gamma model (West and Harrison 1997, chapter 10, p. 357).

The model is completed with a multivariate normal prior for $\boldsymbol{\theta}_0$, $\boldsymbol{\theta}_0 \sim N_p(\mathbf{m}_0, \mathbf{C}_0)$, an independent gamma prior for $\phi^{1/2}$, $\phi^{1/2} \sim Ga(n_\phi/2, s_\phi/2)$, and $\mathbf{W}_t^{-1} \sim Wishart(n_w, \mathbf{S}_w)$, where Wishart$(\nu, V)$ denotes the Wishart distribution with ν degrees of freedom and scale matrix V.

The posterior distribution of the parameters in the model (4) is given by

$$\pi(\boldsymbol{\Theta}, \mathbf{U}, \mathbf{W}, \phi \mid \mathcal{D}_T) \propto \prod_{t=1}^{T} [\pi(y_t|\boldsymbol{\theta}_t, U_t, \phi)\pi(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, \mathbf{W}_t)\pi(U_t)\pi(\mathbf{W}_t)] \quad (5)$$
$$\pi(\phi^{1/2})\pi(\boldsymbol{\theta}_0),$$

where $\mathcal{D}_t = \{\mathcal{D}_{t-1} \cup I_t \cup y_t\}$ represents all information until time t, for $t = 1, 2, \ldots, T$. The quantity I_t represents all external information at time t. If there is no external information at time t, then $I_t = \emptyset$. All prior information is summarized in $\mathcal{D}_0 = I_0$ containing all hyper parameters associated with the prior distributions. The unknown quantities are defined as follows: $\boldsymbol{\Theta} = (\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_T)$, $\mathbf{U} = (U_1, U_2, \ldots, U_T)$, $\mathbf{W} = (\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_T)$.

We can sample from the posterior distribution (5) through an MCMC algorithm. Our starting point is the efficient Gibbs sampler for Bayesian (static) quantile regression proposed by Kozumi and Kobayashi (2011). The dynamic coefficients are then sampled using a FFBS algorithm (Carter and Kohn 1994; Frühwirth-Schnatter 1994; Shephard 1994). The FFBS is a two-step efficient block sampler that draws the states jointly given the parameters for linear Gaussian state-space models. The data is sequentially processed to update numerical summaries of the filtering densities $\pi(\boldsymbol{\theta}_t|\mathcal{D}_t)$ at time t (forward filtering) and then the joint distribution is simulated via the implied backward compositional form (backward sampling). In this case, using standard results about the multivariate Gaussian distribution, it is easily proved that the random vector $(\boldsymbol{\Theta}, \mathcal{D}_T)$ has a Gaussian distribution for and by this way, the marginal and conditional distributions are also Gaussian distributions.

Theorem 2.1 displays the full conditionals and the sampling algorithm.

Theorem 2.1 *Let $\Phi = (\boldsymbol{\Theta}, \mathbf{U}, \mathbf{W}, \phi, t = 1, \ldots, T)$. A Gibbs sampling algorithm for the dynamic quantile model in (4) involves two main steps:*

1. *First, sample $\phi^{1/2}$, U_t, and $\mathbf{W}_t^{-1}$ for $t = 1, \ldots, T$ from their full conditional distributions:*

   (i) $\phi^{1/2} \mid \mathcal{D}_T, \boldsymbol{\Theta}, \mathbf{U} \sim Ga\left(n_\phi^*/2, s_\phi^*/2\right)$, where $n_\phi^* = n_\phi + 3T$ and

$$s_\phi^* = s_\phi + \sum_{t=1}^{T} \frac{(y_t - \mathbf{F}_t'\boldsymbol{\theta}_t - a_\tau U_t)^2}{b_\tau U_t} + 2\sum_{t=1}^{T} U_t.$$

(ii) $U_t \mid \mathcal{D}_T, \boldsymbol{\theta}_t, \phi \sim \mathcal{GIG}\left(\chi_t^*, \kappa_t^*, \frac{1}{2}\right)$, *where* $\chi_t^* = \frac{(y_t - \mathbf{F}_t'\boldsymbol{\theta}_t)^2\phi^{1/2}}{b_\tau}$ *and* $\kappa_t^* = \phi^{1/2}\left(\frac{a_\tau^2}{b_\tau} + 2\right)$ *and* $\mathcal{GIG}$ *is the generalized inverse Gaussian distribution (Jorgensen 2012).*

(iii) $\mathbf{W}_t^{-1} \mid \mathcal{D}_T, \boldsymbol{\theta}_t \sim Wishart\left(n_w^*, \mathbf{S}_w^*\right)$, *where* $n_w^* = n_w + p + 1$ *and* $\mathbf{S}_w^* = \mathbf{S}_w + (\boldsymbol{\theta}_t - \mathbf{G}_t\boldsymbol{\theta}_{t-1})'(\boldsymbol{\theta}_t - \mathbf{G}_t\boldsymbol{\theta}_{t-1})$.

2. *Next, use the FFBS method to sample from* $\pi(\boldsymbol{\theta} \mid \cdot)$*:*

(i) Forward filtering: for $t = 1, \ldots, T$ *calculate*

$$\mathbf{m}_t = E(\boldsymbol{\theta}_t|\mathcal{D}_t) = \mathbf{a}_t + \mathbf{R}_t\mathbf{F}_t q_t^{-1}(y_t - f_t) \text{ and}$$
$$\mathbf{C}_t = V(\boldsymbol{\theta}_t|\mathcal{D}_t) = \mathbf{R}_t - \mathbf{R}_t\mathbf{F}_t q_t^{-1}\mathbf{F}_t'\mathbf{R}_t,$$

with $\mathbf{a}_t = E(\boldsymbol{\theta}_t|\mathcal{D}_{t-1}) = \mathbf{G}_t\mathbf{m}_{t-1}$, $\mathbf{R}_t = V(\boldsymbol{\theta}_t|\mathcal{D}_{t-1}) = \mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}_t' + \mathbf{W}_t$,
$f_t = E(y_t|\mathcal{D}_{t-1}) = \mathbf{F}_t'\mathbf{a}_t + U_t a_\tau$ *and* $q_t = V(y_t|\mathcal{D}_{t-1}) = \mathbf{F}_t'\mathbf{R}_t\mathbf{F}_t + b_\tau U_t \phi^{-1/2}$.

(ii) Backward sampling: sample $\boldsymbol{\theta}_T \sim N_p(\mathbf{m}_T, \mathbf{C}_T)$ *and for* $t = T - 1, \ldots, 0$ *sample* $\boldsymbol{\theta}_t \sim N_p(\mathbf{h}_t, \mathbf{H}_t)$*, where*

$$\mathbf{h}_t = \mathbf{m}_t + \mathbf{C}_t\mathbf{G}_t'\mathbf{R}_{t+1}^{-1}(\boldsymbol{\theta}_{t+1} - \mathbf{a}_{t+1}) \text{ and } \mathbf{H}_t = \mathbf{C}_t - \mathbf{C}_t\mathbf{G}_{t+1}'\mathbf{R}_{t+1}^{-1}\mathbf{G}_{t+1}\mathbf{C}_t.$$

In place of assuming a Wishart prior for $\mathbf{W}_t^{-1}$ it is also possible to use a discount factor $\delta \in (0, 1)$ subjectively assessed, controlling the loss of information. In this case the unique difference is that $\mathbf{R}_t$ is recalculated according to a discount factor δ such as $\mathbf{W}_t = \frac{1-\delta}{\delta}\mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}_t'$. Hence, $\mathbf{R}_t$ can be rewritten as $\mathbf{R}_t = \frac{1}{\delta}\mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}_t'$.

The full conditional distribution of U_t is obtained using Lemma 2.1, which shows that the generalized inverse Gaussian distribution ($\mathcal{GIG}$) is conjugate to the normal distribution in a normal location-scale mixture model.

Lemma 2.1 *Let* $\mathbf{y} = (y_1, \ldots, y_n)$ *be a normal random sample with likelihood function* $\pi(\mathbf{y}|U) = \prod_{i=1}^n N(y_i|a + bU, cU)$ *and suppose that the prior distribution for* U *is* $\mathcal{GIG}(\chi, \kappa, \lambda)$*. Then, the posterior distribution* $U \mid \mathbf{y} \sim \mathcal{GIG}(\chi^*, \kappa^*, \lambda^*)$*, where* $\chi^* = \chi + c^{-1}\sum_{i=1}^n (y_i - a)^2$, $\kappa^* = nb^2c^{-1} + 2\kappa$ *and* $\lambda^* = \lambda - n/2$.

The proof of this result is immediate and is omitted in the text. Moreover, note that a $Ga(\alpha, \beta)$ distribution is a particular case of a $\mathcal{GIG}$ with $\chi = 0$, $\kappa = 2\beta$ and $\lambda = \alpha$.

2.2 Approximated dynamic quantile linear model

On the other hand, considering that data arrive sequentially, we propose an efficient and fast sequential inference procedure obtained with a closed-form solution, in order to update inference on unknown parameters online. It is useful for example in financial applications where one has to estimate hourly the term structure of interest rates as new data continously arrive. In particular, the main interest in the approximated method proposed is that besides being faster than MCMC algorithms, it does not require assessing chain convergence.

The approach also explores the mixture representation of the $\mathcal{AL}$ described in (3). Hence, posterior computation can be conveniently carried out using conventional Bayesian updating, conditional on the gamma random variable U_t. We also use a normal approximation to U_t's distribution in the logarithm scale, introducing explicit dynamic behavior, once again, generalizing the model presented in (4).

The normal approximation to the gamma distribution, described in Bernardo (1981), is presented below and the proof is presented in Appendix A.

Lemma 2.2 *Using the Kullback-Leibler divergence, in a large class of transformations, we have:*

(i) The best transformation, to approximate $\theta \sim Ga(a, b)$ *for a normal distribution is* $\zeta(\theta) = log(\theta)$. *Then* $\zeta \simeq N[E(\zeta), V(\zeta)]$, *where* $E(\zeta) \simeq \log\left(\frac{a}{b}\right) - \frac{1}{2a}$ *and* $V(\zeta) \simeq \frac{1}{a}$.

(ii) If $\zeta \sim N[E(\zeta), V(\zeta)]$ *then* $\theta = exp(\zeta)$ *is such that* $E(\theta) \simeq \exp\left[E(\zeta) + V(\zeta)/2\right]$ *and* $V(\theta) \simeq \exp\left[2E(\zeta) + V(\zeta)\right] V(\zeta)$.

Therefore, an approximate conditional normal dynamic quantile regression is obtained as:

$$\begin{array}{rl} y_t \mid \boldsymbol{\theta}_t, u_t, \phi & \sim N\left(\mathbf{F}_t'\boldsymbol{\theta}_t + a_\tau \phi^{-1/2} \exp(u_t), \phi^{-1} b_\tau \exp(u_t)\right), \\ \boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{t-1}, \mathbf{W}_t, \phi & \sim N_p\left(\mathbf{G}_t\boldsymbol{\theta}_{t-1}, \phi^{-1}\mathbf{W}_t\right), \\ u_t \mid u_{t-1}, W_{u,t}, \phi & \sim N\left(u_{t-1}, \phi^{-1} W_{u,t}\right), \end{array} \tag{6}$$

where $u_t = \log(U_t)$, with $U_t \sim Ga(a, b)$. Model (6) is more flexible than model (4) because it allows u_t to change with time according to a random walk. Furthermore, the scale parameter here is introduced in all the model equations.

The model is completed assuming the following independent prior distributions: $\boldsymbol{\theta}_0 \sim N_p(\mathbf{m}_0, \phi^{-1}\mathbf{C}_0)$, $u_0 \sim N(m_{u,0}, \phi^{-1}C_{u,0})$ and $\phi^{-1} \sim$

$Ga(n_0/2, d_0/2)$. The model in (4) can be viewed as a particular case of proposal (6) assuming that $W_{u,t} = 0, \forall t$, $m_{u,0} = -1/2$ and $C_{u,0} = 1$ and ignoring the multiplicative factor included here just to facilitate analytical expressions.

The inference procedure is described below. First we present all results conditional on both u_t and ϕ and later we integrate out those quantities.

2.2.1 Normal conditional model

We start the inference procedure by exploiting the advantage that, conditional on u_t and ϕ, we have normal distributions in one-step forecast and posterior distributions of $\boldsymbol{\theta}_t$ at time t, so all the properties of a normal model can be used here. The dependence on u_t appears first due to the one-step forecast distribution at time t and it will appear in the posterior distribution as time passes. Let us define $u_{1:t} = (u_1, \ldots, u_t)$. Theorem 2.2 presents the main steps in the inference procedure conditional on u_t.

Theorem 2.2 *Assuming that the states' posterior distribution at time $t-1$ is $\boldsymbol{\theta}_{t-1} \mid \mathcal{D}_{t-1}, u_{1:(t-1)}, \mathbf{W}_t, \phi \sim N[\mathbf{m}_{t-1}, \phi^{-1}\mathbf{C}_{t-1}]$ and the conditions defining model (6), it follows that the prior distribution of $\boldsymbol{\theta}_t$ and the conditional predictive distribution for any time t, given u_t and ϕ, are respectively:*

$$\begin{aligned} \boldsymbol{\theta}_t \mid \mathcal{D}_{t-1}, u_{1:(t-1)}, \mathbf{W}_t, \phi &\sim N_p[\mathbf{a}_t, \phi^{-1}\mathbf{R}_t], \\ y_t \mid \mathcal{D}_{t-1}, u_{1:t}, \mathbf{W}_t, \phi &\sim N[f_t(u_t), \phi^{-1}q_t(u_t)], \end{aligned} \tag{7}$$

with $\mathbf{a}_t = \mathbf{G}_t\mathbf{m}_{t-1}$ and $\mathbf{R}_t = \mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}_t + \mathbf{W}_t$, $f_t(u_t) = \mathbf{F}_t'\mathbf{a}_t + a_\tau\phi^{-1/2}\exp(u_t)$ and $q_t(u_t) = \mathbf{F}_t'\mathbf{R}_t\mathbf{F}_t + b_\tau\exp(u_t)$. The conditional joint covariance between y_t and $\boldsymbol{\theta}_t$, given $D_{t-1}, u_{1:t}, \phi$, is easily obtained as $\mathbf{R}_t\mathbf{F}_t$ completing the joint normal prior for $\boldsymbol{\theta}_t$ and y_t. Therefore, the posterior density of $\boldsymbol{\theta}_t$ follows as:

$$\boldsymbol{\theta}_t \mid \mathcal{D}_t, u_{1:t}, \mathbf{W}_t, \phi \sim N_p[\mathbf{m}_t(u_t), \phi^{-1}\mathbf{C}_t(u_t)], \tag{8}$$

where $\mathbf{m}_t(u_t) = \mathbf{a}_t + \mathbf{R}_t\mathbf{F}_t q_t(u_t)^{-1}(y_t - f_t(u_t))$ and $\mathbf{C}_t(u_t) = \mathbf{R}_t - \mathbf{R}_t\mathbf{F}_t q_t(u_t)^{-1}\mathbf{F}_t'\mathbf{R}_t$.

It is worth pointing out that mean and variance of the predictive and the state posterior distributions are functions of u_t, as reinforced by the notation used. However, since u_t is unknown for all t, we must find those distributions marginal on them. We will do this sequentially in the one-step forecast distribution in (7) for each time t.

2.2.2 Marginalizing on u_t

From now on, we will rewrite the time evolution equation in model (6) as $\boldsymbol{\theta}_t^* \mid \boldsymbol{\theta}_{t-1}^*, \mathbf{W}_t^*, \phi \sim N_{p+1}\left(\mathbf{G}_t^*\boldsymbol{\theta}_{t-1}^*, \phi^{-1}\mathbf{W}_t^*\right)$, where $\boldsymbol{\theta}_t^* = (\boldsymbol{\theta}_t, u_t)'$, $\mathbf{G}_t^* = \text{BlockDiag}\left(\mathbf{G}_t, 1\right)$ and $\mathbf{W}_t^* = \text{BlockDiag}\left(\mathbf{W}_t, W_{u,t}\right)$ with prior distribution given by $\boldsymbol{\theta}_0^* \mid \mathcal{D}_0, \phi \sim (\mathbf{m}_0^*, \phi^{-1}\mathbf{C}_0^*)$, where $\mathbf{m}_0^* = (\mathbf{m}_0, m_{u,0})'$ and $\mathbf{C}_0^* = \begin{pmatrix} \mathbf{C}_0 & \boldsymbol{\Lambda}_0 \\ \boldsymbol{\Lambda}_0' & C_{u,0} \end{pmatrix}$.

Assuming the posterior distribution at time $t-1$, $\boldsymbol{\theta}_{t-1}^* \mid \mathcal{D}_{t-1}, \mathbf{W}_t^*, \phi \sim N_{p+1}(\mathbf{m}_{t-1}^*,\ \phi^{-1}\mathbf{C}_{t-1}^*)$. By evolution, it follows that $\boldsymbol{\theta}_t^* \mid \mathcal{D}_{t-1}, \mathbf{W}_t^*, \phi \sim N_{p+1}(\mathbf{a}_t^*, \phi^{-1}\mathbf{R}_t^*)$, with $\mathbf{a}_t^* = \mathbf{G}_t^*\mathbf{m}_{t-1}^*$ and $\mathbf{R}_t^* = \mathbf{G}_t^*\mathbf{C}_{t-1}^*\mathbf{G}_t^* + \mathbf{W}_t^*$.

In particular, we have that $u_t \mid \mathcal{D}_{t-1}, W_{u,t}, \phi \sim N(a_{u,t}, \phi^{-1}R_{u,t})$ and the result (ii) of Lemma 2.2 leads to this distribution in the original (gamma) scale as:

$$U_t \mid \mathcal{D}_{t-1}, W_{u,t}, \phi \sim Ga(\alpha_t, \beta_t), \tag{9}$$

where $\alpha_t = \phi R_{u,t}^{-1}$ and $\beta_t = \exp(-a_{u,t})\phi R_{u,t}^{-1}$.

Thus, we have that the one-step forecast distribution in (7) can be seen as a normal-gamma mean-variance mixture, with the following different features from those stated in (3): (i) U_t in this case is gamma distributed with shape parameter different from 1; and (ii) $q_t(u_t)$ is a linear function of u_t with non null linear coefficient. These comments lead us to a recent class of distributions, a variant of the $\mathcal{AL}$, as described in Theorem 2.3.

Theorem 2.3 *The one-step ahead forecast distribution, conditional on ϕ and marginalized on u_t arises as the convolution of independent normal and generalized asymmetric Laplace distribution ($\mathcal{GAL}$). It can be represented as*

$$y_t = \zeta_t + \epsilon_t,$$

where $\zeta_t \sim \mathcal{GAL}(\mathbf{F}_t'\mathbf{a}_t, a_\tau\phi^{-1/2}\beta_t^{-1}, b_\tau\phi^{-1}\beta_t^{-1}, \alpha_t)$ and $\epsilon_t \sim N\left(0, \phi^{-1}\mathbf{F}_t'\mathbf{R}_t\mathbf{F}_t\right)$. We will refer to this as an $\mathcal{NGAL}$ distribution.

A brief presentation of the $\mathcal{NGAL}$ distribution, its moments the characteristic function, and the proof of Theorem 2.3 are presented in Appendix B. Although, the $\mathcal{NGAL}$ distribution suffers from a lack of closed-form expressions for its probability density and cumulative distribution functions, they can be efficiently determined using numerical integration as discussed in Appendix B. In particular, the one-step ahead forecast mean and variance marginal on u_t, can be easily obtained through properties of

conditional mean and variance as:

$$\begin{aligned}
E\left(y_t \mid \mathcal{D}_{t-1}, \phi\right) &= E\left[E\left(y_t \mid \mathcal{D}_{t-1}, U_t\right) \mid \mathcal{D}_{t-1}\right] = \mathbf{F}_t'\mathbf{a}_t + a_\tau \phi^{-1/2} E\left(U_t \mid \mathcal{D}_{t-1}\right) \\
&= \mathbf{F}_t'\mathbf{a}_t + a_\tau \phi^{-1/2}\alpha_t/\beta_t = f_t, \text{ and} \\
V\left(y_t \mid \mathcal{D}_{t-1}, \phi\right) &= E\left[V\left(y_t \mid \mathcal{D}_{t-1}, U_t\right) \mid \mathcal{D}_{t-1}\right] + V\left[E\left(y_t \mid \mathcal{D}_{t-1}, U_t\right) \mid \mathcal{D}_{t-1}\right] \\
&= \phi^{-1}\left[\mathbf{F}_t'\mathbf{R}_t\mathbf{F}_t + b_\tau E\left(U_t \mid \mathcal{D}_{t-1}\right)\right] + a_\tau^2 \phi^{-1} V\left(U_t \mid \mathcal{D}_{t-1}\right) \\
&= \phi^{-1}\left(\mathbf{F}_t'\mathbf{R}_t\mathbf{F}_t + b_\tau\alpha_t/\beta_t + a_\tau^2\alpha_t/\beta_t^2\right) = \phi^{-1} q_t.
\end{aligned}$$

The recurrences for posterior mean and variance may also be derived using approaches that do not invoke the normal assumption, since they possess strong optimality properties that are derived when the distributions are only partially specified in terms of means and variances. The Bayes linear estimation procedure, presented in West and Harrison (1997, Chap. 4), provides an alternative estimate that can be viewed as an approximation to the optimal procedure. Theorem 2.4 presents the main steps in the inference procedure now marginal on u_t.

Theorem 2.4 *The joint distribution of $\boldsymbol{\theta}_t^*$ and y_t is partially described using its first and second moments, as follows:*

$$\left(\begin{array}{c} \boldsymbol{\theta}_t^* \\ y_t \end{array} \middle| \mathcal{D}_{t-1}, \mathbf{W}_t^*, \phi\right) \sim \left[\left(\begin{array}{c} \mathbf{a}_t^* \\ f_t \end{array}\right), \phi^{-1}\left(\begin{array}{cc} \mathbf{R}_t^* & \mathbf{A}_t q_t \\ q_t\mathbf{A}_t' & q_t \end{array}\right)\right],$$

where $\mathbf{A}_t = q_t^{-1}\left(\begin{array}{c} \mathbf{R}_t\mathbf{F}_t + \phi^{-1/2} a_\tau \exp(a_{u,t})\boldsymbol{\Lambda}_t \\ \boldsymbol{\Lambda}_t'\mathbf{F}_t + \phi^{-1/2} a_\tau \exp(a_{u,t}) R_{u,t} \end{array}\right)$ *and* $\boldsymbol{\Lambda}_t = \mathbf{G}_t\boldsymbol{\Lambda}_{t-1}$.

The joint covariance between y_t and $\boldsymbol{\theta}_t^*$, given D_{t-1} and ϕ, is obtained using the first order Taylor approximation $\exp(u_t) \approx \exp(a_{u,t})[1 + u_t - a_{u,t}]$.

Through the Bayes linear estimation procedure, we get:

$$\boldsymbol{\theta}_t^* \mid \mathcal{D}_t, \mathbf{W}_t^*, \phi \sim [\mathbf{m}_t^*, \phi^{-1}\mathbf{C}_t^*], \tag{10}$$

where $\mathbf{m}_t^* = \mathbf{a}_t^* + \mathbf{A}_t(y_t - f_t)$ *and* $\mathbf{C}_t^* = \mathbf{R}_t^* - \mathbf{A}_t q_t \mathbf{A}_t'$ *and we can easily return to the normality assumption.*

The discount factor strategy can be used in place of $\mathbf{W}_t^*$.

Although, one of the attractive features of this approach is that estimation and forecasting can be applied sequentially, as new data become available, one can use the backward-recursive algorithm and get the smoothed estimates: $\boldsymbol{\theta}_t^* \mid \mathcal{D}_T, \mathbf{W}_t^*, \phi \sim [\mathbf{h}_t^*, \phi^{-1}\mathbf{H}_t^*]$, where

$$\begin{aligned}
\mathbf{h}_t^* &= \mathbf{m}_t^* + \mathbf{C}_t^*\mathbf{G}^{*\prime}{}_{t+1}\mathbf{R}_{t+1}^{*^{-1}}(h_{t+1}^* - \mathbf{a}_{t+1}^*) \text{ and} \\
\mathbf{H}_t^* &= \mathbf{C}_t^* - \mathbf{C}_t^*\mathbf{G}^{*\prime}{}_{t+1}\mathbf{R}_{t+1}^{*^{-1}}(\mathbf{R}_{t+1}^* - \mathbf{H}_{t+1}^*)\mathbf{R}_{t+1}^{*^{-1}}\mathbf{G}^*{}_{t+1}\mathbf{C}_t^*.
\end{aligned} \tag{11}$$

2.2.3 Estimating ϕ

The steps of the method described above are conditional on ϕ. In the case where ϕ is unknown a practical solution is to use a plug-in estimator for ϕ obtained from the maximum a posteriori estimation.

The posterior distribution of ϕ given the observed data is given by:

$$p(\phi \mid \mathcal{D}_T) = \prod_{t=1}^{T} p(y_t \mid \mathcal{D}_{t-1}, \phi) p(\phi \mid \mathcal{D}_0), \tag{12}$$

where $p(y_t \mid \mathcal{D}_{t-1}, \phi)$ is the predictive distribution conditional on ϕ.

Closed-form expressions for the density function of the family of $\mathcal{NGAL}$ distributions are not available as far as we know. However, the density and the cumulative distribution function can be obtained numerically using the convolution form or the inversion of the characteristic function. In particular, using the convolution to represent the density function, we get:

$$p(y_t \mid \mathcal{D}_{t-1}, \phi) = \int_{-\infty}^{\infty} p_\epsilon(y_t - z) p_\zeta(z) dz, \tag{13}$$

where $p_\zeta(.)$ is the density of the $\mathcal{GAL}$ distribution, and $p_\epsilon(.)$ is the density of a normal distribution with mean 0 and variance c. Also, using the inversion of the characteristic function, we get:

$$p(y_t \mid \mathcal{D}_{t-1}, \phi) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-isy_t} \varphi(s) ds, \tag{14}$$

where $\varphi(.)$ is the characteristic function of the $\mathcal{NGAL}$ distribution.

The integrals (13) and (14) can be evaluated numerically using current quadrature methods. For example, Kuonen (2003) discussed numerical integration using Gaussian quadrature and adaptive rules, which dynamically concentrate the computational work in the sub regions where the integrand is most irregular, and the Monte Carlo method. They concluded that adaptive Gauss-Kronrod quadrature performed best in their examples.

Finally, a brief summary of the algorithm is stated in the following steps:

(1) for $k = 0$ give an initial value $\phi^{-1(0)}$;

(2) calculate for $t = 1, \ldots, T$:

(i) $\mathbf{a}_t^* = \mathbf{G}_t^* \mathbf{m}_{t-1}^*$, $\mathbf{R}_t^* = \mathbf{G}_t^* \mathbf{C}_{t-1}^* \mathbf{G}_t^* + \mathbf{W}_t^*$, $\alpha_t = \phi R_{u,t}^{-1}$ and $\beta_t = \exp(-a_{u,t}) \phi R_{u,t}^{-1}$;

(ii) $f_t = \mathbf{F}_t' \mathbf{a}_t + a_\tau \phi^{-1/2} \alpha_t / \beta_t$ and $q_t = (\mathbf{F}_t' \mathbf{R}_t \mathbf{F}_t + b_\tau \alpha_t / \beta_t + a_\tau^2 \alpha_t / \beta_t^2)$;

(iii) get $p(y_t \mid \mathcal{D}_{t-1}, \phi^{(k)})$ numerically using (13) or (14);

(iv) calculate $\mathbf{m}_t^* = \mathbf{a}_t^* + \mathbf{A}_t(y_t - f_t)$ and $\mathbf{C}_t^* = \mathbf{R}_t^* - \mathbf{A}_t q_t \mathbf{A}_t'$, where $\mathbf{A}_t$ is a function of $\phi^{-1(k)}$.

(3) do $k = k + 1$ and get $\phi^{-1(k)}$ maximizing $p(\phi|\mathcal{D}_t)$ in (12).

(4) repeat (2) and (3) until convergence is achieved.

3 Applications

To illustrate the performance of the proposed model and inference procedures, we apply the method to some synthetic data and well-known time series data. Then, we apply it to the incidence of tuberculosis in Rio de Janeiro, in which it is important to assess if public health policies are effective, not only in reducing the trend in the number of cases, but also the variability of total cases. Moreover, the upper quantiles may be useful to detect an epidemic.

Although the approximated method presents less computing burden and keeps the relevant sequential analysis of the data, we interchange the use of the MCMC approach with the approximate method in the following applications.

The choice of the quantile to be tracked depends on the specific aims of the problem. In each application we arbitrarily fixed a small quantile (10%, 25%), the median, and a large quantile (75%, 90%) to be monitored, with the purpose to illustrate the methodology for different scenarios. However, in some specific contexts, there is a practical rationale behind this choice. For example, on the Value-at-Risk (VaR) estimation, used by financial institutions and their regulators as the standard measure of market risk, τ-th quantile is often set to 0.01 or 0.05. Or in survival analysis, the mean residual life function of a survival time S, given by the expected remaining lifetime given survival up to time t_τ, this is $E(S - t_\tau \mid S > t_\tau)$, it is especially useful when the tail behavior of the distribution is of interest.

3.1 Artificial data examples

In order to assess the efficiency of the proposed sequential procedure and the convergence of the MCMC estimation, two artificial datasets were generated. The proposed model was fitted to these datasets and its estimates were then compared to the true values used in the dataset generation process. In the first study we generate a time series with temporal trend and seasonal

component with one harmonic from a Gaussian DLM and estimate the linear predictor for 3 different quantiles and compare the results obtained from each inferential method. In the second one, the focus was to empirically evaluate the ability of our proposed model in estimating the parameters with a non-normal artificial dataset, for which a data transformation can be considered.

In both studies a non-informative prior distribution is assumed for the parametric vector with: $\mathbf{m}_0 = 0$, $\mathbf{C}_0 = 10^5$, $n_\phi = 0.001$, $s_\phi = 0.001$. We take two approaches to deal with the evolution variance: (i) we set a Wishart prior distribution for $\mathbf{W}_t^{-1}$; and (ii) we apply a discount factor $\delta = 0.95$ setting $\mathbf{W}_t = \mathbf{C}_t(1-\delta)/\delta$ (West and Harrison 1997, p. 51).

The results shown hereafter for the MCMC algorithm correspond to 110,000 MCMC sweeps, after a burn-in of 10,000 iterations and the chain thinning by taking every 10th sample value. As the default FFBS-based approach produces results that use the posterior based on the full data, for comparison purposes the results for the approximated method are based in the smoothing equations (11).

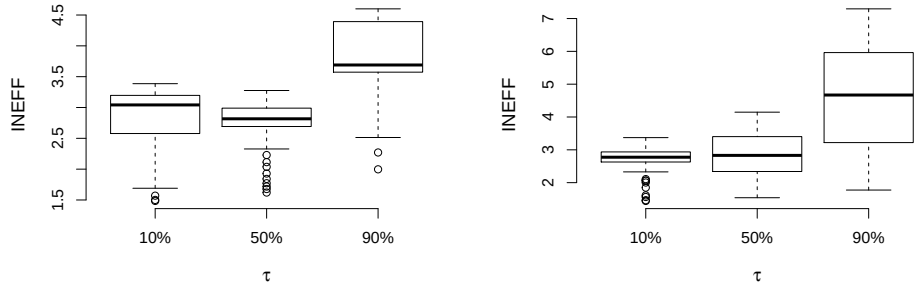
3.1.1 Trend and seasonal DLM

An artificial time series of size $T = 100$ was generated for a Gaussian dynamic linear model, with the specification $\mathbf{F}_t = (1, 0, 1, 0)'$ and $\mathbf{G}_t = \begin{pmatrix} \mathbf{L_2} & \mathbf{0} \\ \mathbf{0} & \mathbf{J_2}(\omega) \end{pmatrix}$, where $\mathbf{L}_2 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $\mathbf{J_2}(\omega) = \begin{pmatrix} cos(\omega) & sin(\omega) \\ -sin(\omega) & cos(\omega) \end{pmatrix}$, with $\omega = \frac{2\pi}{12}$. This corresponds to a second-order polynomial model with a harmonic component. We arbitrarily fixed $V = 49$ and $\mathbf{W} = \begin{pmatrix} \mathbf{W_2} & \mathbf{0} \\ \mathbf{0} & \mathbf{I_2} \end{pmatrix}$, where $\mathbf{W_2} = \begin{pmatrix} 0.02 & 0.01 \\ 0.01 & 0.01 \end{pmatrix}$ and $\mathbf{I}_n$ is an identity matrix of dimension n.

The DQLM was fitted for $\tau = 0.1, 0.5$ and 0.9 and both inference approaches proposed in the paper were considered. In the MCMC algorithm, we assumed an inverse Wishart prior distribution for $\mathbf{W}$ with hyperparameters $n_w = 8$ and $\mathbf{S}_w = 0.1\mathbf{I}_4$. Convergence for all the parameters was achieved and Figures 11 and 12 in Appendix C show, respectively, the trace plots with the posterior distribution of parameters $\boldsymbol{\theta}_t$'s and the histograms of the posterior densities of some elements of the covariance matrix $\mathbf{W}$.

Moreover, Figure 1 reports the inefficiency factor (INEFF) for some of the parameters, defined as the ratio between the numerical variance of the posterior sample mean and the variance of the sample mean, assuming uncorrelated draws from the target distribution. The larger INEFF the less efficient the sampling scheme (Gamerman and Lopes 2006, p. 126). The

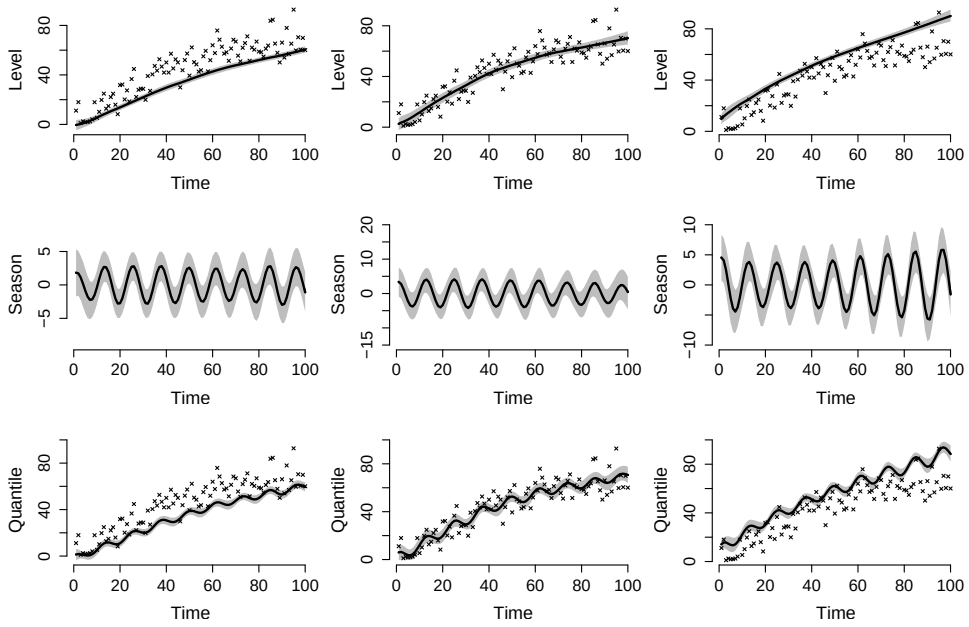
INEFF are below ten for each parameter, which indicates that the sampler is mixing well, mainly for the lower quantile and the median.



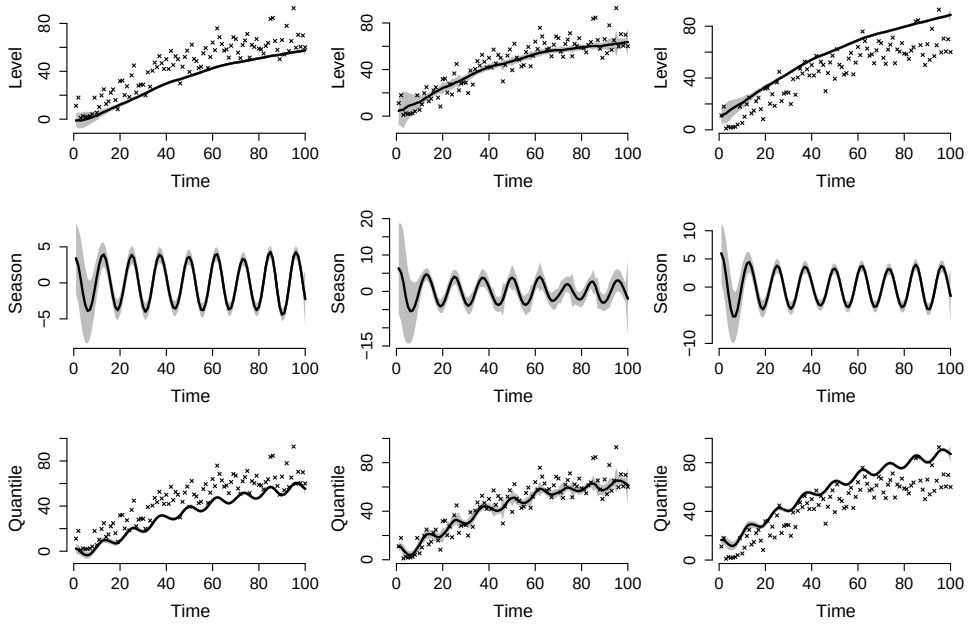
(a) θ_1 (b) θ_3

Figure 1: Box plots of the inefficiencies of θ_{1t} and θ_{3t}, for $t = 1, \ldots, T$ and $\tau = 0.1, 0.5, 0.9$.

Panels of Figure 2 present the posterior summary of the level and seasonal components and the linear predictor $\mathbf{F}_t'\boldsymbol{\theta}_t$ for each quantile (from left to right $\tau = 0.1, 0.5, 0.9$), with the generated time series (points). The posterior mean is represented by the solid line and the 95% credible region by the shaded area.



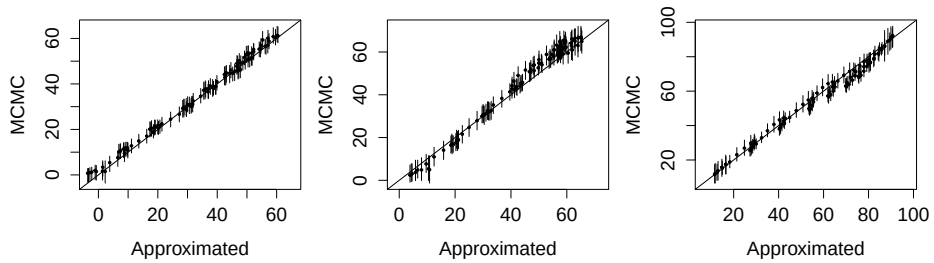
(a) MCMC



(b) Approximated method

Figure 2: Smoothed posterior mean (solid line), 95% credible region (shaded area) of the level and seasonal components for each quantile, based on the MCMC output (a) and in the approximated method (b). From left to right $\tau = 0.1, 0.5, 0.9$.

Figure 3 presents the plot of the estimated values of the linear predictor for each quantile under MCMC approach versus the proposed approximated algorithm. The lengths of the segments represent the 95% credibility interval obtained by MCMC approach. We conclude that both methods produce similar results, but while MCMC takes about 5 minutes for 5,000 sweeps for a specific quantile, the approximated method takes seconds. Both algorithms were implemented in the R programming language, version 3.4.1 (R Core Team 2017), in a computer with an Intel(R) Core(TM) i7-7700 processor 3.60 GHz. This is first evidence of the relevance of the approximated DQLM to deal with a scalable dataset.



(a) 10% (b) 50% (c) 90%

Figure 3: Plot of the estimated values (posterior mean) of the linear predictor under the MCMC inference approach versus the approximated method.

3.1.2 Non-Gaussian artificial dataset

To illustrate how the method works with a non-Gaussian dataset, we generated an artificial dataset from a first order dynamic gamma regression with mean μ_t, scale parameter ϕ and a canonical link function $\eta_t = \log(\mu_t) = \mathbf{F}_t'\boldsymbol{\theta}_t$. In particular, we generated $T = 100$ observations assuming $\mathbf{F}_t = (1, x_t)$, where x_t is an auxiliary variable at time t, $\mathbf{G}_t = \mathbf{I}_{2\times 2}$, $\mathbf{W} = 0.01\mathbf{I}_{2\times 2}$, for all $t = 1, \ldots, T$ and $\phi = 50$. Each auxiliary variable x_t was generated independently from a uniform distribution defined in the interval (2,4).

The model (4) was fitted to the original data and to the log transformed data, for the quantiles 0.10, 0.50, and 0.90. We particularly choose here to do the inference using only the MCMC algorithm. Although quantile regression is invariant to monotonic transformations, that is, the quantiles of the transformed variable are the transformed quantiles of the original variable (Koenker 2005, chapter 2, p. 34), the estimates of all the involved quantities were noticeably better when the data were transformed.

In order to validate the former result, some simulation studies were developed. Several samples were generated from the gamma model. A simple static model is proposed in this exercise, with $\mathbf{F}_t = 1$ and $\mathbf{W}_t = 0$, for all $t = 1, \ldots, T$. Fifty replications of samples of sizes $T = 100$ and $T = 250$ were generated using three different levels of skewness.

Figure 4 reports the empirical nominal coverage of the 95% credibility intervals measured in percentages and the relative mean absolute error (RMAE) for the posterior mean of the quantiles for each case. The RMAE decreases when we apply the $\mathcal{AL}$ to the logarithm of the sample observations and also as the sample size increases or the distribution becomes more symmetrically distributed. The desired nominal level of 95% is best achieved when the logarithm of the sample is considered, mainly as the level of asymmetry decreases. The improvement in the results when using the log transformation is more noticeable as the skeweness decreases.

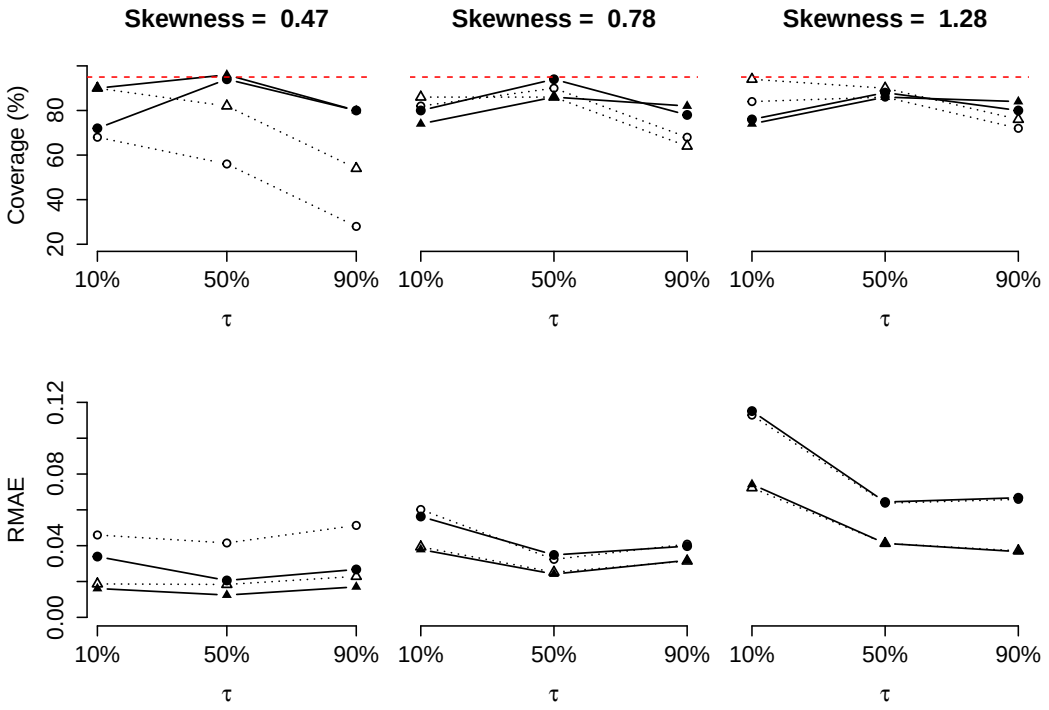


Figure 4: Empirical nominal coverage of the 95% credibility intervals (%) and the MAE for the posterior mean of the quantiles for each case. The symbols • and ▲ represent the results obtained when fitting the quantile regression to the logarithm of the observations for $T = 100$ and $T = 250$, respectively. ∘ and △ represent the same results when the quantile regression is applied to the observations on the original scale for $T = 100$ and $T = 250$, respectively.

Those results encourage us, as a future work, to explore in this context the idea of using the $\mathcal{AL}$ distribution for random effects in the link function, instead of applying it to the transformed response variable.

3.2 Real data example: Nile River flow

In this section we revisit a classic univariate time series previously analyzed in the literature. We apply the proposal to the annual flow of the Nile River at Aswan from 1871 to 1970 (Cobb 1978). This series shows level shifts, so we considered here a model which includes change points or structural breaks. The dataset is available in the R software (R Core Team 2017).

We also apply the proposal to a well-known toy example with the time series of quarterly gas consumption in the UK from 1960 to 1986, in which there is a possible change in the seasonal factor around the third quarter of 1970 in Appendix 4.

The dataset in Figure 5 corresponds to the measurements of the annual flow of Nile River at Aswan (Egypt) from 1871 to 1970. The time series shows level shifts. The construction of the first dam of Aswan started in 1898 and the second big dam was completed in 1971, which caused enormous changes on the Nile flow and in the vast surrounding area. The application of a quantile model for this kind of dataset could be interesting to determine, for example, the return period of a flood. The return period $R(y)$ is defined to be the mean value of a geometric distributed random variable with probability parameter $P(y > y^*) = 1 - \tau$. Then, $R(y) = 1/(1 - \tau)$.

In order to capture these possible level changes we consider here a model that does not assume a regular pattern and stability of the underlying system, but can include change points or structural breaks.

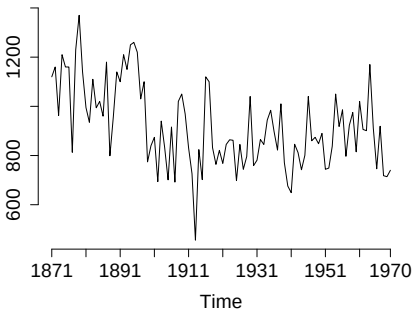


Figure 5: Measurements of the annual flow of Nile River at Aswan from 1871 to 1970.

A simple way to account for observations that are unusually far from their one step-ahead predicted value is to describe the evolution error using a heavy-tailed distribution. The Student-t distribution family is particularly appealing in this respect for two reasons: (i) the Student-t distribution admits a simple representation as a scale mixture of normal distributions, which allows one to treat a DLM with t-distributed observation errors as a Gaussian DLM, conditional on the scale parameters; and (ii) the FFBS algorithm can still be used. Thus, we consider, the DQLM with evolution characterized by the Student-t distribution, given by:

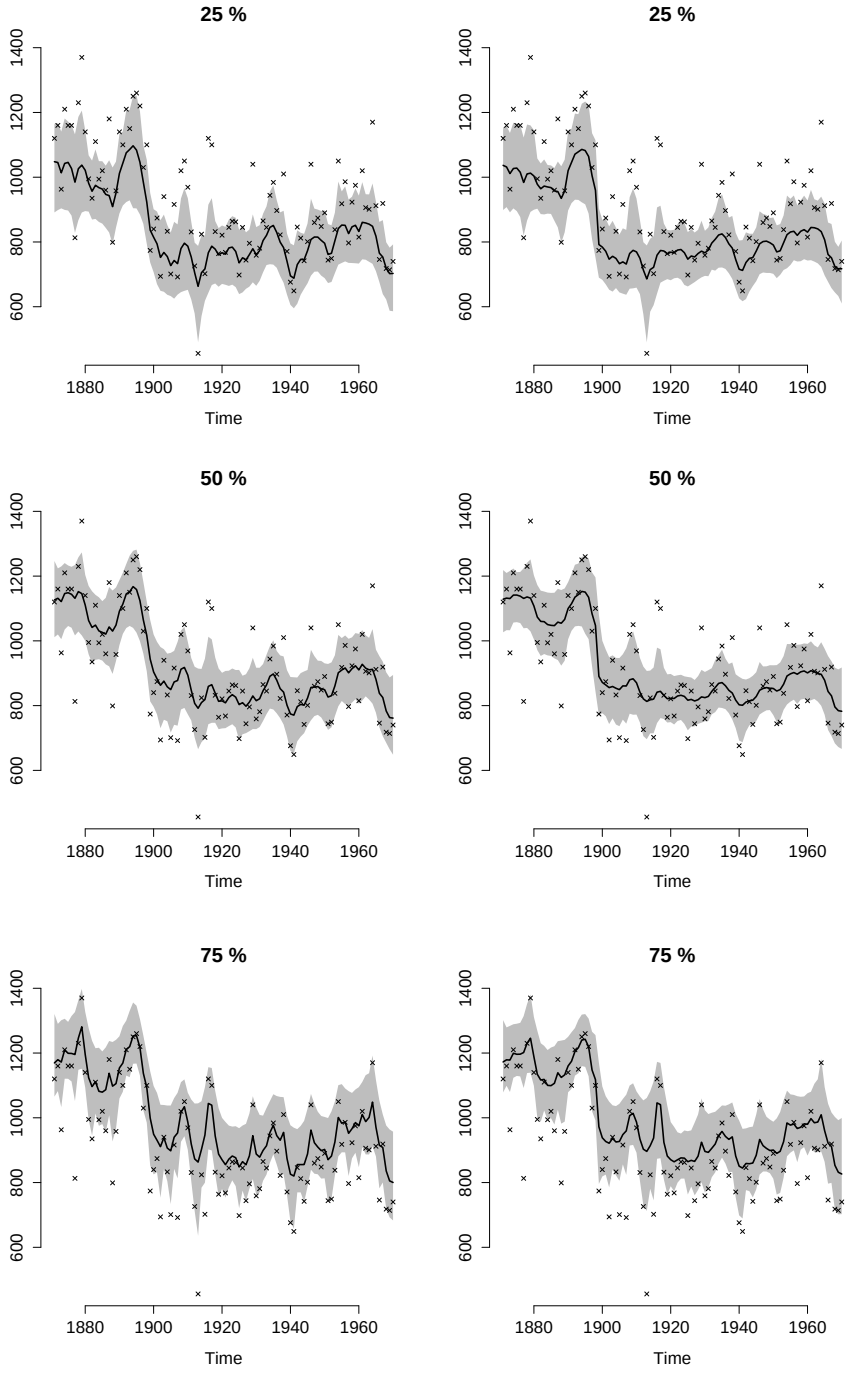
$$\begin{aligned} \boldsymbol{\theta}_t &\sim \mathrm{N}(\mathbf{G}_t\boldsymbol{\theta}_{t-1}, \lambda_t^{-1}\mathbf{W}), \\ \lambda_t &\sim \mathrm{Ga}(\nu/2, \nu/2), t = 1, \ldots, T. \end{aligned} \tag{15}$$

The latent variable λ_t^{-1} can be informally interpreted as the degree of non-Gaussianity of w_t. In fact, values of λ_t^{-1} lower than 1 make larger absolute values of w_t more likely. Hence, the posterior distribution of λ_t^{-1} can be used to flag possible outliers. Through its degree-of-freedom parameter ν, which may also vary on time, different degrees of heaviness in the tails can be attached. This class of models is discussed in Petris et al. (2009).

In this example, we fitted a first-order polynomial dynamic model, thus we assumed $F_t = G_t = 1$ in model (4). We fitted the model (15) and the one with normal evolution, for the 0.25, 0.5, and 0.75-quantiles.

In both cases, for the variance of the states W, we considered a half-Cauchy prior distribution, discussed by Gelman (2006), with scale 25, set as a weakly informative prior distribution. Although we could even assume a prior distribution for this, as described in Petris et al. (2009), we assumed ν known a priori and fixed at 2.5.

Figure 6 shows the linear predictor for each quantile, with its 95% credibility interval represented by the shaded area, obtained from the normal (first column) and Student-t fits (second column). Model (15) results in smoother linear predictors with more accurate credibility intervals.



(a) Normal (b) Student-T

Figure 6: Posterior mean of the linear predictor (represented by the solid line) and the 95% credible region (represented by the shaded area) for each quantile, under normal and student-t evolution.

Table 1 presents the posterior mean of the linear predictor for the 0.25, 0.5, and 0.75-quantiles for the model with normal and Student-t evolution around the year 1899. The abrupt regime change is better captured in the Student-t than in the normal model for all quantiles.

Table 1: Posterior mean of the linear predictor for the 0.25, 0.5, and 0.75-quantiles of the model with normal and Student-t evolution.

	Student-t			Normal		
Year	25%	50%	75%	25%	50%	75%
1896	1064.01	1134.51	1219.03	1046.11	1124.00	1216.96
1897	1004.97	1083.83	1147.69	984.23	1064.32	1137.24
1898	959.51	1047.47	1117.77	922.98	1016.65	1093.48
1899	792.07	890.29	972.25	836.57	943.65	999.19
1900	784.17	871.67	938.27	814.69	903.88	949.23
1901	771.33	863.93	926.85	794.25	882.78	930.59

Figure 7 shows the boxplots of the posterior distribution of λ_t^{-1} in logarithmic scale from 1871 to 1970 for each quantile. Values of $\log \lambda_t^{-1}$ greater than 0 indicate an abrupt regime change. Boxplots in gray do not include the value 0. Thus, it is possible to observe that the model in fact accommodates the outliers. The regime change in 1899 is detected for all the quantile regression models fitted. However, for the 0.75-quantile other outliers were detected.

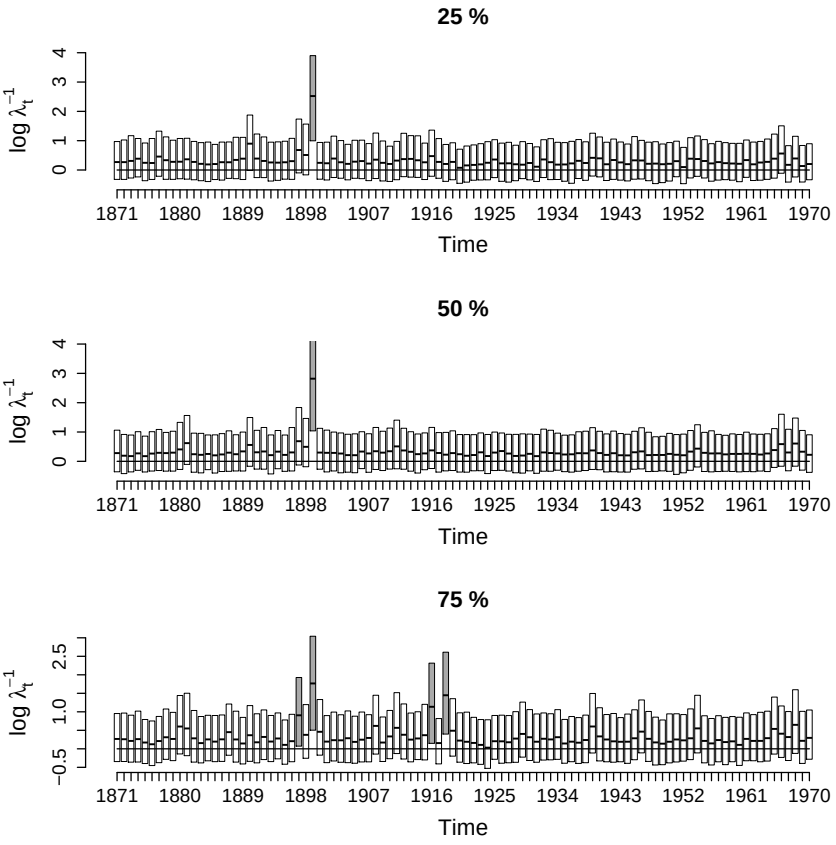


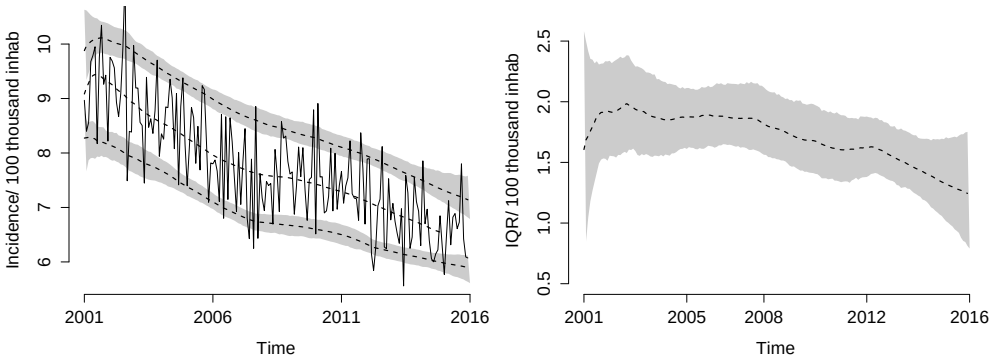
Figure 7: Boxplots of the posterior samples of λ_t^{-1} in log scale for the 0.25, 0.5, and 0.75 quantiles over the years.

3.3 Tuberculosis cases in Rio de Janeiro

According to the World Health Organization (WHO), tuberculosis (TB) is one of the top 10 causes of death worldwide. Brazil is one of the countries with the highest number of cases in the world and since 2003 the disease has been considered a priority by the Brazilian Ministry of Health. As part of the overall effort to reduce the incidence and mortality rate, the Ministry of Health, through the General Office to coordinate the National Tuberculosis Control Program (CGPNCT), prepared a national plan to end tuberculosis as a public health problem in Brazil, reaching at target of less than 10 cases per 100,000 inhabitants by 2035. The plan to end the TB epidemic is a target under the Sustainable Development Goals that requires implementing a mix of biomedical, public health and socioeconomic interventions along with

research and innovation. The World Health Assembly passed a resolution approving with full support the new post-2015 End TB Strategy with its ambitious targets (WHO 2014).

The state of Rio de Janeiro is located in southeastern Brazil, with over 16 million residents in 2017. The Rio de Janeiro state has one of the highest TB rates in the country. In 2015, there were 13,094 new notified cases, representing 15% of new cases for the whole country. We fit the proposed DQLM with a trend component for monthly incidence in Rio de Janeiro from January 2001 to December 2015. Figure 8 (a) presents the posterior mean (dashed line) and the 95% credibility interval (region in gray) for the linear predictor for the 0.1, 0.5, and 0.9-quantiles, under the MCMC procedure. Figure 8 (b) presents the posterior summary of the interquantile range (IQR) between 10% and 90%. It is possible to observe a decreasing pattern for the IQR, mainly after 2008.



(a) (b)

Figure 8: Monthly incidence of TB in Rio de Janeiro state: (a) Posterior mean (dashed line) of the linear predictor for the 0.1, 0.5, and 0.9-quantiles, and their 95% credibility intervals for the 0.1 and 0.9-quantiles; (b) Posterior summary for the interquantile range between the 0.1 and 0.9 quantiles.

One of the targets of the post-2015 global tuberculosis strategy is a 20% reduction in tuberculosis incidence by 2020, compared with 2015, a 50% reduction by 2025, an 80% reduction by 2030 and a 90% reduction in tuberculosis incidence by 2035, compared with 2015. In particular, the interest here is to predict the incidence from January 2016 to December 2020, in order to detect in terms of the median, if the target can be achieved.

In MCMC inference procedure, samples from $\boldsymbol{\theta}_{T+k}$, k a non-negative integer, are obtained by propagating the samples from the posterior

distribution through the evolution equation (4). In the approximated method, this is done after estimating ϕ through maximum posterior estimation. Hence, we get that:

$$\boldsymbol{\theta}^*_{T+k} \mid \mathcal{D}_T \sim [\mathbf{a}^*_{T+k}, \phi^{-1}\mathbf{R}^*_{T+k}], \tag{16}$$

where $\mathbf{a}^*_{T+k} = \mathbf{G}_{T+k}\mathbf{a}^*_{T+k-1}$ and $\mathbf{C}^*_{T+k} = \mathbf{G}_{T+k}\mathbf{R}^*_{T+k-1}\mathbf{G}'_{T+k}$ can be recursively calculated.

In order to capture the steep decline of new cases that represents the government target, we considered beside a linear forecast function in the original scale, a linear forecast for the log-transformation to the dataset and two higher-order models. First, we considered a quadratic growth DLM (West and Harrison 1997, chapter 7, p. 223), which is obtained specifying in equation (2) $\mathbf{F}_t = (1, 0, 0)'$, $\mathbf{G}_t = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$ and $\mathbf{W}_t$ a diagonal matrix.

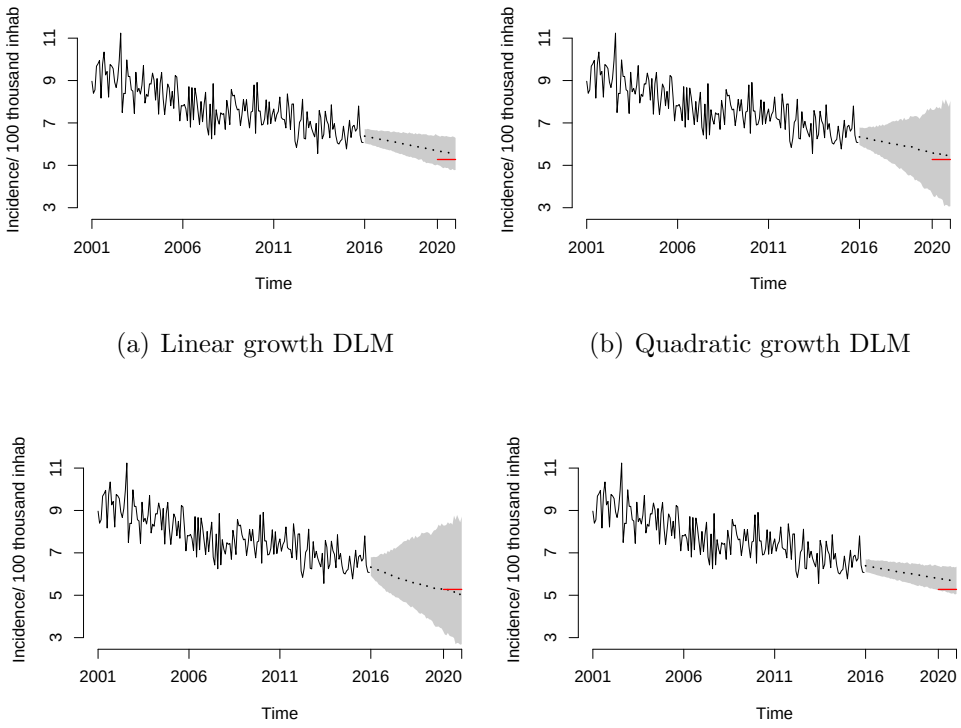
Then, we considered a locally approximation by a second-order Taylor expansion to the non-linear Gompertz-type forecast functions (Harrison et al. 1977). The Gompertz function evolves three parameters and is defined by:

$$g(t) = A_t + B_t \exp(C_t t). \tag{17}$$

To deal with the non-linear form, the quadratic polynomial DLM, previously described, was used for updating the corresponding three parameters, then, the form in (17) was converted back and the full non-linear form used for the forecast function.

Figure 9 presents the forecast from 2016 to 2020 of the median of the month incidence of TB per 100 thousand inhabitants for both alternatives previously described, under the MCMC approach, for the four models previously described. The dashed line represents the posterior mean, the region in gray the 95% credibility interval and the red line indicates the TB reduction targets calculated for Rio de Janeiro state using 2015 as the baseline.

According to all four models, the WHO's 2020 target, estimated by 5.28 cases per month per 100 thousand inhabitants, can be achieved in Rio de Janeiro. The Gompertz model captures the steep decline and it estimates for January 2020 a TB incidence of 5.19 (95%CI = (3.15, 8.29)) TB cases per 100 thousand inhabitants. As in the Gompertz model fitting, we estimate $B_T < 0$ (mean = - 0.99, 95%CI =(-1.12, -0.96)) and $C_T > 0$ (mean = 0.04, 95%CI =(-0.11, 0.18)), $g(t)$ defined in (17) is a non-increasing model converging to A_t as $t \to \infty$. Thus, considering a typical stagnation of the disease incidence, the Gompertz model seems to be the more appropriate to capture such long-term forecasts.



(a) Linear growth DLM

(b) Quadratic growth DLM

(c) Gompertz model

(d) Linear growth DLM - log data

Figure 9: Temporal predictions of monthly incidence of TB in Rio de Janeiro state from January 2016 to December 2020. The line represents the observed time series and the dashed line represents the mean of the predictive distribution for the median (0.5-quantile), the region in gray represents the 95% credibility interval and the red line the WHO's reduction target, under each model considered.

4 Conclusions

In this article we propose a new class of models, named dynamic quantile linear models. For the inference procedure, we develop two approaches: (i) a MCMC algorithm based on Gibbs sampling and FFBS for the model of the location-scale mixture representation of the asymmetric Laplace distribution; and (ii) a faster sequential method, based on approximations using Kullback-Leibler divergence and Bayes linear method. The second inference approach has the advantage of being computationally cheaper. A suggested alternative approach is the use of SMC methods for online state

and parameter estimation in the state-space model proposed. In particular, the approach presented in Yang et al. (2018) could be applied to the DQLM in (2) and (4). The approach considers filtering and smoothing methods via SMC for general state space models in the context of state and fixed parameters, providing a correction to the smoothing methods proposed by Carvalho et al. (2010). However, as illustrated by Yang et al. (2018) the filtering and smoothing methods with SMC could have almost the same computational time of a MCMC.

We evaluated the DQLM in artificial and real datasets. In the simulation study, we applied our model in a Gaussian example with trend and seasonal components where the DQLM performed well, and the approximate DQLM was a computationally efficient alternative to MCMC. We also applied our model in a non-Gaussian example generated by a gamma model encouraging the investigation of introducing the $\mathcal{AL}$ model in the link function, or in the response variable. In the classic real Nile River data example, we illustrated by fitting a model for outliers and structural breaks that the detection of occasional abrupt changes differs depending on the quantile of interest.

The application to the tuberculosis data in Rio de Janeiro, Brazil, illustrates the practical importance of evaluating quantiles instead of the mean in the context of forecasting. It also encourages us to extend the proposal to joint modeling for the quantiles and a dynamic quantile hierarchical model. In particular, our method can be applied to any infectious disease, once it is important to assess whether public health policies are effective, not only in reducing the trend in the number of cases, but also the variability of number of total cases. Moreover, the upper quantiles can be useful for early detection of an outbreak (epidemic). If the distribution of the number of cases (represented by the quantiles) is much higher than usual, it is a strong indication that attention is required.

References

R. Alhamzawi, K. Yu, and J. Pan. Prior elicitation in Bayesian quantile regression for longitudinal data. *Journal of Biometrics & Biostatistics*, 2 (3):1–7, 2011. doi: 10.4172/2155-6180.1000115.

G. W. Bassett Jr and H.-L. Chen. Portfolio style: Return-based attribution using quantile regression. In *Economic Applications of Quantile Regression*, pages 293–305. Springer, 2002.

L. S. Bastos and D. Gamerman. Dynamic survival models with spatial frailty. *Lifetime Data Analysis*, 12(4):441–460, 2006.

J. M.-M. B. Bernardo. *Bioestadística: una perspectiva bayesiana*. Number 61: 311. Vicens-vives, 1981.

Z. Cai. Regression quantiles for time series. *Econometric Theory*, 18(01): 169–192, 2002.

C. K. Carter and R. Kohn. On gibbs sampling for state space models. *Biometrika*, 81(3):541–553, 1994.

C. M. Carvalho, M. S. Johannes, H. F. Lopes, N. G. Polson, et al. Particle learning and smoothing. *Statistical Science*, 25(1):88–106, 2010.

G. W. Cobb. The problem of the nile: conditional solution to a changepoint problem. *Biometrika*, 65(2):243–251, 1978.

C. Da-Silva, H. S. Migon, and L. Correia. Dynamic bayesian beta models. *Computational Statistics & Data Analysis*, 55(6):2074–2089, 2011.

A. Doucet, N. De Freitas, and N. Gordon. An introduction to sequential monte carlo methods. In *Sequential Monte Carlo methods in practice*, pages 3–14. Springer, 2001.

J. Durbin and S. J. Koopman. A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89(3):603–616, 2002.

M. V. Fernandes, A. M. Schmidt, and H. S. Migon. Modelling zero-inflated spatio-temporal processes. *Statistical Modelling*, 9(1):3–25, 2009.

M. A. Ferreira, S. H. Holan, and A. I. Bertolde. Dynamic multiscale spatiotemporal models for gaussian areal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(5):663–688, 2011.

T. C. O. Fonseca and M. A. R. Ferreira. Dynamic multiscale spatiotemporal models for poisson data. *Journal of the American Statistical Association*, 112(517):215–234, 2017. doi: 10.1080/01621459.2015.1129968.

S. Frühwirth-Schnatter. Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15(2):183–202, 1994.

D. Gamerman and H. F. Lopes. *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. Chapman and Hall/CRC, 2006.

D. Gamerman, A. R. Moreira, and H. Rue. Space-varying regression models: specifications and simulation. *Computational Statistics & Data Analysis*, 42(3):513–533, 2003.

A. Gelman. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, 1(3):515–534, 2006.

P. J. Harrison and C. F. Stevens. Bayesian forecasting. *Journal of the Royal Statistical Society. Series B (Methodological)*, 38(3):205–247, 1976.

P. J. Harrison, T. Leonard, and T. N. Gazard. An application of multivariate hierarchical forecasting. Technical Report 15, Department of Statistics, University of Warwick., 1977.

B. Jorgensen. *Statistical properties of the generalized inverse Gaussian distribution*, volume 9. Springer Science & Business Media, 2012.

R. Koenker. *Quantile regression*. Number 38. Cambridge university press, 2005.

R. Koenker and G. Bassett. Regression quantiles. *Econometrica*, 46(1):33–50, January 1978.

R. Koenker and Z. Xiao. Quantile autoregression. *Journal of the American Statistical Association*, 101(475):980–990, 2006.

S. Kotz, T. Kozubowski, and K. Podgorski. *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance*. Springer Science & Business Media, 2001.

H. Kozumi and G. Kobayashi. Gibbs sampling methods for Bayesian quantile regression. *Journal of Statistical Computation and Simulation*, 81(11): 1565–1578, 2011.

A. A. Kudryavtsev. Using quantile regression for rate-making. *Insurance: Mathematics and Economics*, 45(2):296–304, 2009.

D. Kuonen. Numerical integration in S-PLUS or R: A survey. *Journal of Statistical Software*, 8(13):1–14, 2003.

R. Li, X. Huang, and J. Cortes. Quantile residual life regression with longitudinal biomarker measurements for dynamic prediction. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 65(5):755–773, 2016.

H. F. Lopes, E. Salazar, D. Gamerman, et al. Spatial dynamic factor analysis. *Bayesian Analysis*, 3(4):759–792, 2008.

Y. Luo, H. Lian, and M. Tian. Bayesian quantile regression for longitudinal data models. *Journal of Statistical Computation and Simulation*, 82(11): 1635–1649, 2012.

H. S. Migon and L. C. Alves. Multivariate dynamic regression: modeling and forecasting for intraday electricity load. *Applied Stochastic Models in Business and Industry*, 29(6):579–598, 2013.

H. S. Migon, D. Gamerman, H. F. Lopes, and M. A. Ferreira. *Handbook of statistics. Bayesian statistics: modeling and computation*, volume 25, chapter Bayesian dynamic models, pages 553–588. Elsevier, 2005.

B. Neelon, F. Li, L. F. Burgette, and S. E. Benjamin Neelon. A spatiotemporal quantile regression model for emergency department expenditures. *Statistics in Medicine*, 34(17):2559–2575, 2015.

L. Peng, J. Xu, and N. Kutner. Shrinkage estimation of varying covariate effects based on quantile regression. *Statistics and Computing*, 24(5):853–869, 2014.

G. Petris, S. Petrone, and P. Campagnoli. *Dynamic Linear Models with R*. Springer-Verlag, 2009.

R. Prado and M. West. *Time series: modeling, computation, and inference*. CRC Press, 2010.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017. URL `http://www.R-project.org/`.

M. A. Rahman. Bayesian quantile regression for ordinal models. *Bayesian Analysis*, 11(1):1–24, 2016.

R. R. Ravines, A. M. Schmidt, H. S. Migon, and C. D. Rennó. A joint model for rainfall–runoff: the case of rio grande basin. *Journal of Hydrology*, 353 (1-2):189–200, 2008.

W. J. Reed. The normal-Laplace distribution and its relatives. In *Advances in Distribution Theory, Order Statistics, and Inference*, pages 61–74. Springer, 2006.

B. J. Reich. Spatiotemporal quantile regression for detecting distributional changes in environmental processes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61(4):535–553, 2012.

B. J. Reich and L. B. Smith. Bayesian quantile regression for censored data. *Biometrics*, 69(3):651–660, 2013.

B. J. Reich, M. Fuentes, and D. B. Dunson. Bayesian spatial quantile regression. *Journal of the American Statistical Association*, 106(493):6–20, 2011.

N. Shephard. Partial non-Gaussian state space. *Biometrika*, 81(1):115–131, 1994.

M. A. O. Souza, H. S. Migon, and J. B. M. Pereira. Extended dynamic generalized linear models: the two-parameter exponential family. *Computational Statistics & Data Analysis*, 121(5):164–179, 2018.

K. Sriram, P. Shi, and P. Ghosh. A Bayesian quantile regression model for insurance company costs data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 179(1):177–202, 2016.

J. W. Taylor. A quantile regression approach to estimating the distribution of multiperiod returns. *The Journal of Derivatives*, 7(1):64–78, 1999.

M. West. Robust sequential approximate bayesian estimation. *Journal of the Royal Statistical Society (Ser. B)*, 43:157–166, 1981.

M. West and J. Harrison. *Bayesian Forecasting and Dynamic Models*. Springer, 1997.

M. West, P. J. Harrison, and H. S. Migon. Dynamic generalized linear models and bayesian forecasting. *Journal of the American Statistical Association*, 80(389):73–83, 1985.

WHO. Draft global strategy and targets for tuberculosis prevention, care and control after 2015. Documentation for World Health Assembly 67 A67/11, World Health Organization, 2014.

B. Yang, J. R. Stroud, and G. Huerta. Sequential monte carlo smoothing with parameter estimation. *Bayesian Analysis*, 2018.

Y. Yang, H. J. Wang, and X. He. Posterior inference in Bayesian quantile regression with asymmetric Laplace likelihood. *International Statistical Review*, 84(3):327–344, 2016.

K. Yu and R. A. Moyeed. Bayesian quantile regression. *Statistics & Probability Letters*, 54(4):437–447, 2001.

Y. Yu. Bayesian quantile regression for hierarchical linear models. *Journal of Statistical Computation and Simulation*, 85(17):3451–3467, 2015.

Y. R. Yue and H. Rue. Bayesian inference for additive mixed quantile regression models. *Computational Statistics & Data Analysis*, 55(1):84–96, 2011.

Z. Y. Zhao, M. Xie, and M. West. Dynamic dependence networks: Financial time series forecasting and portfolio decisions. *Applied Stochastic Models in Business and Industry*, 32(3):311–332, 2016.

Appendix 1: Proof of Lemma 2.2

Lemma 2.2

(i) Let us find a tractable monotone transformation $\zeta = \zeta(\theta)$ to induce normality in the sense of minimizing the divergence measure between $p(\zeta)$ and its normal approximation. In particular we have $\theta \sim Ga(a, b)$. This is equivalent to finding ζ that minimizes the expected value

$$k[p(\zeta), q(\zeta)] = \int p(\zeta) \log \frac{p(\zeta)}{q(\zeta)} d\zeta,$$

where $q(\zeta)$ is the density function of the $N[E(\zeta), V(\zeta)]$ distribution and $p(\zeta) = p(\theta)/|d\zeta/d\theta|$.

Thus, we want a function ζ which minimizes

$$\begin{aligned} k[p(\zeta), q(\zeta)] &= \int p(\zeta) \log p(\zeta) d\zeta - \int p(\zeta) \log q(\zeta) d\zeta \\ &= \int p(\zeta) \log p(\zeta) d\zeta + \frac{1}{2} \log(2\pi V(\zeta)) \int p(\zeta) d\zeta + \int p(\zeta) \left[\frac{(\zeta - E(\zeta))^2}{2V(\zeta)} \right] d\zeta \\ &= \int p(\zeta) \log p(\zeta) d\zeta + \frac{1}{2} \log(2\pi V(\zeta)). \end{aligned}$$

We consider the class of transformations $\zeta = \zeta(\theta)$, such that $d\zeta/d\theta = \theta^{-\alpha}$, $0 \leq \alpha \leq 1$, which contains as particular cases the standard transformations:

$$\zeta(\theta) = \begin{cases} \theta, & \text{for } \alpha = 0, \\ \log \theta, & \text{for } \alpha = 1, \\ (1-\alpha)\theta^{1-\alpha}, & \text{for } \alpha < 1. \end{cases}$$

We have

$$\begin{aligned} \int p(\zeta) \log p(\zeta) d\zeta &= \int p(\theta) \log p(\theta) d\theta - \int \log \zeta'(\theta) p(\theta) d\theta \\ &= C_1 + \alpha \int \log \theta p(\theta) d\theta = C_1 + \alpha E(\log \theta) \approx C + \alpha \left(\log \frac{a}{b} - \frac{1}{2a} \right). \end{aligned}$$

Moreover, using that $V(\zeta) \approx [\zeta'(E(\theta))]^2 V(\theta)$, we get $\log(2\pi V(\zeta)) = 2\alpha \log(b/a) + C_2$. Then,

$$k[p, q] = \alpha \left(\log \frac{b}{a} - \frac{1}{2a} \right) + C,$$

which is a decreasing function in α for the particular values that a and b can assume in the method. It follows that progressively better normalizing transformations are obtained for larger values of α. Thus, in this class of transformations, $\zeta(\theta) = \log\theta$ is the one which minimizes the divergence measure between $p(\zeta)$ and its normal approximation.

(ii) If we have $\zeta \sim N[\mu, \sigma^2]$, such that $\mu = E(\zeta)$ and $\sigma^2 = V(\zeta)$, then $\theta = \exp(\zeta)$ has distribution $LN[\mu, \sigma^2]$ and we desire the better $Ga(a, b)$ to approximate this lognormal distribution in the sense of Kullback-Leibler divergence. That is, we want a and b which minimize

$$k[p(\theta), q(\theta)] = \int p(\theta)\log p(\theta)d\theta - \int p(\theta)\log q(\theta)d\theta,$$

or maximizes

$$f(a, b) = \int p(\theta)\log q(\theta)d\theta,$$

where $q(\theta)$ and $p(\theta)$ are the density functions of the $Ga(a, b)$ and $LN(\mu, \sigma^2)$ distribution, respectively.

Thus,

$$\begin{aligned} f(a,b) = \int p(\theta)\log q(\theta)d\theta &= \int \left(a\log b - \log(\Gamma(a)) + (a-1)\log(\theta) - b\theta\right) p(\theta)d\theta \\ &= a\log b - \log(\Gamma(a)) + (a-1)E_{LN}(\log(\theta)) - bE_{LN}(\theta) \\ &= a\log b - \log(\Gamma(a)) + (a-1)\mu - b\exp\left(\mu + \sigma^2/2\right). \end{aligned}$$

Deriving it with respect to a and b and setting equal to zero, we get that the maximum of $f(a, b)$ is achieve when:

$$a = \sigma^{-2} \text{ and } b = \sigma^{-2}\exp\left[-\left(\mu + \frac{\sigma^2}{2}\right)\right],$$

for $\sigma^2 < 1$, in order to guarantee the existence of the mode of the gamma distribution.

Appendix 2: $\mathcal{NGAL}$ distribution

This distribution was first presented in Reed (2006), who described some of its properties. A random variable Y has a $\mathcal{NGAL}$ distribution if its characteristic function is equal to

$$\varphi_y(s) = \left(\frac{\alpha\beta\exp(i\delta s - \gamma s^2/2)}{(\alpha - is)(\beta - is)}\right)^{\rho}, \tag{18}$$

where α, β, ρ and γ are positive parameters and δ is real. We write $Y \sim \mathcal{NGAL}(\delta, \gamma, \alpha, \beta, \rho)$ to indicate that Y follows such a distribution. Using the cumulants of the distribution it is possible to obtain:

$$E(Y) = \rho\left(\delta + \frac{1}{\alpha} - \frac{1}{\beta}\right) \text{ and } Var(Y) = \rho\left(\gamma + \frac{1}{\alpha^2} + \frac{1}{\beta^2}\right).$$

Also, the coefficients of kurtosis and skewness are, respectively, given by:

$$\frac{k_4}{k_2^2} = \frac{6(\alpha^4 + \beta^4)}{\rho(\gamma\alpha^2\beta^2 + \alpha^2 + \beta^2)^2} \text{ and } \frac{k_3}{k_2^{3/2}} = \frac{2(\beta^3 - \alpha^3)}{\rho^{1/2}(\gamma\alpha^2\beta^2 + \alpha^2 + \beta^2)^2}.$$

Figure 10 compares the density curves of $\mathcal{NGAL}$ for some values of the parameters. It is possible to observe that as σ^2 increases, the density becomes both wider and flatter. The parameters α and β affect, respectively, the upper and lower tail behavior of the $\mathcal{NGAL}$ distribution: small values of α and β correspond, respectively, to a fat upper and lower tails, while as they increase, the upper and lower tails of the distribution reduce to those of a normal distribution. Finally, as ρ increases both mean and variance increase.

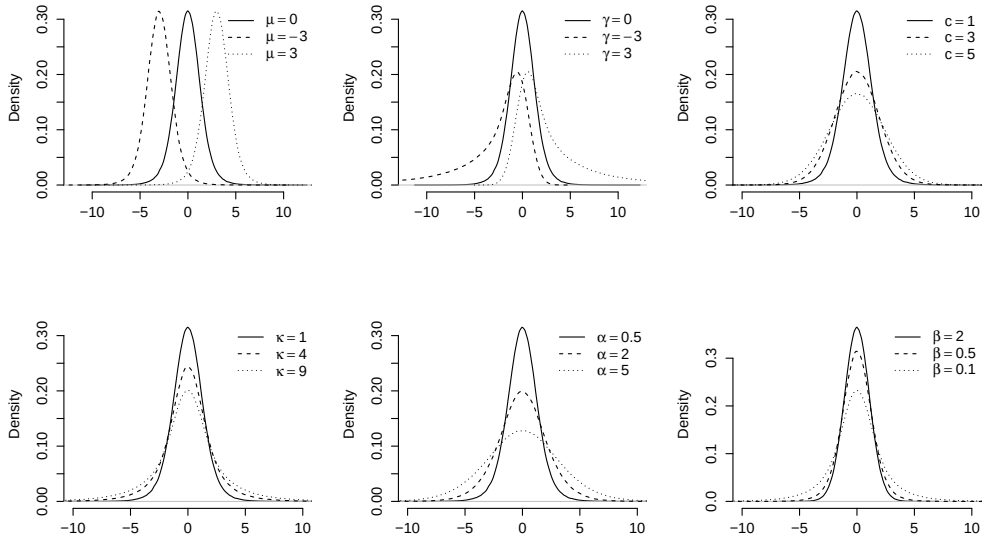


Figure 10: Density function of the $\mathcal{NGAL}$ distribution with some different values of the parameters.

Finally, we need to prove that the distribution described in Theorem 2.3 is a $\mathcal{NGAL}$ distribution. In order to facilitate the notation, let us omit the

index t and call: $\theta = \mathbf{F}_t'\mathbf{a}_t$, $\mu = a_\tau \phi^{-1/2}\beta_t^{-1}$, $\sigma^2 = \phi^{-1}b_\tau\beta_t^{-1}$, $c = \phi^{-1}\mathbf{F}_t'\mathbf{R}_t\mathbf{F}_t$, $\rho = \alpha_t$ and $w = U_t$. Thus, we have:

$$\begin{aligned} y \mid w &\sim N\left(\theta + \mu w, c + \sigma^2 w\right), \\ w &\sim Ga(\rho, 1). \end{aligned}$$

Conditional on w, we obtain the ch.f of y as:

$$\varphi_y(s) = E\left[E\left(e^{isy} \mid w\right)\right] = e^{is\theta}\int_0^\infty e^{is\mu w} E\left(e^{is(c+\sigma^2 w)^{1/2}z}\right) g(w)dw$$

$$\begin{aligned} &= e^{is\theta}\frac{1}{\Gamma(\alpha)}\int_0^\infty e^{is\mu w} e^{-\frac{1}{2}s^2(c+\sigma^2 w)} w^{\alpha-1}e^{-w}dw \\ &= e^{is\theta-\frac{1}{2}s^2c}\frac{1}{\Gamma(\alpha)}\int_0^\infty w^{\alpha-1}e^{-w\left(1+\frac{1}{2}\sigma^2 s^2 - i\mu s\right)}dw \\ &= e^{is\theta-\frac{1}{2}s^2c}\left(\frac{1}{1+\frac{1}{2}\sigma^2 s^2 - i\mu s}\right)^\rho. \end{aligned} \tag{19}$$

Thus, $\varphi_Y(s) = \varphi_\epsilon(s).\varphi_\zeta(s)$, where $\varphi_\epsilon(s)$ is the $\mathrm{N}(0, c)$ ch.f and $\varphi_\zeta(s)$ is the $\mathcal{GAL}(\theta, \mu, \sigma, \rho)$ ch.f. Thus, it follows that the NGAL distribution is that of the convolution of the normal $\mathrm{N}(0, c)$ and $\mathcal{GAL}(\theta, \mu, \sigma, \rho)$ distributions, that is $Y = \zeta + \epsilon$, for independent ζ and ϵ.

Moreover, the ch.f in (19) can be expressed as stated in (18), for $\alpha = \frac{2\kappa}{\sqrt{2}\sigma}$, $\beta = \frac{2}{\sqrt{2}\sigma\kappa}$, $\delta = \theta/\rho$, $\gamma = c/\rho$ and the additional parameter $\kappa > 0$ is related to μ and σ as follows:

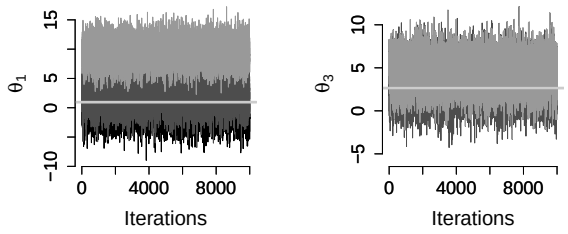
$$\kappa = \frac{\sqrt{2\sigma^2 + \mu^2} - \mu}{\sqrt{2}\sigma}, \text{ while } \mu = \frac{\sigma}{\sqrt{2}}\left(\frac{1}{\kappa} - \kappa\right),$$

showing that it is $\mathcal{NGAL}$ distributed.

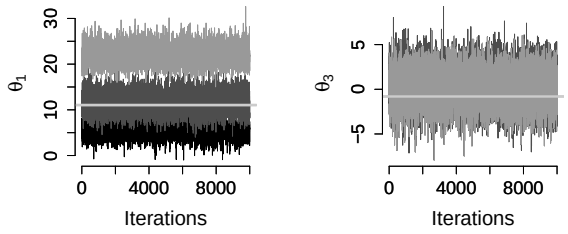
Appendix 3: Assessment by MCMC

Figure 11 shows the trace plot with the posterior distribution of parameters $\boldsymbol{\theta}_t$'s for some times for each quantile regression fitted. The chains in black are obtained for quantile 0.10, in dark gray for quantile 0.50 and in light gray for 0.90. The line represents the true value of each component of $\boldsymbol{\theta}_t$ used in the data generating process.

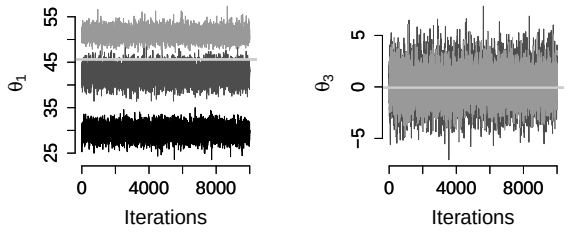
Figure 12 presents histograms of the posterior densities of some elements of the covariance matrix $\mathbf{W}$ for the median, where the true value used in the data generation process is given by the vertical dashed line. The hyperparameters appear to be well estimated.



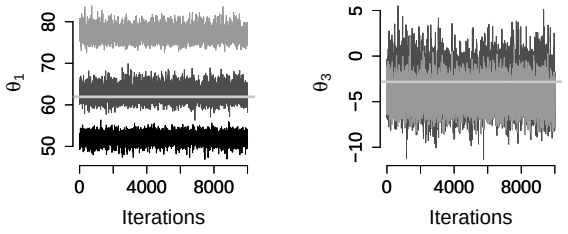
(a) $t = 1$



(b) $t = 10$



(c) $t = 40$



(d) $t = 80$

Figure 11: Trace plot with the posterior densities of θ_t's for some times and σ for each quantile regression fitted.

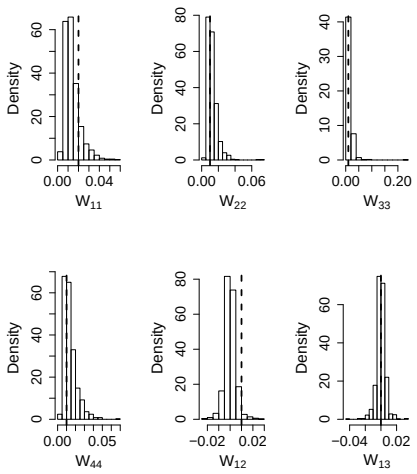


Figure 12: Histograms with the posterior densities of the hyperparameters in the diagonal of the covariance matrix **W** for the median.

Appendix 4: A toy example: UK gas consumption

The following dataset consists of quarterly UK gas consumption from 1960 to 1986 (see Figure 13). The plot of the data suggests a possible change in the seasonal factor around the third quarter of 1970.

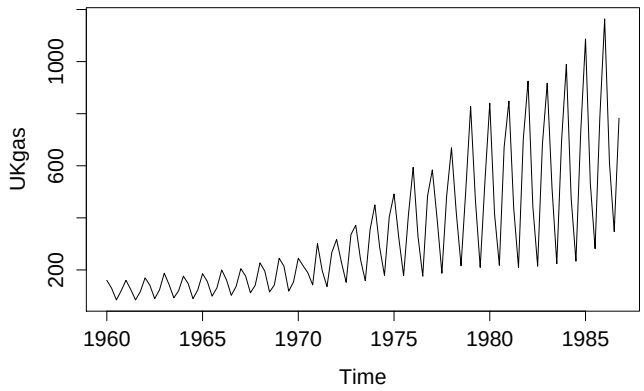


Figure 13: Quarterly UK gas consumption from 1960 to 1986, in millions of therms.

We employ a model built on a local linear trend plus a quarterly seasonal component DLM to analyze the data, that is, we consider $\mathbf{F}_t = (1, 0, 1, 0, 0)'$ and $\mathbf{G}_t = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$, for all $t = 1, \ldots, 108$, in model (4).

We fit the DQLM for $\tau = 0.10, 0.50, 0.90$ to the real dataset on the log scale. Figure 14 provides a plot of the posterior mean with the MCMC algorithm (in black) and approximated method (in blue) of the trend and seasonal components, together with 95% credibility intervals using MCMC. It is possible to observe the change in the seasonal factor around the third quarter of 1970. Moreover, in terms of the posterior mean both methods are very close to each other.

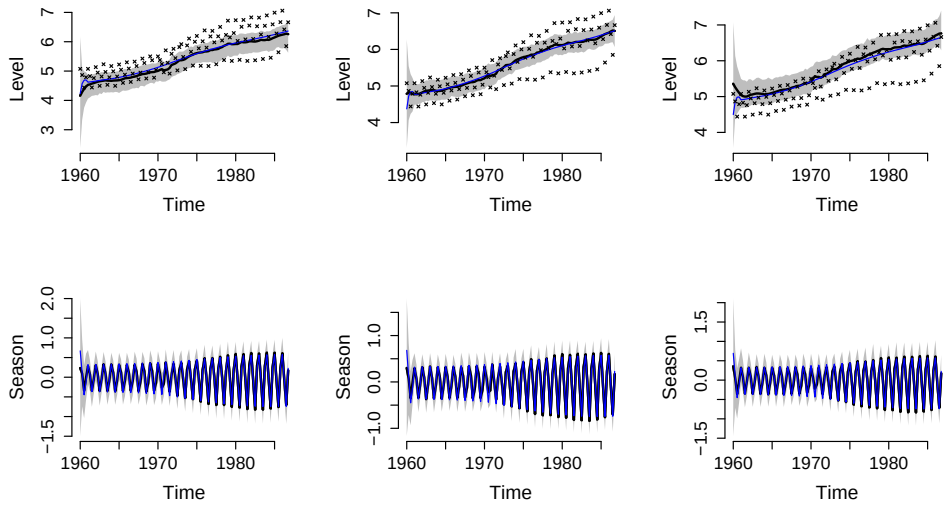


Figure 14: Posterior mean (solid line), 95% credible region (shaded area) for the level and seasonal components for each quantile.