Inference on dynamic models for non-Gaussian random fields using INLA

R. X. Cortes[a], T. G. Martins[b], M. O. Prates[a] and B. A. Silva[c]

[a]*Department of Statistics, Universidade Federal de Minas Gerais*
[b]*Department of Mathematical Sciences, Norwegian University of Science and Technology*
[c]*Department of Sociology, Universidade Federal de Minas Gerais*

Abstract. Robust time series analysis is an important subject in statistical modeling. Models based on Gaussian distribution are sensitive to outliers, which may imply in a significant degradation in estimation performance as well as in prediction accuracy. State-space models, also referred as Dynamic Models, is a very useful way to describe the evolution of a time series variable through a structured latent evolution system. Integrated Nested Laplace Approximation (INLA) is a recent approach proposed to perform fast approximate Bayesian inference in Latent Gaussian Models which naturally comprises Dynamic Models. We present how to perform fast and accurate non-Gaussian dynamic modeling with INLA and show how these models can provide a more robust time series analysis when compared with standard dynamic models based on Gaussian distributions. We formalize the framework used to fit complex non-Gaussian space-state models using the R package INLA and illustrate our approach with a simulation study and a Brazilian homicide rate dataset.

1 Introduction

Robust estimation of time series analysis is an important and challenging field of statistical application from either frequentist (Bustos and Yohai, 1979, Denby and Martin, 1979) or Bayesian perspectives (West, 1981). Robust methods are useful when there is a small percentage of data that do not follow a proposed model. Dynamic Linear Models (DLM) and Generalized Dynamic Linear Models (DGLM), also referred as state-space models, are a broad class of parametric models that generalizes regression and time series models with time varying parameters, where both the parameter variation and the observed data are described in an evolutionary structured way (Migon et al., 2005). Dynamic models are composed by an observational equation and one or more system equations in which the error terms are usually chosen to follow a Gaussian distribution. However, it is well known that the Gaussian distribution is very sensitive to outliers, which may produce degradation in the estimation performance (Fox, 1972). Therefore, one might be interested in

building a more flexible model based on heavy-tailed distributions rather than the usual Gaussian. Such models fall into the class of *non-Gaussian dynamic models* (see Kitagawa (1987), Durbin and Koopman (2000) for a detailed description as well as applications of this class of models).

Integrated Nested Laplace Approximation (INLA) is an approach proposed by Rue, Martino and Chopin (2009) to perform approximate fully Bayesian inference in the class of latent Gaussian models (LGMs). LGMs is a broad class and include many of the standard models currently in use by the applied community, e.g., stochastic volatility, disease mapping, log-Gaussian Cox process and generalized linear models. As opposed to the simulation-based methods, like Markov Chain Monte Carlo (MCMC), INLA performs approximate inference using a series of deterministic approximations that take advantage of the LGM structure to provide fast and accurate approximations. Moreover, it avoids known problems with commonly used simulation-based methods, e.g., difficulty in diagnosing convergence, additive Monte Carlo errors, and high demand in terms of computational time. Even for dynamic models within the class of LGMs, it was not possible to fit most of them using the available tools in the `INLA` package for `R`, hereafter denoted as `R-INLA`. Ruiz-Cárdenas, Krainski and Rue (2012) presented a general framework which enabled users to use `R-INLA` to perform fully Bayesian inference for a variety of state-space models. However, their approach does not include the class of *non-Gaussian dynamic models* where the errors of the system equations have a non-Gaussian distribution as, for example, the heavy-tailed distributions.

One of the key assumptions of the INLA approach is that the latent field follows a Gaussian distribution. However, Martins and Rue (2014) have shown a way to extend INLA to cases where some independent components of the latent field have a non-Gaussian distribution. Their approach transfer the non-Gaussianity of the latent field to the likelihood function and it has shown to produce satisfactory results as long as this distribution is not far from Gaussian. Distributions that add flexibility around a Gaussian as *near-Gaussian* distributions are referred as being, for example, unimodal and symmetric.

The contribution of this paper is three folded: (1) To extend INLA for non-Gaussian latent models with dependency structure, specifically for non-Gaussian DLMs; (2) To present in a simple manner how to use `R-INLA` to perform non-Gaussian DLMs modelling.

To accomplish these issues, we introduce a reparametrization of the non-Gaussian DLM and combine it with the computational framework provided by `R-INLA` to introduce how to model dependent non-Gaussian latent field in the `R-INLA` setup; (3) We analyze the Brazilian homicides rates using a robust approach. The analysis indicates that, in most of our application scenarios, the robust method outperforms the traditional Gaussian approach.

The paper is organized as following: Section 2 introduces the methodology of our approach, presenting how to perform fast Bayesian inference using `R-INLA` for non-Gaussian DLMs. Section 3 presents our simulation study to compare two

competitor models using some quality measures. In Section 4, we present the study over the Brazilian Homicide data, explaining our findings. Finally, in Section 5 we discuss some final remarks and future research.

2 Methodology

This section will describe our approach to handle non-Gaussian DLM within `R-INLA`. Although valid for DGLM, we have chosen to illustrate our extension using a DLM to facilitate the presentation. To apply the extension for DGLM, a simple change in the Gaussian likelihood is necessary, which is a trivial modification under `R-INLA`. Section 2.1 will define a general DLM of interest, and show that it fits the class of LGM only if the error terms of the system equations are Gaussian distributed. Section 2.2 will review the INLA methodology, including the recent extension that allows INLA to be applied to models where some components of the latent field have non-Gaussian distribution. Section 2.3 gives an overview of a generic approach to fit dynamic models using `R-INLA` through an augmented model structure. Finally, Section 2.4 extend the approaches presented in Sections 2.2 and 2.3 and show how this extension can be exploited to fit non-Gaussian DLM within `R-INLA`.

2.1 Models

The INLA approach performs approximate Bayesian inference in latent Gaussian models where the first stage is formed by the likelihood function with conditional independence properties given the latent field $\mathbf{x}$ and possible hyperparameters $\boldsymbol{\theta}_1$, where each data point $\{y_t, t = 1, \dots, n_d\}$ is connected to one element in the latent field x_t. In this context, the latent field $\mathbf{x}$ is formed by linear predictors, random and fixed effects, depending on the model formulation. Assuming that the elements of the latent field connected to the data points, that is, the linear predictors $\{\eta_t, t = 1, \dots, n_d\}$, are positioned on the first n_d elements of $\mathbf{x}$, we have:

- *Stage* 1. $\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}_1 \sim \pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}_1) = \prod_{t=1}^{n_d} \pi(y_t|x_t, \boldsymbol{\theta}_1)$.

The conditional distribution of the $\mathbf{x}$ given some possible hyperparameters $\boldsymbol{\theta}_2$ forms the second stage of the model and has a joint Gaussian distribution,

- *Stage* 2. $\mathbf{x}|\boldsymbol{\theta}_2 \sim \pi(\mathbf{x}|\boldsymbol{\theta}_2) = \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}(\boldsymbol{\theta}_2), \mathbf{Q}^{-1}(\boldsymbol{\theta}_2))$,

where $\mathcal{N}(\cdot; \boldsymbol{\mu}, \mathbf{Q}^{-1})$ denotes a multivariate Gaussian distribution with mean vector $\boldsymbol{\mu}$ and a precision matrix $\mathbf{Q}$. In most applications, the latent Gaussian field have conditional independence properties, which translates into a sparse precision matrix $\mathbf{Q}(\boldsymbol{\theta}_2)$, which is of extreme importance for the numerical algorithms used by INLA. The latent field $\mathbf{x}$ may have additional linear constraints of the form $\mathbf{A}\mathbf{x} = \mathbf{e}$ for an $q \times n_d$ matrix $\mathbf{A}$ of rank q, where q is the number of constraints and n_d

the size of the latent field. The hierarchical model is then completed with an appropriate prior distribution for the m-dimensional hyperparameter of the model $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$:

- *Stage* 3. $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$.

The structure of a non-Gaussian DLM is composed by an observation equation describing the relationship between the observations $\mathbf{y}$ $\{y_t; t = 1, \ldots, n_d\}$, which are connected to a linear combination of the state parameters $\mathbf{a}$ $\{a_t; t = 1, \ldots, n_d\}$, and a system of equations describing the evolution of $\mathbf{a}$. For example:

$$y_t = a_t + v_t, \qquad v_t \sim N(0, \theta_1),$$
$$a_t = a_{t-1} + w_t, \qquad w_t \sim \pi(\cdot)$$

in which a_t is the state vector at time t, θ_1 is the Gaussian variance of v_t and the noises w_t could follow a non-Gaussian distribution. We emphasize that the structure described above could be more flexible allowing any linear combination and addition of covariates. Furthermore, this is an extension over the traditional DLM where we now can have a non-Gaussian distribution for the noise w_t in the system equation.

If w_t is assumed to be Gaussian, this structure falls naturally into the class of LGMs (see Section 2.3). To help understand the INLA review of Section 2.2, we can rewrite a LGM using a hierarchical structure with three stages. To elucidate the understanding of notation in our examples, we highlight that state vector $\mathbf{a}$ does not necessarily corresponds to the latent field $\mathbf{x}$. Since our approach lies in a augmented likelihood function, the dimension of the latent field $\mathbf{x}$ is larger than the dimension of $\mathbf{y}$ and $\mathbf{a}$.

However, if a Gaussian distribution is not assumed for w_t, it is no longer possible to write the model as a hierarchical structure with the Gaussian assumption in the second stage. To accommodate the non-Gaussian DLM, it is necessary to expand the class of LGMs defined early to allow that nodes of the latent field have non-Gaussian distributions. We then rewrite stage 2 of the hierarchical model as

- *Stage* 2^{new}. $\underbrace{(\mathbf{x}_{\text{G}}, \mathbf{x}_{\text{NG}})}_{\mathbf{x}} |\boldsymbol{\theta}_2 \sim \pi(\mathbf{x}|\boldsymbol{\theta}_2) = \mathcal{N}(\mathbf{x}_{\text{G}}; \mathbf{0}, \mathbf{Q}^{-1}(\boldsymbol{\theta}_2)) \times \prod_t \pi(\mathbf{x}_{\text{NG}_t}|\boldsymbol{\theta}_2),$

where $\mathbf{x}_{\text{G}}$ and $\mathbf{x}_{\text{NG}}$ represent the Gaussian and independent non-Gaussian terms of the latent field, respectively. As a result, the distribution of the latent field is not Gaussian, which precludes the use of INLA to fit this class of models.

Section 2.2 summarizes how to perform inference, within the R-INLA framework, on models where the non-Gaussian components of the latent field belong to the class of near-Gaussian distributions. Later, in Section 2.4 we introduce how to perform inference when the non-Gaussian components have a dependent structure, specifically belonging to the class of non-Gaussian DLMs.

2.2 INLA review

Using the hierarchical representation of LGMs given in Section 2.1, we have that the joint posterior distribution of the unknowns is

$$\pi(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y}) \propto \pi(\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta}) \prod_{t=1}^{n_d} \pi(y_t|x_t, \boldsymbol{\theta})$$

$$\propto \pi(\boldsymbol{\theta})|\mathbf{Q}(\boldsymbol{\theta})|^{n/2} \exp\left[-\frac{1}{2}\mathbf{x}^T\mathbf{Q}(\boldsymbol{\theta})\mathbf{x} + \sum_{t=1}^{n_d} \log\{\pi(y_t|x_t, \boldsymbol{\theta})\}\right].$$

The approximated posterior marginals of interest $\tilde{\pi}(x_t|\mathbf{y})$, $t = 1, \ldots, n_d$ and $\tilde{\pi}(\theta_j|\mathbf{y})$, $j = 1, \ldots, m$ returned by INLA have the following form

$$\tilde{\pi}(x_t|\mathbf{y}) = \sum_u \tilde{\pi}\left(x_t|\boldsymbol{\theta}^{(u)}, \mathbf{y}\right)\tilde{\pi}\left(\boldsymbol{\theta}^{(u)}|\mathbf{y}\right)\Delta\boldsymbol{\theta}^{(u)}, \tag{2.1}$$

$$\tilde{\pi}(\theta_s|\mathbf{y}) = \int \tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})\, d\boldsymbol{\theta}_{-s}, \tag{2.2}$$

where $\{\tilde{\pi}(\boldsymbol{\theta}^{(u)}|\mathbf{y})\}$ are the density values computed during a grid exploration on $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$, for given approximations of $\pi(x_t|\boldsymbol{\theta}, \mathbf{y})$ and $\pi(\boldsymbol{\theta}|\mathbf{y})$.

Looking at equations (2.1)–(2.2), we can see that the method can be divided into three main tasks. First, propose an approximation $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ to the joint posterior of the hyperparameters $\pi(\boldsymbol{\theta}|\mathbf{y})$, second propose an approximation $\tilde{\pi}(x_t|\boldsymbol{\theta}, \mathbf{y})$ to the marginals of the conditional distribution of the latent field given the data and the hyperparameters $\pi(x_t|\boldsymbol{\theta}, \mathbf{y})$ and last explore $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ on a grid and use it to integrate out $\boldsymbol{\theta}$ in equation (2.1) and $\boldsymbol{\theta}_{-j}$ in equation (2.2).

The approximation used for the joint posterior of the hyperparameters $\pi(\boldsymbol{\theta}|\mathbf{y})$ is

$$\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) \propto \frac{\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\pi_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})}\Big|_{\mathbf{x}=\mathbf{x}^*(\boldsymbol{\theta})}, \tag{2.3}$$

where $\pi_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ is a Gaussian approximation to the full conditional of $\mathbf{x}$, and $\mathbf{x}^*(\boldsymbol{\theta})$ is the mode of the full conditional for $\mathbf{x}$, for a given $\boldsymbol{\theta}$. The full conditional of the latent field when dealing with LGMs is given by

$$\pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y}) \propto \exp\left\{-\frac{1}{2}\mathbf{x}^T\mathbf{Q}(\boldsymbol{\theta})\mathbf{x} + \sum_{t\in\mathcal{T}} g_t(x_t)\right\}, \tag{2.4}$$

where $\mathcal{T}$ is an index set and $g_t(x_t) = \log\pi(y_t|x_t, \boldsymbol{\theta_1})$. The Gaussian approximation used by INLA is obtained by matching the modal configuration and the curvature at the mode. The good performance of INLA is highly dependent on the appropriateness of the Gaussian approximation in equation (2.4) and this turns out to be the case when dealing with LGMs because the Gaussian prior assigned to the latent field has a non-negligible effect on the full conditional, specially in terms of shape and correlations. Besides, the likelihood function is usually well behaved and not

very informative on **x**. It is very important to note that equation (2.3) is equivalent to the Laplace approximation of a marginal posterior distribution (Tierney and Kadane, 1986), and it is exact if $\pi(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})$ is Gaussian, in which case INLA gives exact results up to small integration error due to the numerical integration of equations (2.1) and (2.2).

For approximating $\pi(x_t|\boldsymbol{\theta}, \mathbf{y})$, three options are available in R-INLA. The so called Laplace, Simplified Laplace and Gaussian which are ordered in terms of accuracy. We refer to Rue, Martino and Chopin (2009) for a detailed description of these approximations and Martins et al. (2012) on how to compute equation (2.2) efficiently.

Martins and Rue (2014) have demonstrated how INLA can be used to perform inference in latent models where some independent components of the latent field have a non-Gaussian distribution, in which case the latent field is no longer Gaussian. Their approach approximates the distribution of the non-Gaussian components $\pi(\mathbf{x}_{\mathrm{NG}}|\boldsymbol{\theta}_2)$ by a Gaussian distribution $\pi_{\mathrm{G}}(\mathbf{x}_{\mathrm{NG}}|\boldsymbol{\theta}_2)$ and corrects this approximation with the correction term

$$\mathrm{CT} = \pi(\mathbf{x}_{\mathrm{NG}}|\boldsymbol{\theta}_2)/\pi_{\mathrm{G}}(\mathbf{x}_{\mathrm{NG}}|\boldsymbol{\theta}_2)$$

in the likelihood. Taking into consideration the above approximation and correction term, we can rewrite our latent model with the following hierarchical structure:

- *Stage* 1. $\mathbf{z}|\mathbf{x}, \boldsymbol{\theta} \sim \pi(\mathbf{z}|\mathbf{x}, \boldsymbol{\theta}) = \prod_{t=1}^{n_d+k} \pi(z_t|x_t, \boldsymbol{\theta})$, where

$$\pi(z_t|x_t, \boldsymbol{\theta}) = \begin{cases} \pi(y_t|x_t, \boldsymbol{\theta}_1), & \text{for } 1 \le t \le n_d, \\ \pi(x_{\mathrm{NG}_t}|\boldsymbol{\theta}_2)/\pi_{\mathrm{G}}(x_{\mathrm{NG}_t}|\boldsymbol{\theta}_2), & \text{for } n_d < t \le n_d + k \end{cases}$$

and **z** is an augmented response vector with $z_t = y_t$ if $t \le n_d$ and $z_t = 0$ if $n_d < t \le n_d + k$, where k is the length of $\mathbf{x}_{\mathrm{NG}}$. It is important to emphasize that Stage 1 above is not the likelihood function, but expressing the model using this form makes the practical definition of the non-Gaussian latent model within the R-INLA framework easier to understand.

The latent field has now a Gaussian approximation replacing the non-Gaussian distribution of $\mathbf{x}_{\mathrm{NG}}$,

- *Stage* 2. $\underbrace{(\mathbf{x}_{\mathrm{G}}, \mathbf{x}_{\mathrm{NG}})}_{\mathbf{x}}|\boldsymbol{\theta}_2 \sim \pi(\mathbf{x}|\boldsymbol{\theta}_2) = \mathcal{N}(\mathbf{x}_{\mathrm{G}}; \mathbf{0}, \mathbf{Q}^{-1}(\boldsymbol{\theta}_2)) \times \pi_{\mathrm{G}}(\mathbf{x}_{\mathrm{NG}}|\boldsymbol{\theta}_2),$

which means that $\pi(\mathbf{x}|\boldsymbol{\theta}_2)$ is now Gaussian distributed.

Martins and Rue (2014) have shown that the main impact of this strategy occurs in the Gaussian approximation to the full conditional of the latent field that now takes the form

$$\pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y}) \propto \exp\left\{ -\frac{1}{2}\mathbf{x}^T \mathbf{Q}(\boldsymbol{\theta})\mathbf{x} + \sum_{t=1}^{n_d} g_t(x_t) + \sum_{t=n_d+1}^{n_d+k} h_t(x_t) \right\}, \qquad (2.5)$$

where $g_t(x_t) = \log \pi(y_t|x_t, \boldsymbol{\theta})$ as before and

$$h_t(x_t) = \log \mathrm{CT}_t = \log \pi(\mathbf{x}_{\mathrm{NG}_t}|\boldsymbol{\theta}_2) - \log \pi_{\mathrm{G}}(\mathbf{x}_{\mathrm{NG}_t}|\boldsymbol{\theta}_2).$$

The key for a good accuracy of INLA depends on the behavior of $h_t(x_t)$ which is influenced by the distribution $\pi(\mathbf{x}_{\mathrm{NG}}|\boldsymbol{\theta}_2)$ of the non-Gaussian components and by the Gaussian approximation $\pi_{\mathrm{G}}(\mathbf{x}_{\mathrm{NG}_t}|\boldsymbol{\theta}_2)$ to this non-Gaussian distribution. Also good results are obtained when $\pi_{\mathrm{G}}(\mathbf{x}_{\mathrm{NG}_t}|\boldsymbol{\theta}_2)$ is chosen to be a zero mean and low precision Gaussian distribution such that

$$\pi_{\mathrm{G}}(\mathbf{x}_{\mathrm{NG}_t}|\boldsymbol{\theta}_2) \propto \text{constant}$$

and $\pi(\mathbf{x}_{\mathrm{NG}_t}|\boldsymbol{\theta}_2)$ is not too far away from a Gaussian, for which they coined the term *near-Gaussian* distributions. This means that the application of INLA within the context of non-Gaussian DLM will yield accurate results as long as these components are distributed according to a flexible distribution around the Gaussian, as in the Student's t case for example, which is unimodal and symmetric.

2.3 `R-INLA` for DLM

In this section, we present a simple dynamic model to illustrate the framework to perform fast Bayesian inference within `R-INLA`. The INLA approach could be used to estimate any dynamic structure that could be written as a latent Gaussian model described in Section 2.1, however the approach presented here is motivated to overcome some limitations of `R-INLA`. Suppose as a Toy example the following first order univariate dynamic linear model

$$y_t = a_t + v_t, \qquad v_t \sim N(0, \theta_1), t = 1, \ldots, n_d, \qquad (2.6)$$

$$a_t = a_{t-1} + w_t, \qquad w_t \sim N(0, \theta_2), t = 2, \ldots, n_d. \qquad (2.7)$$

It is possible to fit the model given by equations (2.6) and (2.7) using the standard latent models available in `R-INLA` and we are aware that the corresponding model could be estimated through the well-known Kalman Filter (Kalman, 1960). However, this simple model is useful to illustrate the framework used in this paper, which allow us to fit more complex dynamic models that would otherwise not be available through `R-INLA`. The presented approach involves an augmented model structure in which the system equations are treated as observation equations.

The key step is to equate to zero the system equations of the state-space model, so that

$$0 = a_t - a_{t-1} - w_t, \qquad w_t \sim N(0, \theta_2), t = 2, \ldots, n_d. \qquad (2.8)$$

Then it is possible to build an augmented model by merging the "faked zero observations" from equation (2.8) to the actual observations $\{y_t, t = 1, \ldots, n_d\}$ of equation (2.6). In addition, the "faked observations" are assumed to follow a Gaussian distribution with high and fixed precision to represent the fact that those artificial

observations are deterministically known. Instead of using this Gaussian distribution with high and fixed precision and mean given by $\phi_t = a_t - a_{t-1} - w_t$ for the artificial observations, as in Ruiz-Cárdenas, Krainski and Rue (2012), we use in what follows a Gaussian with variance θ_2 and mean $\eta_t^* = a_t - a_{t-1}$. This is an equivalent representation and will make it easier to describe in Section 2.4 the extension of this approach to dynamic models with non-Gaussian error terms in the system equations.

To complete the model definition, note that there is no information about a_t beyond the temporal evolution given by equation (2.7), and so we only need to know the perturbations w_t, $t = 2, \ldots, n_d$ to estimate the states a_t, since $\{w_t\}$ are the only stochastic term in system equation (equation (2.7)). This characteristic of dynamic models allow to represent the dependence structure as a function of the independent perturbation terms. To represent this within R-INLA, let $\mathbf{a} = \{a_1, \ldots, a_{n_d}\}$ be formed by independent random variables each following a Gaussian distribution with fixed and low precision and encode the temporal evolution present in equation (2.8) using the copy feature available in R-INLA (Martins et al., 2012). Finally, inverse-gamma priors are assigned to the variances θ_1 and θ_2. The reason to use this augmented model is that it allows us to encode the dynamic evolution of equation (2.7) using standard generic tools available in R-INLA, instead of requiring the implementation of a different dynamic structure for each possible type of dynamic model.

2.4 `R-INLA` for non-Gaussian DLM

We now present how to perform fast Bayesian inference on non-Gaussian DLM through the R-INLA package. We first formalize the augmented model described in Section 2.3 and the likelihood correction described in Section 2.2 in this framework. We then show how our approach can be exploited to fit non-Gaussian DLM using R-INLA. The results of formalizing our approach overcomes the limitation assumption of independence for the non-Gaussian components in the latent field and, moreover, generalizes the DLM class of models.

The augmented model approach described in Section 2.3 can be represented using a hierarchical framework. Similar to Section 2.2, assume we have an augmented response vector $\mathbf{z}$ with $z_t = y_t$ if $t \leq n_d$ and $z_t = 0$ if $n < t \leq 2n_d - 1$ and

- *Stage* 1. $\mathbf{z}|\mathbf{x}, \boldsymbol{\theta} \sim \pi(\mathbf{z}|\mathbf{x}, \boldsymbol{\theta}) = \prod_{t=1}^{2n_d-1} \pi(z_t|x_t, \boldsymbol{\theta})$, where

$$\pi(z_t|x_t, \boldsymbol{\theta}) = \begin{cases} \pi(y_t|x_t, \theta_1), & \text{for } 1 \leq t \leq n_d, \\ \pi(z_t|x_t, \theta_2), & \text{for } n < t \leq 2n_d - 1 \end{cases} \tag{2.9}$$

with $\pi(y_t|x_t, \theta_1) \stackrel{d}{=} \mathcal{N}(y_t; x_t, \theta_1)$ and $\pi(z_t|x_t, \theta_2) \stackrel{d}{=} \mathcal{N}(0; x_t, \theta_2)$. Note that, as mentioned in Section 2.3, we have used a Gaussian distribution with variance given by θ_2 as the likelihood for the artificial zero observations. Internally, for R-INLA, the $(4n_d - 1)$-dimensional latent field is defined as

- *Stage* 2. $\mathbf{x} = (\eta_1, \ldots, \eta_{n_d}, \eta_{n_d+1}, \ldots, \eta_{2n_d-1}, a_1, \ldots, a_{n_d})$,

where $\mathbf{a}$ is given independent Gaussian priors with low and fixed precision, $\eta_t = a_t + s_t$ is the linear predictor connected to the observation y_t, for $t = 1, \ldots, n_d$ and $\eta_t = a_t - a_{t-1} + s_t$ is the linear predictor connected with the artificial zero observations, for $t = n_d + 1, \ldots, 2n_d - 1$, and s_t is a small noise represented by a Gaussian distribution with zero mean and high and fixed precision to eliminate a rank deficiency in the above representation of $\mathbf{x}$. Finally, priors are assigned to the hyperparameters of the model:

- *Stage* 3. $\theta_1 \sim \text{IG}(a_v, b_v), \theta_2 \sim \text{IG}(a_w, b_w)$.

By comparing this hierarchical representation with the likelihood correction approach described in Section 2.2, we note that we are approximating the distribution of the state vector $\mathbf{a} = \{a_t, t = 1, \ldots, n_d\}$, defined by equation (2.7), which is originally given by a Gaussian with precision matrix $Q_a = \theta_2^{-1}\mathbf{R}$, with

$$\mathbf{R} = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 2 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 & -1 & 0 \\ 0 & 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & 0 & \cdots & 0 & -1 & 1 \end{pmatrix},$$

by a very low precision independent Gaussian distribution. In the correction approach

$$\pi(\mathbf{a}) \propto \text{constant}$$

and this approximation is corrected in the likelihood function by adding the following correction term

$$\text{CT} = \prod_{t=n_d+1}^{2n_d-1} \pi(z_t | x_t, \theta_2)$$

with $\pi(z_t | x_t, \theta_2)$ defined in equation (2.9). Note that this representation also corresponds to those "faked zero observations" of equation (2.8). once we have identified this, observe that the log likelihood $g_t(x_t)$ and the log correction term $h_t(x_t)$ in equation (2.5) both have quadratic forms, which implies that the full conditional of the latent field $\pi(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})$ is Gaussian distributed, meaning that R-INLA gives exact results up to a small integration error, as mentioned in Section 2.2.

Next, assume the following non-Gaussian DLM,

$$y_t = a_t + v_t, \qquad v_t \sim N(0, \theta_1), t = 1, \ldots, n_d, \tag{2.10}$$

$$a_t = a_{t-1} + w_t, \qquad w_t \sim t(0, \tau, \nu), t = 2, \ldots, n_d, \tag{2.11}$$

which can be written in a hierarchical structure

$$y_t | a_t, \theta_1 \sim N(a_t, \theta_1),$$

$$a_t | a_{t-1}, \tau, \nu \sim t(a_{t-1}, \tau, \nu),$$

$$\tau \sim \pi(\tau), \qquad \nu \sim \pi(\nu), \qquad \theta_1 \sim \pi(\theta_1)$$

highlighting the fact that the latent field is no longer Gaussian. Note that the $\boldsymbol{\theta_2} = (\tau, \nu)$ and the distribution of v_t in equation (2.10) could have a non-Gaussian distribution as well, since non-Gaussian likelihood functions are already standard in R-INLA, but using a Gaussian here makes the final impact of the non-Gaussianity of w_t more easily visible and analyzed in Section 3. As mentioned in Section 1, the motivation of using heavier tailed distributions such as Student-t in the noise of the latent system is to robustify the model. Robustifying the model means that the dynamic system is less sensitive to different types of outliers. By allowing this higher flexibility of w_t we can better handle what is called *innovative outliers* in time series literature (Fox, 1972, Masreliez and Martin, 1977, Mc Quarrie and Tsai, 2003).

By a similar argument made in Section 2.3, we note that we only need to know the stochastic terms $\{w_t, t = 2, \ldots, n_d\}$ and the system dynamics in equation (2.11) to estimate $\mathbf{a} = (a_1, \ldots, a_{n_d})$. Consequently, if we include those pieces of information in the likelihood function through a correction term, we can assign independent Gaussian priors with zero mean and low and fixed precisions for $\mathbf{a}$, which will lead to the following hierarchical model:

- *Stage* 1. $\mathbf{z} | \mathbf{x}, \boldsymbol{\theta} \sim \pi(\mathbf{z} | \mathbf{x}, \boldsymbol{\theta}) = \prod_{t=1}^{2n_d - 1} \pi(z_t | x_t, \boldsymbol{\theta})$, where

$$\pi(z_t | x_t, \boldsymbol{\theta}) = \begin{cases} \pi(y_t | x_t, \theta_1), & \text{for } 1 \leq t \leq n_d, \\ \pi(z_t | x_t, \tau, \nu), & \text{for } n_d < t \leq 2n_d - 1 \end{cases}$$

with $\pi(y_t | x_t, \theta_1) \overset{d}{=} \mathcal{N}(y_t; x_t, \theta_1)$ and $\pi(z_t | x_t, \tau, \nu) \overset{d}{=} t(0; x_t, \tau, \nu)$.

- *Stage* 2. $\mathbf{x} = (\eta_1, \ldots, \eta_{n_d}, \eta_{n_d+1}, \ldots, \eta_{2_d n - 1}, a_1, \ldots, a_{n_d})$,

where $\mathbf{a}$ and $\boldsymbol{\eta}$ are the same as defined earlier. Finally, priors are assigned to the hyperparameters of the model:

- *Stage* 3. $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$ with $\boldsymbol{\theta} = (\theta_1, \tau, \nu)$.

We see that the hierarchical model above is very similar to the one presented by equation (2.9) and the difference is on the correction term

$$\text{CT} = \prod_{t=n_d+1}^{2n_d-1} \pi(z_t | x_t, \tau, \nu),$$

which is no longer Gaussian distributed, leading to a full conditional of the form equation (2.5) with a non-Gaussian log correction term $h_t(x_t)$. As we have showed,

this configuration fits the framework summarized in Section 2.2 and therefore we can apply the results to the context of non-Gaussian dependent latent fields, specifically DLMs. Thus, R-INLA provides accurate results for non-Gaussian DLM, as long as the non-Gaussian distribution attributed to the error terms of the system equations are not too far from a Gaussian distribution, as discussed in Section 1. This assumption is satisfied by the Student-t distribution, as well as for other distributions that corrects the Gaussian in terms of skewness and/or kurtosis.

3 Simulation study

In this section, we present the results of a Monte Carlo simulation for the Toy Example defined in Section 2.4 (see equations (2.10) and (2.11)) to better understand the benefits of fitting a non-Gaussian DLM with INLA. Moreover, we investigate the property of different model selection criteria available from R-INLA in this context. We have chosen to perform a contamination study similar to the ones presented in Pinheiro, Liu and Wu (2001) and in Martins and Rue (2014) where the noise w_t from equation (2.11) is contaminated with the following mixture of Gaussian distributions

$$w_t \sim (1 - p) \times N(0, \theta_2) + p \times f \times N(0, \theta_2), \qquad t = 1, \ldots, n_d,$$

where p is the expected percentage of *innovative outliers* in the latent system and f is a fixed value indicating the magnitude of the contamination. We have generated all possible scenarios with $n_d = 100, 250, 500$, $p = 0, 0.05, 0.1, 0.15, 0.20, 0.25$ and $f = 2, 4, 8$, resulting in a total of 54 different scenarios. For each of them, 1000 datasets were simulated and analyzed. The true variance parameter of the observational and system noises are set to $\theta_1, \theta_2 = 2$.

In R-INLA, the Student's t likelihood is parametrized in terms of its marginal precision τ and degrees of freedom ν. This is advantageous because the precision parameter under the Gaussian and the Student's t distribution possess the same interpretation allowing the same prior to be used for τ whether we refer to the Gaussian or to the Student's t model. In this Monte Carlo experiment, we have used a Gamma[1] prior with shape and rate parameters given by 1 and 2.375 for both the observational and system noise precision parameters. The prior for ν is based on the framework of (Martins et al., 2014). In their context, the prior is design for the flexibility parameters, which in this case is the degrees of freedom ν, in such way that the basic model plays a central role in the more flexible one. In our context, it means that the prior for the degrees of freedom is constructed such that the mode of the prior happens to be in the value that recovers the Gaussian model and deviations from the Gaussian model are penalized based on the distance between the basic and the flexible model. The prior specification consists in the

---

[1]If $X \sim \text{Gamma}(a, b)$ then $E(X) = \frac{a}{b}$.

choice of the degree of flexibility (df) parameter, $0 < \mathrm{df} < 1$, which represents the percentage of prior mass attributed to the degrees of freedom between 2 and 10. We have set $\mathrm{df} = 0.3$ in our applications. We refer to (Martins et al., 2014) for more details about priors for flexibility parameters.

As mentioned in the Introduction, a model based on the student's t distribution is expected to be more robust with respect to outliers in a contaminated data setup when compared to a similar model based on Gaussian distributions. To assess the gain in performance of the more flexible model based on the student's t distribution, we will compute the mean squared error (MSE), the conditional predictive ordinate (CPO) (Gelfand, Dey and Chang, 1992, Dey, Chen and Chang, 1997) and deviance information criteria (DIC) (Spiegelhalter et al., 2002). The intuition behind the CPO criterion is to choose a model with higher predictive power measured in terms of predictive density.

For the jth simulated dataset of a given scenario, let a_{tj} be the true latent variable at time t. We will denote by $\widehat{a}_{tj,G}$ and $\widehat{a}_{tj,T}$ the posterior mean of a_{tj} computed by the Gaussian and student's t model, respectively. The student-t model efficiency over the Gaussian one to estimate a_{tj} for each dataset j is defined by

$$E_j = \frac{\sum_{t=1}^{n_d}(\widehat{a}_{tj,G} - a_{tj})^2}{\sum_{t=1}^{n_d}(\widehat{a}_{tj,T} - a_{tj})^2} - 1,$$

which can be viewed as ratio of the respective MSEs centered at 0.

Figure 1 represents the median over $\{E_j, j = 1, \ldots, 1000\}$ for each scenario. The results were as expected. There were slight efficiency improvements for close contamination patterns while the efficiency gains become larger as we move to
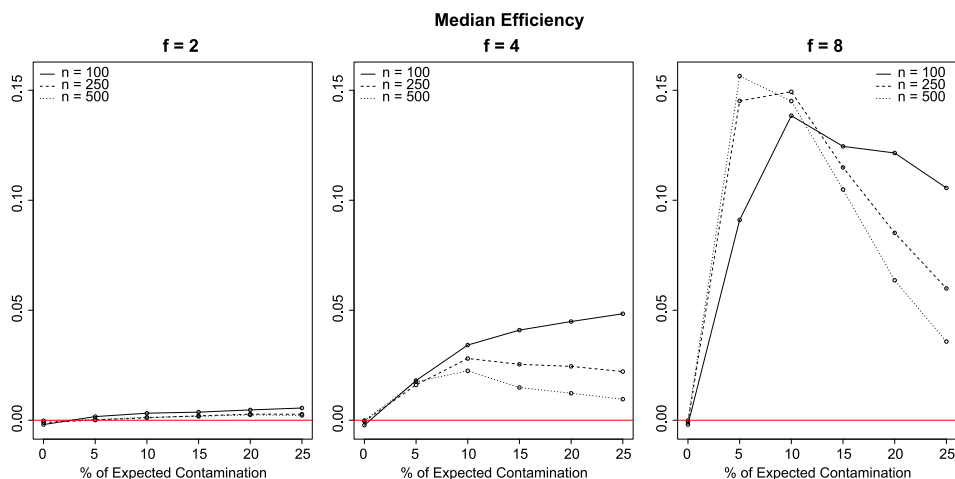

Figure 1 *Median of efficiencies for magnitude $f = 2$ (left), $f = 4$ (center) and $f = 8$ (right), $n_d = 100$ (solid line), $n_d = 250$ (dashed line) and $n_d = 500$ (dotted line). We have the percentage of expected contamination in the x-axis and the median of efficiency in the y-axis.*

higher contamination patterns, reaching efficiency gains greater than 15% for some critical scenarios. The efficiency gains are higher for moderate expected contamination percentage, around 10% in our case, and this non-monotonic behavior can be explained by the fact that once the data becomes too much contaminated, not even the more flexible model based on the student's t distribution can continue to give increasingly better results when compared to the Gaussian model, although the more flexible one continues to improve upon it.

To compare the model fitting, we use the DIC as well as the CPO criteria. First, we define the relative DIC (RDIC) as

$$\text{RDIC}_j = \frac{\text{DIC}_{Gj} - \text{DIC}_{tj}}{\text{DIC}_{tj}}, \tag{3.1}$$

for each one of the simulated data, $j = 1, \ldots, 1000$. In the top part of Figure 2, we plot the median of RDIC values obtained by the fitted Gaussian model and by fitted the student's t model for each scenario. From this figure, we observe the same pattern of Figure 1.

The summary statistic provided by the CPO criteria is called logarithm of the pseudo marginal likelihood (LPML) which evaluates the predictive power of a model. Therefore, to compare both models the LPML difference is used. To make it comparable to other goodness-of-fit measures, for example, DIC, we define the -LPML by

$$\text{-LPML}_j = -\left(\sum_{i=1}^{n_d} \log\{\pi(y_i|\mathbf{y}_{-i})\} \right)_j,$$

where j is the jth dataset in a given scenario. In this definition, lower values of -LPML indicates better predictive power. In order to compare both approaches, we have computed the logarithm of the Pseudo Bayes Factor (lPsBF) (Geisser and Eddy, 1979) for each iteration. This measure is defined as

$$\text{lPsBF}_j = \text{-LPML}_{tj} - (\text{-LPML}_{Gj}) = \text{LPML}_{Gj} - \text{LPML}_{tj}.$$

To make the comparison equivalent to the RDIC presented in equation (3.1), we define the relative lPsBF (RPsBF) as

$$\text{RPsBF}_j = \frac{\text{LPML}_{Gj} - \text{LPML}_{tj}}{\text{LPML}_{tj}}.$$

From the bottom part of Figure 2, all conclusions from the MSE and RDIC can be applied in the context of the RPsBF measure, but the gain becomes more evident. Moreover, we can see from the bottom part of Figure 2 that when the simulated scenario is stable with low expected proportion and low contamination, the median of the RPsBF is small and not significant. However, for larger sample size and contamination it is showed that the student-t approach is preferable for most of the scenarios and highlights this choice when the magnitude of the contamination increase reaching values of this median relative difference even higher than
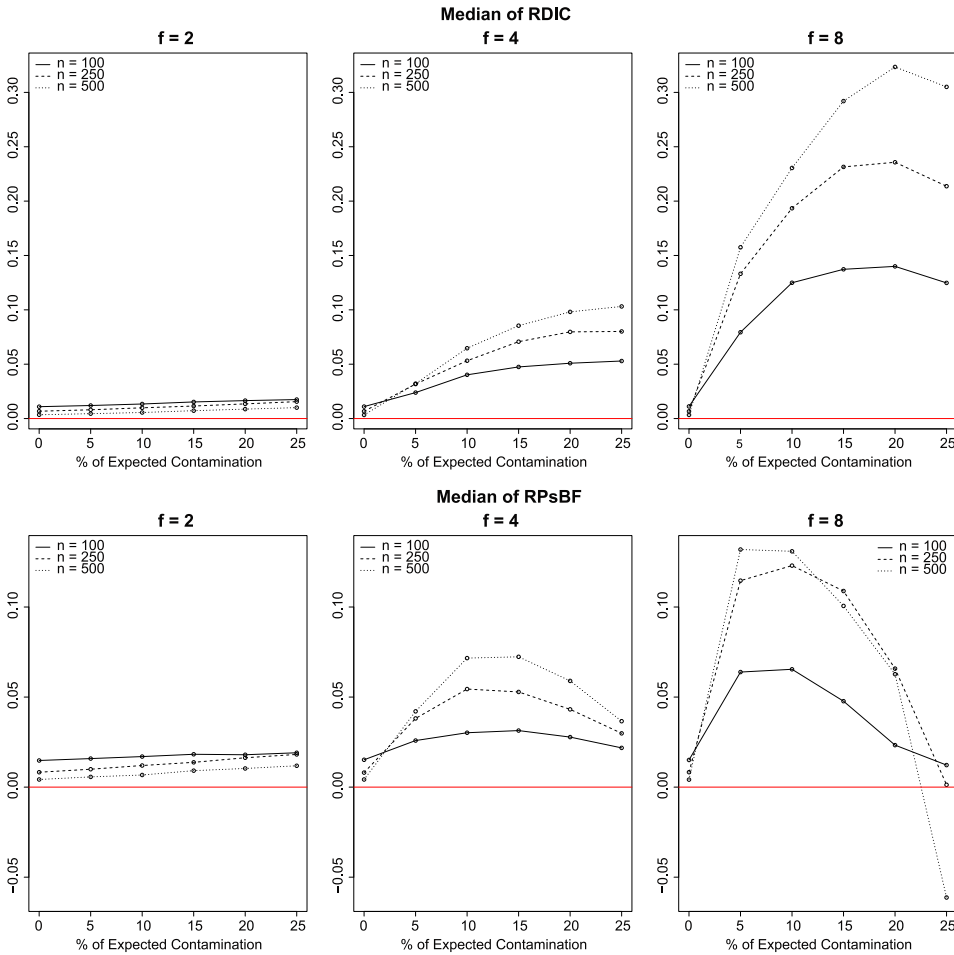
Figure 2 *Top*: *Median of RDIC, in the y-axis, for all scenarios*; *Bottom*: *Median of RPsBF, in the y-axis, for all scenarios. We have the percentage of expected contamination in the x-axis and all scenarios are*: $f = 2$ (*left*), $f = 4$ (*center*) *and* $f = 8$ (*right*), $n_d = 100$ (*solid line*), $n_d = 250$ (*dashed line*) *and* $n_d = 500$ (*dotted line*).

10% in some cases. One curious fact observed is that in the most critical scenario where $p = 0.25$, $f = 8$ and $n_d = 500$ the RPsBF values pointed incisively to the Gaussian approach, indicating that, since the generation process has too much contamination and generates to many innovative outliers, even the student-t approach is not able to control for this behavior producing predictive measures that are less accurate.

From the simulation study, we can conclude that the more flexible model is preferred over the traditional one in most of the scenarios analyzed, and the gap

between the models are higher when a moderate number of innovative outliers are involved.

4 Data application

The goal of this section is to analyze Brazilian Homicide rates with approximate Bayesian Inference for Dynamic Models using R-INLA. We have accessed data of death in Brazilian cities from 1980 to 2010 made freely available by the Brazil's public healthcare system—DATASUS.[2] This database of homicides, not natural death,[3] has a standard national structure containing age, sex, civil state, occupation, naturality, local residence of the victim and local of occurrence of the event that was used to compute the total number of homicides by city. The time series under study represents the number of events standardized by each city population. R-INLA applied to this longitudinal data offers a singular opportunity, not common in studies of this nature in Brazil, of evaluating the impact of safety public policies on crime pattern, specially on temporal pattern. In addition, numerous international studies point to the relevance of the impact assessment of crime prevention policies (Sherman and Weisburd, 1995, Sherman et al., 1997, Sherman, 1998).

Homicides studies is a broad field of sociological research (see, for example, Jacobs and Richardson (2008)). Homicides represent a specific criminal category that, although with less cases than property crimes, as burglary and robbery, it generates strong population demand for public policies of prevention. In this sense, studies that deal with the temporal dynamic of homicides try to associate, in a general way, this behavior with economic, social and political factors. For instance, high Brazilian homicide rates could be due to high levels of unemployment, poverty and economic inequality (Mir, 2004). Other factors, like age structure of population (Graham and Bowling, 1995, Flood-Page et al., 2000) and disordered population growth, or inequality in social conditions, are considered in order to explain these rates, mainly in the largest cities (Mayhew and Levinger, 1976, Blau and Blau, 1982). Considering disordered population growth as a feasible explanation for criminal raising, researchers from the Chicago sociological school, since the beginning of the 20th century, have been observing how the urban environment produces criminogenic conditions as a result of ecological factors (Wirth, 1938, Burguess, 1925, Harris, 1976). Urbanization and social differentiation produces an impersonal environment causing weakened social controls and permitting deviant behavior to occur more frequently. In this context, it is important the *Social Disorganization Theory* (Shaw and McKay, 1942) in the political perspective. While the fast urban growth is linked to social disorganization, Gaviria and Pages (2002) shows the negative relation between urban growth and trust level of police

---

[2]www.datasus.gov.br/.
[3]External causes.

and judiciary power. Their results suggest that rapid urbanization can, under some circumstances, set in motion a dynamic of mutual reinforcement between increasing crime rates and crumbling law enforcement institutions. In addition, alteration in the criminal historic behavior is associated to *law-enforcement* elements, such as increase of the number of police officers, expenses with safety policies and increase of imprisonment rates. Analyzing data from New York City, Zimring (2007) concluded that "there is a strong evidence that changing the number of cops, as well policing tactics, has a important impact in crime" (pg. 151).

Specifically in Brazil, Goertzel and Kahn (2009), while studying the behavior of the strong decline of homicides rates in São Paulo state since 2000s, have concluded that more repressive police models and disarmament policies reduced substantially homicides and other violent crimes in the state. Statistically, it was expected that, in regions where some policies of criminal prevention and control were applied, the rates could suffer from sudden structural changes. This fact requires a robust approach to model them, such as the assumption of heavy-tailed distribution for the latent system noise to handle possible innovative outliers as discussed in Section 3.

The model adapted was similar to equations (2.6) and (2.7). However, to model each time series we considered the state capital cities grouped according to a specific criterion. Specifically, we have used the following capital division:

- Group 1 (G1)—São Paulo and Rio de Janeiro.
- Group 2 (G2)—Belo Horizonte, Recife, Vitória and Porto Alegre.
- Group 3 (G3)—All the 21 remaining capitals.

The main characteristic that motivated this division consists on the different urbanization processes occurred in Brazil. As demonstrated in Santos (2005), the complex urban and territorial Brazilian organization bears deep differences between regions in terms of urbanization. In that sense, the cities division groups were adapted from the study regarding the urbanization process of the Brazilian society between 1940 and 2010, in particular, concerning the evolution of urban population in the metropolitan clusters in this period (de Brito and de Pinho, 2012). For each group created, we have one tendency estimated for the capitals and one tendency estimated for all first order spatial neighbours (those which share border with the capital). Thus, for each group, there is a model for the capitals and another one for the capitals neighbours, totaling six models.

We can see in Figure 3 the homicide rate analysis for G1, G2 and G3. According to this graph, some features can be pointed such as the presence of unusual rates. Two cities in G2 between 1995 and 2000 present high level in homicide rates compared to their usual levels. Moreover, we also highlight unusual values for some series in G3, first observed in two series close to 1990 and second a sudden raise after 2005 in one time series. Therefore, since we have these features a robust approach to estimate such data is justified.
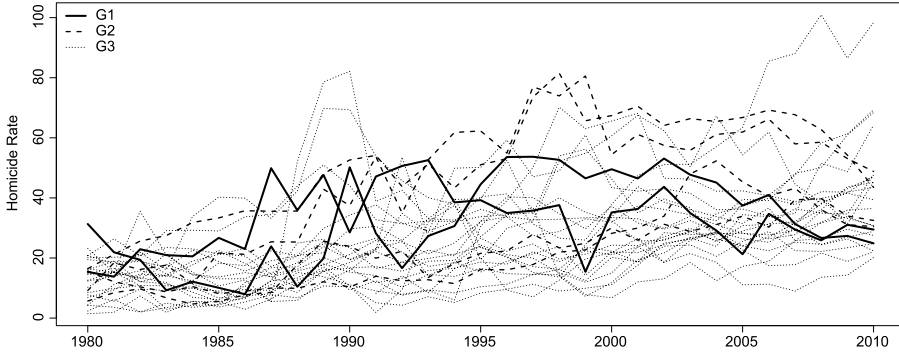
Figure 3 *Capital Homicide Rate data for each group.* G1: *São Paulo and Rio de Janeiro*; G2: *Belo Horizonte, Recife, Vitória and Porto Alegre*; G3: *All the* 21 *remaining capitals.*

Let Y_{til} be the Homicide Rate of city $i = 1, \ldots, n_l$ in group $l = 1, \ldots, 6$ at time t, then we have the following model:

$$Y_{til} = a_{tl} + v_{til}, \qquad v_{til} \sim N(0, \theta_{1il}), t = 1, \ldots, 31,$$

$$a_{tl} = a_{(t-1)l} + w_{tl}, \qquad w_{tl} \sim (0, \theta_{2l}), t = 2, \ldots, 31,$$

where w_{tl} is either Gaussian or student-t and n_l is the number of cities in group l.

To complete the model specification, it is necessary to specify the priors for the hyperparameters. In our case, we have to specify priors for the precision of each group in the observational equation and a single prior for the system equation precision. The prior of the precision of the observational equation is created to cover with high probability the variances of all Brazilian cities, setting a prior $\theta_{1il}^{-1} \sim \text{Gamma}(5, 500), i = 1, \ldots, n_l$ and $l = 1, \ldots, 6$ we cover with 90% of probability the values between the 25th and 75th quantiles of the cities sample variance. For the precision prior for the latent equation, a prior $\theta_{2l}^{-1} \sim \text{Gamma}(1, 0.1)$ is used for all cases. Finally, the same prior set in the simulation study for the degrees of freedom v assuming 30% of probability of prior mass for v values between 2 and 10 is used.

It is important to emphasize that the latent state represented by the vector **a**, can be interpreted as the non-observed mean tendency of the cities of each modeling.

To understand the temporal trends, sociological literature analyzes the context of how interpersonal violence spreads. Analyzing a historical time series of more than 30 years of delinquency and crime, Shaw and McKay (1942) verified that not only crime, but several social problems were related to a disorganized social environment. The theoretical approach developed by these authors helps to understand the effects of a unplanned urbanization process in criminal behavior. In largest context like Latin America de urban crime is, to an important extent, a conse-
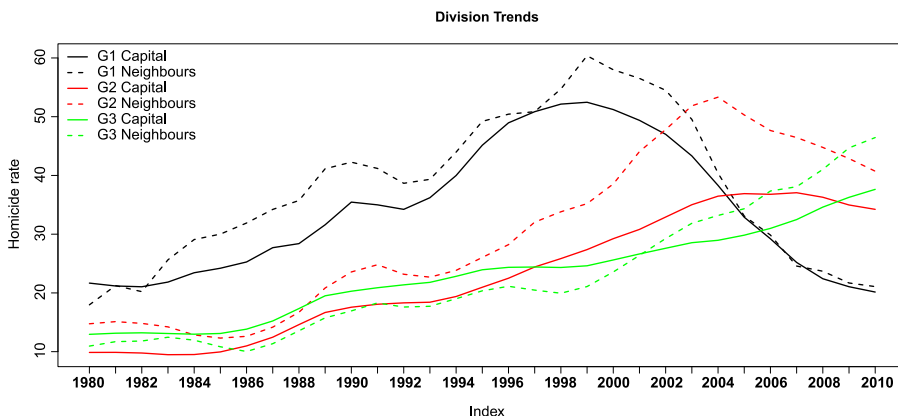
Figure 4 *Division posterior tendencies*, **a**, *for each group of Student-t approach*. G1: *São Paulo and Rio de Janeiro*; G2: *Belo Horizonte*, *Recife*, *Vitória and Porto Alegre*; G3: *All the* 21 *remaining capitals*.

quence of the inability of many cities in the region to keep up with the increasing demands for public safety brought about by a hasty and disorderly urbanization process (Gaviria and Pages, 2002). So, in Brazilian big cities, the economic development was followed by the appearance of urban enclaves (such as slums) where the impairment of the traditional mechanisms of social control promotes an environment of differentiated criminal opportunities (Sutherland, Cressey and Luckenbill, 1992). For more information about *Social Disorganization Theory*, we refer to Shaw and McKay (1942) and Kubrin and Weitzer (2003).

Figure 4 presents the posterior tendency, **a**, of each group capital and neighbours for the adjusted student-*t* model. The tendency of the first group tells us that its urbanization process started earlier when compared to the other groups because the homicide rate started to get higher first. Another aspect was that until 2000, all the tendencies were nearly linear for all groups. However, a reverse tendency was observed as result of investments and criminal control policies established in G1 (Goertzel and Kahn, 2009). The other two groups are heading towards the same behavior, but they still didn't show it in such an evident way, as seen in G1, the effect of safety policies. In G2 some states already adopted some safety measures, for example, Minas Gerais, which capital is Belo Horizonte, with the creation of *Integração da Gestão em Segurança Pública* (IGESP) in May 2005, and Pernambuco, which capital is Recife, with the creation of the *Pacto Pela Vida* in May 2007.

In order to assess the goodness of fit of our robust approach, we computed the -LPMLs and the DICs for every model, as can be seen in Table 1. From Table 1, all criteria, -LPMLs and DICs, pointed to the robust approach assuming the Student-*t* distribution for the system noise. To verify the evidence that the robust approach

Table 1 *Quality measures of the division*. G1: *São Paulo and Rio de Janeiro*; G2: *Belo Horizonte*, *Recife*, *Vitória and Porto Alegre*; G3: *All the* 21 *remaining capitals*

		G1		G2		G3	
		Gaussian	Student-*t*	Gaussian	Student-*t*	Gaussian	Student-*t*
-LPML	Capitals	229.09	227.81	496.72	496.45	2481.70	2473.55
	Neighbours	3536.78	3517.38	2379.98	2375.89	3026.54	3002.48
DIC	Capitals	460.22	459.64	997.44	996.60	4960.98	4946.32
	Neighbours	7063.15	7030.68	4753.67	4746.94	6040.70	6004.27

Table 2 *Pseudo Bayes Factor criteria*

2PsBF	Evidence against Gaussian
$(-1, 1]$	Worth mention
$(1, 5]$	Positive
$(5, 9]$	Strong
$(9, \infty)$	Very strong

outperforms significantly the traditional one, in the real data set, we chose to analyze Table 2 which was proposed and used in Prates et al. (2010). From Table 2, we can see that the PsBF, which is the -LPML difference as presented in Section 3, have a positive evidence against the Gaussian model for the capital modeling in G1 (2PsBF = 2.56) and a strong evidence against the Gaussian model for the neighboring modeling in G2 (2PsBF = 8.18). However, the bigger gain in favor of the Student-*t* model is verified for the neighboring modeling in G1 and G3 as well as the capitals in this last group, with all having strong evidence in terms of the -LPML differences. Since for each group we have many first order neighbours time series, for G1, G2 and G3 there are 30, 24 and 152 neighbours respectively, analysis suggests that as the number of cities increases there is a demand for a more robust approach.

In our real data application, the gain in terms of predictive power was very clear. We also should point out that there is evidence of deviation from Gaussianity when we look to the posterior distribution of the ν in the student-*t* model. From Table 3, we can see the posterior median and 95% credible intervals (CI). The median measure indicates that the student-*t* distribution is concentrated in medium values of the degrees of freedom but the 95% CI are highly asymmetric reaching very high values.

5 Conclusions

This paper describes how to perform Bayesian inference using `R-INLA` to estimate non-Gaussian Dynamic Models when the evolution noise has a non-Gaussian

Table 3 *Posterior measures of the ν of Student-t approach*. G1: *São Paulo and Rio de Janeiro*; G2: *Belo Horizonte*, *Recife*, *Vitória and Porto Alegre*; G3: *All the* 21 *remaining capitals*

Group	Type	Median	95% Credible interval
G1	Capitals	35.53	(5.84; 441.30)
	Neighbours	28.46	(6.04; 339.38)
G2	Capitals	39.93	(6.73; 533.77)
	Neighbours	31.69	(5.56; 445.12)
G3	Capitals	32.23	(5.14; 434.54)
	Neighbours	55.91	(21.69; 556.67)

distribution. Such models can be viewed as part of latent hierarchical models where a non-Gaussian Random Field is assumed for the latent field and, therefore, invalidates the direct use of the INLA methodology that requires that the latent field must be Gaussian.

Using a random walk example, we presented how to use an augmented structure to overcome the Gaussian limitation of INLA for the latent field. The key to understand why our approach works relies on the fact that we approximate the non-Gaussian latent field through a Gaussian distribution and corrects this approximation in the likelihood function trying to minimize the loss of this approximation for dependent models. We discussed and explained the reasons to make this approximative approach and, specially, where in the R-INLA calculations it will impact.

Through simulations, we showed the necessity of more robust models when the time series suffer sudden structural changes. From our results, we observe that Gaussian models are sensitive to structural changes while our approach assuming a student-*t* field is robust. Specifically, our simulation study presented an incisive demand to avoid the usual Gaussian assumption in most contaminated scenarios. There are indication that some public policies for crime control can generate a positive effect in crime's temporal tendencies allowing the presence of structural changes identified with our proposed approach. It is evident that other control factors might help to confirm this hypothesis, however it is very likely that investments in security policies, such as those implemented in G1 and G2, have contribution in the dynamic observed. Our homicide rate application pointed-out, as expected, that public policies could play an important role to explain homicides dynamics through a robust approach due to the characteristic of these kind of data. Although we analyzed homicide rate because of their sociological impacts, we are aware that this extension would also be well justified in other fields such as stochastic volatility models (see, for example, Jacquier, Polson and Rossi, 2003).

As mentioned in Section 2, a natural extension of the model class presented is the DGLM, where one could assume a non-Gaussian distribution for the observed data and, consequently, impacting equation (2.5) which both $g_t(x_t)$ and

$h_t(x_t)$ could have a non-quadratic form. This extension is investigated in a different manuscript. The main advantage of the model structure presented here is that it allows users to fit basically any complex structured non-Gaussian dynamic model with fast and good accuracy using a friendly tool already available.

We believe that the applied community can make good use of this methodology when necessary. For real time series data is not rare to observe structural breaks and a robust approach, as the one presented, may be more adequate to adjust this type of data. Furthermore, we have formalized how to use the `R-INLA` software for non-Gaussian dynamic models in a simple way.

References

Blau, J. R. and Blau, P. M. (1982). The cost of inequality: Metropolitan structure and violent crime. *American Sociological Review* **47**, 114–129.

Brito, F. A. and Pinho, B. A. T. D. (2012). A dinâmica do processo de urbanizacão no Brasil. Technical Report #464, Belo Horizonte UFMG-CEDEPLAR.

Burguess, E. (1925). *The Growth of the City. Inside the City Book*. Chicago: Univ. Chicago Press.

Bustos, O. H. and Yohai, V. J. (1979). Robust estimates for arma models. *Journal of the American Satistical Association* **81**, 155–168. MR0830576

Denby, L. and Martin, R. D. (1979). Robust estimation of the 1st order autoregressive parameter. *Journal of the American Satistical Association* **74**, 140–146.

Dey, D. K., Chen, M. H. and Chang, H. (1997). Bayesian approach for nonlinear random effects models. *Biometrics* **53**, 1239–1252.

Durbin, J. and Koopman, S. (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. *Journal of the Royal Statistical Society, Series B* **62**, 3–56. MR1745604

Flood-Page, C., Campbell, S., Harrington, V. and Miller, J. (2000). Youth Crime: Findings from the 1998/99 Youth Lifestyles Survey. London: HMSO. Report from Home Office Research Study 209.

Fox, A. J. (1972). Outliers in time series. *Journal of the Royal Statistical Society, Series B* **3**, 350–363. MR0331681

Gaviria, A. and Pages, C. (2002). Patterns of crime victimization in latin American cities. *Journal of Development Economics* **67**, 181–203.

Geisser, S. and Eddy, W. F. (1979). A predictive approach to model selection (Corr: V75 p765). *Journal of the American Statistical Association* **74**, 153–160. MR0529531

Gelfand, A. E., Dey, D. K. and Chang, H. (1992). Model determination using predictive distributions, with implementation via sampling-based methods (Disc: P160-167). In *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.). *Proceedings of the Fourth Valencia International Meeting*, 147–159. Oxford: Clarendon Press. MR1380275

Goertzel, T. and Kahn, T. (2009). The great São Paulo homicide drop. *Homicide Studies* **13**, 398–410.

Graham, J. and Bowling, B. (1995). *Young People and Crime*. London: HMSO. Report from Home Office Research Study 145.

Harris, K. D. (1976). Cities and crime: A geographic model. *Criminology* **14**, 369–386.

Jacobs, D. and Richardson, A. M. (2008). Economic inequality and homicide in the developed nations from 1975 to 1995. *Homicide Studies* **12**, 28–45.

Jacquier, E., Polson, N. G. and Rossi, P. E. (2003). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics* **122**, 185–212. MR2083256

Kalman, R. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering* **82**, 35–45.

Kitagawa, G. (1987). Non-Gaussian state-space modeling of nonstationary time series. *Journal of the American Statistical Association* **82**, 1032–1041. MR0922169

Kubrin, C. E. and Weitzer, R. (2003). New directions in social disorganization theory. *Journal of Research in Crime and Delinquency* **40**, 374–402.

Martins, T. G., Simpson, D., Lindgren, F. and Rue, H. (2012). *Bayesian Computing with INLA: New Features*. NTNU, Norway: Department of Mathematical Sciences.

Martins, T. G. and Rue, H. (2014). Extending integrated nested laplace approximation to a class of near-gaussian latent models. *Scandinavian Journal of Statistics* **41**, 893–912. MR3277029

Martins, T. G., Simpson, D. P., Riebler, A., Rue, H. and Søorbye, S. H. (2014). Penalising model component complexity: A principled. *practical approach to constructing priors*, 1–34. Available at arXiv:1403.4630. MR3277029

Masreliez, C. J. and Martin, R. D. (1977). Robust Bayesian estimation for the linear model and robustyfing the Kalman filter. *IEEE Transactions on Automatic Control* **22**, 361–371. MR0453124

Mayhew, B. H. and Levinger, R. L. (1976). Size and the density of interaction in human aggregates. *American Journal of Sociology* **82**, 86–110.

Mc Quarrie, A. D. and Tsai, C. (2003). Outlier detections in autoregressive models. *Journal of Computational and Graphical Statistics* **12**, 450–471. MR1983164

Migon, H., Gamerman, D., Lopes, H. and Ferreira, M. (2005). Dynamic models. *Handbook of Statistics* **25**, 553–588. MR2490539

Mir, L. (2004). A guerra civil: estado e trauma. São Paulo: Geração Editorial.

Pinheiro, J., Liu, C. and Wu, Y. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate *t* distribution. *Journal of Computational and Graphical Statistics* **10**, 249–276. MR1939700

Prates, M., Dey, D., Willig, M. and Yan, J. (2010). Intervention analysis of hurricane effects on snail abundance in a tropical forest using long-term spatiotemporal data. *Journal of Agricultural, Biological, and Environmental Statistics* **16**, 142–156. MR2767764

Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71**, 319–392. MR2649602

Ruiz-Cárdenas, R., Krainski, E. and Rue, H. (2012). Direct fitting of dynamic models using integrated nested laplace approximations–INLA. *Computational Statistics & Data Analysis* **56**, 1808–1828. MR2892379

Santos, M. (2005). *A Urbanizacão Brasileira*. São Paulo: USP Publisher.

Shaw, C. R. and McKay, H. (1942). *Juvenile Delinquency in Urban Areas*. Chicago: Univ. Chicago Press.

Sherman, L. W., Gottfredson, D. C., MacKenzie, D. L., Eck, J., Reuter, P. and Bushway, S. D. (1997). Preventing crime: What works, what doesn't, what's promising. Washington, DC: National Institute of Justice, US Department of Justice.

Sherman, L. and Weisburd, D. (1995). General deterrent effects of police patrol in crime "hot spots": A randomized, controlled trial. *Justice Quarterly* **12**, 625–648.

Sherman, L. W. (1998). Evidence-based policing. Ideas in American Policing, July. Washington, DC: Police Foundation.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* **64**, 583–639. MR1979380

Sutherland, E., Cressey, D. and Luckenbill, D. (1992). *Principles of Criminology*. General Hall, A Division of Rowman and Littlefield Publishers.

Tierney, L. and Kadane, J. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association* **81**, 82–86. MR0830567

West, M. (1981). Robust sequential approximate Bayesian estimation. *Journal of the Royal Statistical Society. Series B* **43**, 157–166. MR0626760

Wirth, L. (1938). Urbanism as a way of life. *American Journal of Sociology* **44**, 1–24.
Zimring, F. E. (2007). *The Great American Crime Decline*. London: Oxford Univ. Press.

R. X. Cortes
M. O. Prates
Department of Statistics
Universidade Federal de Minas Gerais
Presidente Antônio Carlos 6627
Belo Horizonte
Brazil
E-mail: renanxcortes@gmail.com
        marcosop@gmail.com

T. G. Martins
Department of Mathematical Sciences
Norwegian University of Science
    and Technology
Høgskoleringen 1
7491 Trondheim
Norway
E-mail: thigm85@gmail.com

B. A. Silva
Department of Sociology
Universidade Federal de Minas Gerais
Presidente Antônio Carlos 6627
Belo Horizonte
Brazil
E-mail: braulio.fas@gmail.com