KDGLM: COMPUTATIONAL EFFICIENCY AND SCALABILITY IN GENERALIZED DYNAMIC MODELS

Mariane Branco Alves - DME - IM - UFRJ

mariane@im.ufrj.br

Joint work with Silvaneo Vieira dos Santos Junior e Helio dos Santos Migon

1 INTRODUCTION

2 DYNAMIC GENERALIZED LINEAR MODEL

3 OUR PROPOSAL: SEQUENTIAL LEARNING ON k-DGLMS

4 APPLICATIONS

5 REFERENCES

INTRODUCTION

- Our focus is time series modeling.
- Our aim is to operate with models that are not limited by restrictive hypotheses such as stationary.
- Another relevant point: modeling the data in their natural scale, without the need for transformations, for example, for induction of normality.
- We adopt the class of Dynamic Bayesian Models

INTRODUCTION

- By relaxing certain restrictive assumptions, there is no analytical solution for the cycle of updates required for Bayesian Inference in this class.
- We will work with our own proposal for updating information in this context.Essential aspects in our proposal:
  - Preservation of the sequential aspect of the analysis;
  - Computational efficiency.

$\Rightarrow$ An R package was developed based on a novel methodology for Bayesian updating in dynamic generalyzed models: kDGLM (https://github.com/silvaneojunior/kDGLM).

A TYPICAL NON-STATIONARY TIME SERIES

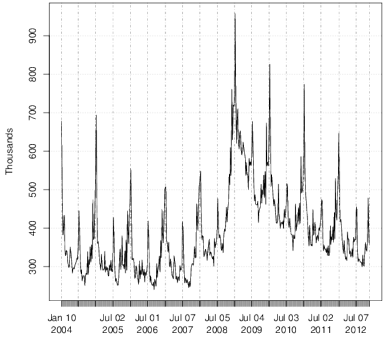


FIGURE: Weekly initial claims for unemployment in the US.

BAYESIAN DYNAMIC MODELS

- Dynamic models are built based on structural components with clear interpretation, in order to describe:
  - temporal trends;
  - seasonal patterns;
  - regressor effects.
- Each structural component can be assigned temporal dynamics
- The Bayesian inferential approach allows for naturally accommodating external information to the observed data.

DYNAMIC LINEAR MODELS

Denote all the information available to the analyst at time t by D_t. The general form of a dynamic linear model is:

$$\begin{aligned} \text{Obs. eq. : } \mathbf{y}_t &= \mathbf{F}_t'\theta_t + \nu_t \qquad \nu_t \sim N[0, \mathbf{V}_t] \\ \text{Evol. eq: } \theta_t &= \mathbf{G}_t\theta_{t-1} + \omega_t \qquad \omega_t \sim N[0, \mathbf{W}_t] \\ \text{Initial information : } (\theta_0 \mid D_0) &\sim N[\mathbf{m}_0, \mathbf{C}_0] \end{aligned} \tag{1}$$

- $\mathbf{y}_t$ response vector ($r \times 1$);
- θ_t latent states vector ($n \times 1$);
- $\mathbf{F}_t$ known design matrix ($n \times r$);
- $\mathbf{G}_t$ known evolution matrix ($n \times n$);
- $\mathbf{V}_t$ observational covafriance matrix ($r \times r$);
- $\mathbf{W}_t$ matriz de covariâncias conhecida ($n \times n$);

The model is defined by $\{\mathbf{F}, \mathbf{G}, \mathbf{V}, \mathbf{W}\}_t$.

EX: LINEAR TREND + REGRESSION DYNAMIC MODEL

$$\text{EQ. OBS.}: y_t = \begin{pmatrix} 1 & 0 & X_t \end{pmatrix} \begin{pmatrix} \mu_t \\ \beta_t \\ \gamma_t \end{pmatrix} + v_t, \quad v_t \sim N(0, V_t)$$

$$\text{EQ. SIST.}: \begin{pmatrix} \mu_t \\ \beta_t \\ \gamma_t \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ \beta_{t-1} \\ \gamma_{t-1} \end{pmatrix} + \omega_t$$

$$\text{INFO. INICIAL}: \begin{pmatrix} \mu_0 \\ \beta_0 \\ \gamma_0 \end{pmatrix} \sim N\left[\begin{pmatrix} m_0 \\ b_0 \\ g_0 \end{pmatrix}, \mathbf{C}_0 \right]$$

ONDE

$$\omega_{\mathbf{t}} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} W_{1,t} + W_{2,t} & W_{2,t} & 0 \\ W_{2,t} & W_{2,t} & 0 \\ 0 & 0 & W_{3,t} \end{pmatrix} \right]$$

Bayesian updating

$$
\begin{array}{ccccc}
\theta_{t-1} \mid D_{t-1} & \overset{evolution}{\rightarrow} & \theta_t \mid D_{t-1} & \overset{updating}{\rightarrow} & \theta_t \mid D_t \\
\text{posterior} & & \text{prior} & & \text{posterior} \\
 & & \downarrow & & \downarrow \\
 & & Y_t \mid D_{t-1} & & Y_{t+h} \mid D_t \\
 & & \text{prediction} & & \text{prediction}
\end{array}
$$

For uni-variate and multivariate DLMs, the updating cycle has analytical solution (see West & Harrison, chaps 4, 16).

Unknown components in $\mathbf{F}_t$, $\mathbf{G}_t$ or non-Gaussian responses make it unfeasible to obtain an analytical solution for the update process.

Bayesian Updating

(A) Prior distribution at time t:

$$
\begin{aligned}
p(\theta_t|D_{t-1}) &= \int p(\theta_t, \theta_{t-1}|D_{t-1})d\theta_{t-1} \\
&= \int p(\theta_t|\theta_{t-1}, D_{t-1})p(\theta_{t-1}|D_{t-1})d\theta_{t-1}
\end{aligned}
$$

(B) One-step-ahead prediction for Y_t:

$$
\begin{aligned}
p(Y_t|D_{t-1}) &= \int p(Y_t, \theta_t|D_{t-1})d\theta_t \\
&= \int p(Y_t|\theta_t, D_{t-1})p(\theta_t|D_{t-1})d\theta_t
\end{aligned}
$$

BAYESIAN UPDATING

(C) Posterior distribution at time t:

$$p(\theta_t|D_t) \propto p(Y_t|\theta_t, D_{t-1})p(\theta_t|D_{t-1})$$

Once the posterior distribution for the states is available at time t, there may be interest in the predictive distribution h steps ahead:

$$p(Y_{t+h}|D_t) \propto p(Y_{t+h}|\theta_{t+h}, D_t)p(\theta_{t+h}|D_t), \quad h = 1, 2, \ldots.$$

The estimates of latent states can be revised in light of all available information through smoothed distributions:

$$p(\theta_{T-h}|D_T) \quad h = 1, 2, \ldots.$$

beginframeDynamic Generalized Models

- In practical applications, violation of the Gaussian assumption for the response are frequent.

- We focus on the class of Dynamic Generalized Linear Models (DGLM), defined by West et al(1985), assuming responses distributed in the **uni-parametric** exponential family

- Bayesian inference must be approximated.

Dynamic Generalized Models

- In this work, we consider DGLMs for k-parametric exponential families and assume that k dynamic predictors may be specified; for instance:
  1. different predictor structures for the mean and precision on a normal dynamic model;
  2. k predictors for $d = k + 1$ categorical counts of a multinomial response.
- We propose an inferential approach for k-parametric uni or multivariate DGLMs based on information geometry, focusing on sequential Bayesian learning and real-time inference (naturally dealing with monitoring, intervention and discounting strategies)

SEQUENTIAL ANALYSIS: INTERVENTION

- D_0: Prior information; $D_t = \{D_0, y_1, \ldots, y_t\}$ (if the learning system is closed to external information)
- Intervention effects: $I_t = \{h_t, H_t\}$ or $\xi_t \sim N[h_t, H_t]$, where h_t and H_t are subjectively evaluated
  $\Rightarrow$ Available informationl:$\{D_t, I_t\}$.
- Instantaneous effects: $I_t = \{y_t \text{ is missing}\} \simeq \{V_t^{-1} \rightarrow 0\}$

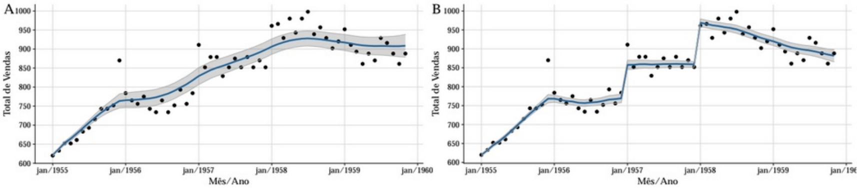


FIGURE: Smoothed mean and 95% credibility interval **A**: Without **B**: With intervention.

1 Introduction

2 Dynamic Generalized Linear Model

3 Our Proposal: sequential learning on k-DGLMs

4 Applications

5 References

DGLM - DEFINITION

A Dynamic Generalized Linear Model (DGLM) is defined, for each time t, by three equations:

$$
\begin{aligned}
p(y_t|\psi_t) &= c(y_t)exp\{\mathbf{H}'\psi_t - b(\psi_t)\} && \text{Obs. Eq.}\\
\boldsymbol{g}(\boldsymbol{\eta}_t) &= \boldsymbol{F}_t'\boldsymbol{\theta}_t = \boldsymbol{\lambda}_t && \text{Predictor}\\
\boldsymbol{\theta}_t &= \mathbf{G}_t'\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(\mathbf{0}, \mathbf{W_t}). && \text{Evol. Eq.}
\end{aligned}
$$

- $\mathbf{H}'(y_t) = (h_1(y_t), \ldots, h_k(y_t))$ is a vector of sufficient statistics for $\boldsymbol{\eta}_t$;
- $\boldsymbol{\psi}_t' = (\psi_{1t}, \ldots, \psi_{kt})$, $\psi_{it} = c_i\phi_i(\eta_t)$, $i = 1, \ldots, k$. $\boldsymbol{\psi}_t$ is the natural or canonical parameter .
- $\mathbf{F}_t$ is a known $(p \times k)$ dynamic regression matrix;
- $\mathbf{G}_t$ is a $(p \times p)$ state evolution matrix;
- $\mathbf{W}_t$ is a $(p \times p)$ evolution covariance matrix.
- $\boldsymbol{g}$ is an invertible link function. If $g_i(\boldsymbol{\eta}_t) = c_i\phi_i(\boldsymbol{\eta}_t) = \psi_{it}$, $i = 1, \ldots, k$, $\boldsymbol{g}$ is the canonical link.

The conjugate prior for ψ_t is given by:

$$p(\psi_t|\boldsymbol{\tau}) = [K(\tau_0, \boldsymbol{\tau})]^{-1} \exp\left\{\tau'\psi_t - \tau_0 b(\psi_t)\right\},$$

$\boldsymbol{\tau}' = (\tau_1, \ldots, \tau_k)$.

- Once the conjugate prior for ψ_t is adopted, posterior and predictive distributions are analytically available.
- In a DGLM, interest lies not only on ψ_t, but also on the states $\boldsymbol{\theta}_t$, for which no conjugate form is available.

DGLM - APPROACHES FOR APPROXIMATING THE POSTERIOR

- MCMC: high computational cost, sequential aspect of the analysis is lost.
  (*e.g. Frühwirth-Schnatter, 1994; Carter and Kohn, 1994; Gamerman, 1998, 1997; Shephard and Pitt, 1997; Durbin and Koopman, 2002*);
- Conjugate Updating : computational efficiency, but restricted to one-parametric exponential families.
  (*West et al, 1985*)
- Bi-parametric exponential family, GMM
  (*Souza et al, 2016*)
- Formulations based on local level models: lack flexibility for the predictor structure.
  (*e.g. Gamerman et al, 2013*)

1 Introduction

2 Dynamic Generalized Linear Model

3 Our Proposal: sequential learning on k-DGLMs

4 Applications

5 References

OUR PROPOSAL: SEQUENTIAL LEARNING ON k-DGLMS

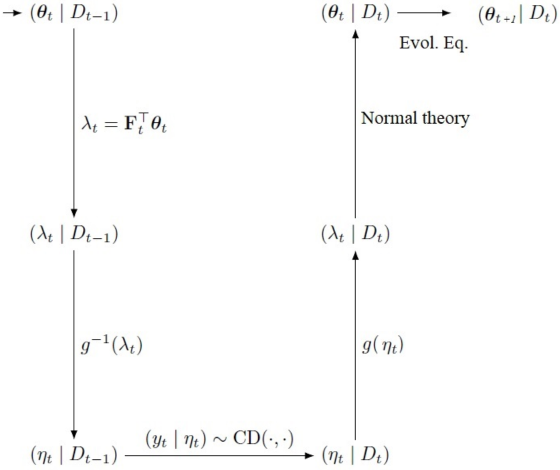
Remember that

$$\boldsymbol{g}(\boldsymbol{\eta}_t) = \boldsymbol{F}_t'\boldsymbol{\theta}_t = \boldsymbol{\lambda}_t.$$

1. At time t=1, assume a p-dimensional Normal prior for the states $\boldsymbol{\theta}_t$, inducing a k-variate normal prior for the vector of linear predictors $\boldsymbol{\lambda}_t$;
2. obtain the conjugate prior for $\boldsymbol{g}(\boldsymbol{\eta}_t)$ that is the best approximation for the normal prior of $\boldsymbol{\lambda}_t$;
3. Update $\psi_t = \boldsymbol{g}(\boldsymbol{\eta}_t)$ through conjugacy properties, obtaining its posterior distribution;
4. obtain the Normal posterior for $\boldsymbol{\lambda}_t$ that is the best approximation for the posterior of $\boldsymbol{g}(\boldsymbol{\eta}_t)$;
5. apply normal theory properties to obtain the updated distribution of $\boldsymbol{\theta}_t$, given $\boldsymbol{\lambda}_t$;
6. repeat steps 1 to 5 for $t = 2, \ldots, T$.

OUR PROPOSAL: SEQUENTIAL LEARNING ON k-DGLMS

Let D_t denote the set of available information at time t.



RECONCILING PRIOR SPECIFICATIONS: PROJECTION THEOREM

Kullback-Leibler (KL) divergence between distributions p and q:

$$D_{KL}[p:q] = KL[p(\eta|\boldsymbol{\tau}'); q(\eta|\boldsymbol{\tau})] = \int p(\eta|\boldsymbol{\tau}')\, log(\frac{p(\eta|\boldsymbol{\tau}')}{q(\eta|\boldsymbol{\tau})})\, d\eta.$$

THEOREM (PROJECTION THEOREM - V. AMARI, 2016))

Let $p(\eta)$ be a probability distribution on a set $\mathcal{F}$. Let S be an exponential family over $\mathcal{F}$. The distribution $q(\eta)$ that minimizes the divergence $D_{KL}[p:q]$, $q(\eta) \in S$ is such that

$$E_q(H_q) = E_p(H_q),$$

where H_q is the vector of sufficient statistics under distribution q.

RECONCILING PRIOR SPECIFICATIONS: PROJECTION THEOREM

- $D_{KL}[p : q] \neq D_{KL}[q : p]$, so the minimization of the KL divergence must be reevaluated for posterior distributions.
- Although we operate on the minimization of the divergence between two prior/posterior densities, the method resumes to an optimization problem on the parametric space.
- Specifically, we search for the parameters $\boldsymbol{\tau}_t$ of the conjugate prior/posterior that minimize the KL divergence between the conjugate specification and the one induced by the normal assumption for the states and perform the reciprocal operation for the posterior updating of the linear predictors.

FILTERING ALGORITHM

arXiv: Alves, Migon, Marotta, Santos Jr (2023). k-parametric Dynamic Generalized Linear Models: a sequential approach via Information Geometry

Algorithm 1: Filtering and one-step ahead prediction in DGLMs via Information Geometry

Step 1: Evolution. Given m_{t-1}, C_{t-1} *and* $F_t'\theta_t = \boldsymbol{\lambda}_t$.

$$\mathrm{a}_t = \mathrm{G}_t\mathrm{m}_{t-1} \quad \text{and} \quad R_t = G_tC_{t-1}G_t' + W_t$$
$$f_t = F_t'a_t \quad \text{and} \quad Q_t = F_t'R_tF_t$$

Step 2: Given f_t *and* Q_t*, solve for* τ_t*:*

Step 2.1: Obtain the vector of sufficient statistics $\mathbf{H}_q(\boldsymbol{\eta}_t)$ of the conjugate distribution $q(\boldsymbol{\eta}_t|\boldsymbol{\tau}_t)$.

Step 2.2: Obtain $\boldsymbol{\tau}_t$ such that $E_p(\mathbf{H}_q(\boldsymbol{\eta}_t)) = E_q(\mathbf{H}_q(\boldsymbol{\eta}_t))$ where the induced distribution is $p(\boldsymbol{\eta}_t|\boldsymbol{f}_t, \boldsymbol{Q}_t)$.

Step 3: Update posterior moments using Bayes' Theorem

$$\boldsymbol{\tau}_t^* = \boldsymbol{\tau}_t + \boldsymbol{h}(\boldsymbol{y}_t)$$

Step 4: Given $\boldsymbol{\tau}_t^*$*, solve for* $\boldsymbol{f}_t^*$ *and* $\boldsymbol{Q}_t^*$*:*

Identify the resulting posterior distributions, then equate its moments considering the sufficient statistics of the Normal$(\boldsymbol{f}_t^*, \boldsymbol{Q}_t^*)$, the updated posterior for the linear predictor.

$$E_p(\mathbf{H}_q(\boldsymbol{\lambda}_t)) = E_q(\mathbf{H}_q(\boldsymbol{\lambda}_t))$$

where $p(\boldsymbol{\lambda}_t|D_t) \sim p_{\boldsymbol{\eta}_t|D_t}(\boldsymbol{g}(\boldsymbol{\eta}_t))\left|\nabla_{\boldsymbol{\lambda}_t}\boldsymbol{g}(\boldsymbol{\eta}_t)\right|$ and $\mathbf{H}_q = [\boldsymbol{\lambda}_t, \boldsymbol{\lambda}_t^2]$.

Step 5: Obtain state posterior moments

$$m_t = a_t + R_tF_t\mathrm{Q}_t^{-1}(f_t^* - f_t) \quad \text{and}$$
$$C_t = R_t + R_tF_tQ_t^{-1}[Q_t^* - Q_t]\,Q_t^{-1}F_t'R_t$$

SMOOTHING AND PREDICTION ALGORITHM

Algorithm 2: Smoothing and J steps-ahead Forecast Distributions

Given $\boldsymbol{a}_t, \boldsymbol{R}_t, \boldsymbol{m}_t$ and $\boldsymbol{C}_t$, obtain the smooth the posterior moments for $t = 1, \ldots, T$:

$$\boldsymbol{m}_t^s = \boldsymbol{m}_t + \boldsymbol{C}_t \boldsymbol{G}'_{t+1} \boldsymbol{R}_{t+1}^{-1} \left(\boldsymbol{m}_{t+1}^s - \boldsymbol{a}_{t+1}\right) \quad \text{and}$$
$$\boldsymbol{C}_t^s = \boldsymbol{C}_t + \boldsymbol{C}_t \boldsymbol{G}'_{t+1} \boldsymbol{R}_{t+1}^{-1} \left(\boldsymbol{C}_{t+1}^s - \boldsymbol{R}_{t+1}\right) \boldsymbol{R}_{t+1}^{-1} \boldsymbol{G}_{t+1} \boldsymbol{C}_t$$

where $\boldsymbol{m}_T^s = \boldsymbol{m}_T, \quad \boldsymbol{C}_T^s = \boldsymbol{C}_T.$

Given $\boldsymbol{m}_T, \boldsymbol{a}_T, \boldsymbol{R}_T, \boldsymbol{C}_T, \boldsymbol{F}_{t+j}$ and $\mathbf{G}_{t+j}$, obtain the j-steps ahead prior for the linear predictor:

$$\boldsymbol{f}_t(j) = \boldsymbol{F}'_{t+j} \boldsymbol{a}_t(j), \quad j = 1, \ldots, J;$$
$$\mathbf{Q}_t(j) = \boldsymbol{F}'_{t+j} \boldsymbol{R}_t(j) \boldsymbol{F}_{t+j}, \quad j = 1, \ldots, J.$$

where

$\boldsymbol{a}_t(j) = \mathbf{G}_{t+j} \boldsymbol{a}_t(j-1); \boldsymbol{R}_t(j) = \boldsymbol{G}_{t+j} \boldsymbol{R}_t(j-1) \mathbf{G}'_{t+j} + \boldsymbol{W}_{t+j},$ with $\boldsymbol{a}_t(0) = \boldsymbol{m}_T; \boldsymbol{R}_t(0) = \boldsymbol{C}_T.$

Once the j-steps ahead prior distribution for the linear predictor is obtained $\rightarrow$ obtain the best conjugate approximation for the canonical parameters $\rightarrow$ $p(y_{t+j}|D_t)$ is obtained directly by conjugacy properties.

PARTICULAR CASES

We present the detailed computations involved in the particular cases:

- Bernoulli ($k = 1$);
- Poisson ($k = 1$);
- Normal with dynamic predictive structure both for the mean and precision ($k = 2$);
- Gamma ($k = 2$)
- Multinomial on $d = k + 1$ categories
- Bivariate normal ($k = 5$)

Work in progress:

- Generalization to m-variate normal, $m > 2$
- Dirichlet for $d = k + 1$ categories

R PACKAGE: KDGLM

The R package kDGLM, enables sequential analysis with: sequential

- filtering;
- smoothing;
- h-steps-ahead predictions;
- intervention;
- automated monitoring.

The package also enables to fit:

- autoregressive and transfer function models;
- multivariate models though latent common factors;

1 Introduction

2 Dynamic Generalized Linear Model

3 Our Proposal: sequential learning on k-DGLMs

4 Applications

5 References

Poisson DGLM for quarterly sales

We fit the following model through our proposal and the Conjugate Updating by West et al (1985):

$$\begin{aligned} y_t|\eta_t &\sim Poisson(\eta_t) \\ log(\eta_t) &= \boldsymbol{F}_t'\boldsymbol{\theta}_t \\ \boldsymbol{\theta}_t &= \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(\mathbf{0}, \boldsymbol{W}_t), \end{aligned}$$

with $\boldsymbol{F}_t' = [1, 0, 1, 0, 1, 0]$; $\boldsymbol{G}_t = diag\,[\boldsymbol{G_0}, \boldsymbol{G_1}, \boldsymbol{G_2}]$ and

$$\boldsymbol{G_0} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \boldsymbol{G_k} = \begin{bmatrix} \cos(kw) & sen(kw) \\ -sen(kw) & \cos(kw) \end{bmatrix}, w = 2\pi/4 \text{ and } k = 1, 2.$$

POISSON DGLM FOR QUARTERLY SALES

The same data were fit adopting the proposal by Gamerman et al(2013), which accommodates a single dynamic component (local level model):

$$
\begin{aligned}
y_t|\eta_t &\sim Poisson(\eta_t) \\
\eta_t &= \alpha_t \exp(\beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t}), \\
\omega \frac{\alpha_t}{\alpha_{t-1}} \mid \alpha_{t-1}, D_{t-1}, \varphi &\sim Beta[\omega a_{t-1}, (1-\omega) a_{t-1}] \\
\Rightarrow \alpha_t, \mid D_{t-1}, \varphi &\sim Gamma(a_{t|t-1}, b_{t|t-1})
\end{aligned}
$$

$x_{1t} = t, x_{2t} = cos(wt), x_{3t} = sin(wt)$ where t=$1, \ldots, 35$ and $w = 2\pi/4$
R Package: NGSSEML (Santos et al., 2021)

POISSON DGLM FOR QUARTERLY SALES

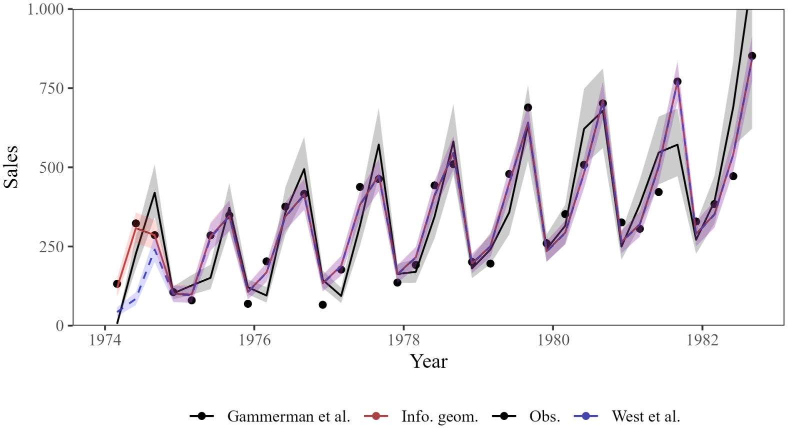


FIGURE: One step ahead prediction.

POISSON DGLM FOR QUARTERLY SALES

Metric	Local level	Conj. Upd.	Our prop.
MAE	82.808	**33.263**	33.496
RAE	0.267	0.122	**0.121**
MSE	10,569.117	**1,581.828**	1,594.727
LPL	-214.006	-	**-190.600**
Comp. time*	818.803s	**0.024s**	**0.024s**

TABLE: Model comparison based on the smoothed mean response function and several metrics. *Time measured using R 4.2.0 running under Windows 10 x64 (build 19044) in an Intel(R) Core(TM) i7-8700K CPU @ 3.70GHz with 16 GB RAM @ 2400 MHz.

Note that West et al (1985) are not able to deal with k-parametric exponential families, for $k > 1$.

STOCHASTIC VOLATILITY MODELS GAMMA FORMULATION

A traditional model for the returns is:

$$y_t = e^{h_t/2}\epsilon_t, \quad \epsilon_t \sim N(0,1)$$

$$\Rightarrow \ln(y_t^2) = h_t + a_t,$$

where a_t follows a log-χ^2 distribution.

$$\iff y_t^2|h_t = e^{h_t}\epsilon_t^2 \sim Gamma(1/2, 1/2e^{-h_t}).$$

GAMMA AND NORMAL FORMULATIONS

We propose two approaches:

- modelling the returns at their original scale, using a normal model with dynamic predictors for the mean and precision;
- a dynamic gamma model for the square of the returns, with shape parameter $\phi = 1/2$.

GAMMA DYNAMIC MODEL FOR THE SQUARED RETURNS

$$y_t^2|h_t = e^{h_t}\epsilon_t^2 \sim Gamma(1/2, 1/2e^{-h_t}).$$

$$\begin{aligned} \ln(\mu_t) &= \lambda_t = h_t \\ h_t &= h + s_t \\ s_t &= \gamma s_{t-1} + \omega_t \end{aligned}$$

γ an autoregressive parameter estimated following the ideas in West Harrison (1997, Chap.13,pp. 492-97).

NORMAL DYNAMIC MODEL FOR THE RETURNS AT THEIR ORIGINAL SCALE

$$\begin{aligned} y_t &\sim N(\mu_t, \phi_t^{-1}) \\ ln(\phi_t) &= -h_t \\ h_t &= h + s_t \end{aligned}$$

and predictive structure

$$\boldsymbol{\lambda}_t = \begin{bmatrix} \lambda_{1t} \\ \lambda_{2t} \end{bmatrix} = \begin{bmatrix} \mu_t \\ -h_t \end{bmatrix},$$

and evolution equations

$$\begin{bmatrix} \mu_t \\ s_t \end{bmatrix} = \begin{bmatrix} \mu_{t-1} \\ \gamma s_{t-1} \end{bmatrix} + \begin{bmatrix} \omega_{1t} \\ \omega_{2t} \end{bmatrix}$$

($\omega_{1t} = 0$ w.p. 1)

An univariate analysis, based on artificial data Gamma x Normal formulations

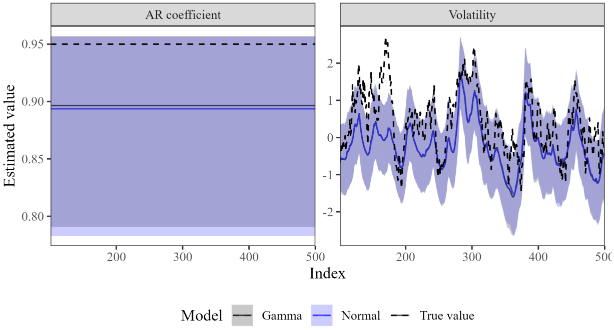


Figure: Gamma and Normal formulations: AR coefficient and Smoothed volatility

An univariate analysis, based on artificial data Gamma x Normal formulations

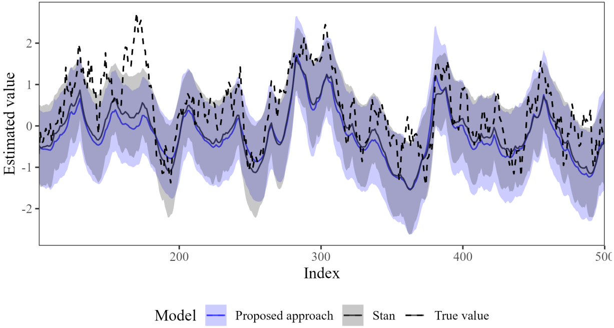
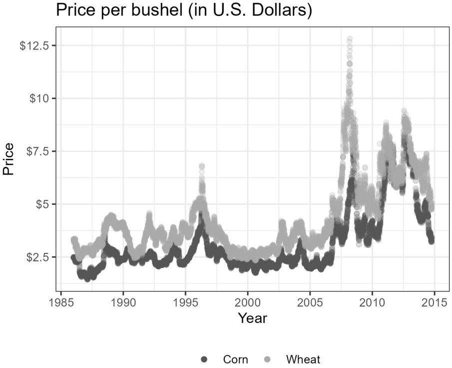
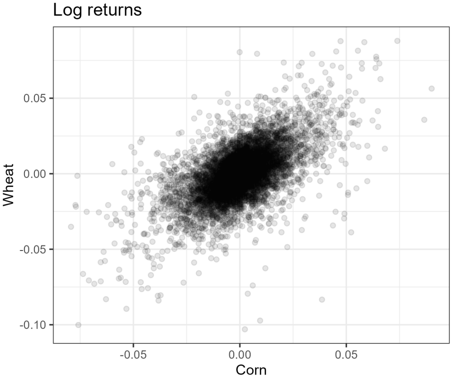


FIGURE: Smoothed volatility via Normal formulation: our proposal and STAN

A bivariate dynamic normal model for the joint returns of wheat and corn



A bivariate dynamic normal model for the joint returns of wheat and corn



A BIVARIATE NORMAL MODEL FOR THE RETURNS OF CORN AND WHEAT

Suppose that we have a sequence of pairs of random variables y_{1t},y_{2t} so that:

$$y_{1t}, y_{2t} \sim \mathcal{N}(\mu_{1t}, \mu_{2t}, \sigma^2_{2t}, \sigma^2_{2t}, \rho_t),$$

$$\begin{bmatrix} \mu_{1t} \\ \mu_{2t} \\ \ln(\sigma^2_{2t}) \\ \ln(\sigma^2_{2t}) \\ logit(\frac{(\rho_t+1)}{2}) \end{bmatrix} = \lambda_t = F_t'\theta_t$$

A BIVARIATE NORMAL MODEL FOR THE RETURNS OF CORN AND WHEAT

Notice that the proposed approach can only fit univariate Normal models. Let's say that we have observed all data until time t, and that $\pi(\theta_t|\mathcal{D}_{t-1})$ is known. It follows that:

$$\begin{aligned}\pi(\theta_t|\mathcal{D}_t) &= \pi(\theta_t|\mathcal{D}_{t-1}, y_{1t}, y_{2t}) \\ &\propto f(y_{1t}, y_{2t}|\theta_t) \times \pi(\theta_t|\mathcal{D}_{t-1}) \\ &= f(y_{2t}|y_{1t}, \theta_t) \times f(y_{1t}|\theta_t) \times \pi(\theta_t|\mathcal{D}_{t-1})\end{aligned}$$

A bivariate normal model for the returns of corn and wheat

Notice that $f(y_{1t}|\theta_t) \times \pi(\theta_t|\mathcal{D}_{t-1})$ can be obtained using our approach and, since $f(y_{1t}|\theta_t) \times \pi(\theta_t|\mathcal{D}_{t-1}) \propto \pi(\theta_t|\mathcal{D}_{t-1}, y_{1t})$, we can write:

$$\begin{aligned}
\pi(\theta_t|\mathcal{D}_t) &= f(y_{2t}|y_{1t}, \theta_t) \times f(y_{1t}|\theta_t) \times \pi(\theta_t|\mathcal{D}_{t-1}) \\
&\propto f(y_{2t}|y_{1t}, \theta_t) \times \pi(\theta_t|\mathcal{D}_{t-1}, y_{1t}),
\end{aligned}$$

A BIVARIATE NORMAL MODEL FOR THE RETURNS OF CORN AND WHEAT

Then, treating $\pi(\theta_t|\mathcal{D}_{t-1}, y_{1t})$ as a prior for θ_t (in the sense that it is the knowledge we have of θ prior to observing y_{2t}), we can use our approach to obtain $f(y_{2t}|y_{1t}, \theta_t) \times \pi(\theta_t|\mathcal{D}_{t-1}, y_{1t})$, since:

$$y_{2t}|y_{1t} \sim \mathcal{N}(\bar{\mu}_t, \bar{\sigma}_t^2),$$

where:

$$\bar{\mu}_t = \mu_{2t} + \rho_t \frac{\sigma_{2t}}{\sigma_{1t}}(y_{1t} - \mu_{1t}),$$
$$\bar{\sigma}_t^2 = (1 - \rho_t^2)\sigma_{2t}^2,$$

A BIVARIATE NORMAL MODEL FOR THE RETURNS OF CORN AND WHEAT

Notice that μ_{1t}, μ_{2t}, σ^2_{1t},σ^2_{2t} and ρ_t are functions of θ_t, so we can also write $\bar{\mu}_t$ and $\bar{\sigma}^2_t$ as functions of θ_t.

But then we find one difficulty: to use the proposed approach to obtain $f(y_{2t}|y_{1t}, \theta_t) \times \pi(\theta_t|\mathcal{D}_{t-1}, y_{1t})$, we would need to write $\bar{\mu}_t$ and $\ln(\bar{\sigma}^2_t)$ as linear functions of θ_t.

To solve this problem, we apply a linearization to $\bar{\mu}_t$ and $\ln(\bar{\sigma}^2_t)$

A bivariate dynamic normal model for the joint returns of wheat and corn

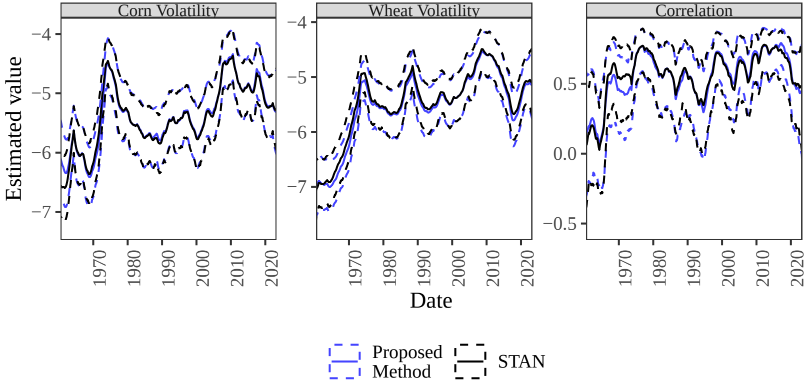


Figure: Estimation for the volatility for Corn and Wheat (left and center) and the correlation between the series (right).

A BIVARIATE NORMAL MODEL FOR THE RETURNS OF CORN AND WHEAT
COMPUTATIONAL TIMES

Fitting the last 5 years of returns: (1,204 observations):
STAN: 1414.46 secs (23.57 mins)
Our proposal: 2.52 secs

Fitting the last 30 years of returns (7,252 observations):
Our proposal: 16.64 secs

MULTINOMIAL DGLM FOR VARICELLA HOSPITAL ADMISSIONS IN BRAZIL

$Y_t|\eta_{1t},\ldots,\eta_{kt} \sim$ Multinomial$(\eta_{1t},\eta_{2t},\ldots,\eta_{dt})$ where $\eta_{dt} = (1-\eta_{1t}-\ldots-\eta_{kt})$, $d = k+1$.

$$\begin{bmatrix} \lambda_{1t} \\ \vdots \\ \lambda_{kt} \end{bmatrix} = \begin{bmatrix} log\left(\frac{\eta_{1t}}{\eta_{d,t}}\right) \\ \\ log\left(\frac{\eta_{kt}}{\eta_{d,t}}\right) \end{bmatrix} = F_t'\theta_t.$$

Consider the same structure for each category:

$$\begin{aligned} F_{jt} &= = [1,0,1,0,1,1]\,, \;\; j = 1,\ldots,k \\ G_{jt} &= blockdiag\left[G_{trend}, G_{seasonality}, 1, 1\right] \end{aligned}$$

$$G_{trend} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, G_{seasonality} = \begin{bmatrix} \cos(\omega) & sin(\omega) \\ -sin(\omega) & \cos(\omega) \end{bmatrix}, \omega = \frac{2\pi}{12}.$$

$F_t = blockdiag\left[F_{1t},\ldots,F_{kt}\right]$, $G_t = blockdiag\left[G_{1t},\ldots,G_{kt}\right]$.

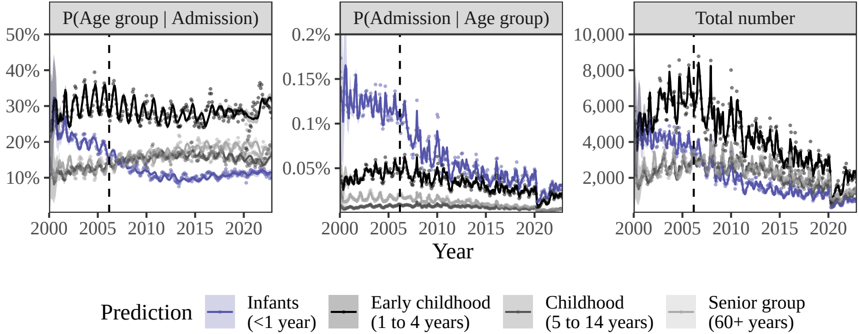


Figure: Observed values (solid circles) compared with one-step-ahead prediction.

MULTIVARIATE POISSON DGLM FOR GASTROENTERITIS HOSPITAL ADMISSIONS IN BRAZIL

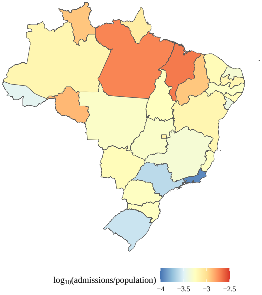


FIGURE: The $\log_{10}$ rate of hospital admission on Brazil by state, from 2010 to 2022.

MULTIVARIATE POISSON DGLM FOR GASTROENTERITIS HOSPITAL ADMISSIONS IN BRAZIL

$$
\begin{aligned}
Y_{it}|\eta_{it} &\sim Poisson(\eta_{it}) \\
\ln\{\eta_{it}\} &= \lambda_{it} = \theta_{1,t} + u_{i,t} + S_{i,t} + \epsilon_{i,t}, \\
\theta_{1,t} &= \theta_{1,t-1} + \theta_{2,t-1} + \omega_{1,t}, \\
\theta_{2,t} &= \theta_{2,t-1} + \omega_{2,t}, \\
\begin{bmatrix} u_{i,t} \\ v_{i,t} \end{bmatrix} &= R \begin{bmatrix} u_{i,t-1} \\ v_{i,t-1} \end{bmatrix} + \begin{bmatrix} \omega^u_{i,t} \\ \omega^u_{i,t} \end{bmatrix}, \\
\epsilon_t &\sim \mathcal{N}(0, \sigma^2_t), \\
S_{1,1}, ..., S_{r,1} &\sim CAR(\tau),
\end{aligned}
$$

where $r = 27$ is the number of areas within our dataset.

Multivariate Poisson DGLM for gastroenteritis hospital admissions in Brazil

- Currently, the kDGLM package does not offer support for sequential estimation of τ, the parameter associated with the CAR prior.
- A study is being developed to address this limitation.
- For now, we conducted a sensitivity analysis to determine an optimal value for τ. The optimal value was $\tau = 0.005$.

MULTIVARIATE POISSON DGLM FOR GASTROENTERITIS HOSPITAL ADMISSIONS IN BRAZIL

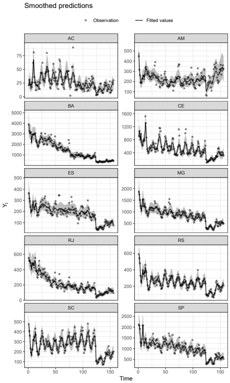


FIGURE: The time series of hospital admissions by gastroenteritis of some Brazilian states and smoothed means, from 2010 to 2022.

Multivariate Poisson DGLM for gastroenteritis hospital admissions in Brazil

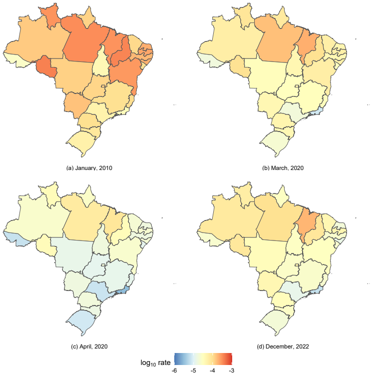


Figure: The smoothed estimation for the $\log_{10}$ hospital admissions rate by gastroenteritis in Brazilian states at 4 key moments

MULTIVARIATE POISSON DGLM FOR GASTROENTERITIS HOSPITAL ADMISSIONS IN BRAZIL

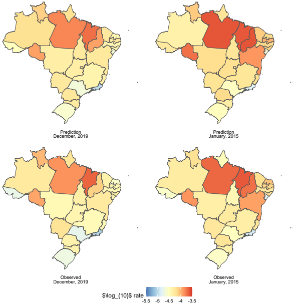


FIGURE: The upper row displays the $\log_{10}$ predictions for the hospital admissions rate due to gastroenteritis in Brazilian states for January 2015 and December 2019, based on information available up to 3 months prior. The lower row shows the observed $\log_{10}$ rate of hospital admissions.

1. Introduction
2. Dynamic Generalized Linear Model
3. Our Proposal: sequential learning on k-DGLMs
4. Applications
5. References

REFERENCES

- Alves, M.B., Migon, H. S., Santos Jr, S. V., Marotta. R. (2024).k-parametric Dynamic Generalized Linear Models: a sequential approach via Information Geometry. https://doi.org/10.48550/arXiv.2201.05387
- Amari, S. (2016). Information geometry and its applications. Springer
- Berry, L. R. and West, M. (2020). Bayesian forecasting of many count-valued time series. Journal of Business Economic Statistics, 38(4):872—887.
- Cargnoni, C., M, P., and West, M. (1997). Bayesian forecasting of multinomial time series through conditionally Gaussian dynamic models. Journal of the American Statistical Association, 92(438):640— 647.
- Carter, C. K. and Kohn, R. (1994). On Gibbs sampling for state space models. Biometrika, 81(3):541—553.
- Durbin, J. and Koopman, S. J. (2002). A simple and efficient simulation smoother for state space time series analysis. Biometrika, (894):603—15.
- Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. Journal of time series analysis, 15(2):183—202.
- Gamerman, D. (1997). Sampling from the posterior distribution in generalized linear mixed models. Statistics and Computing, 7(1):57â68.
- Gamerman, D. (1998). Markov chain Monte Carlo for dynamic generalised linear models. Biometrika, 85(1):215—227.
- Gamerman, D., dos Santos, T. R., and Franco, G. C. (2013). A non-Gaussian family of state-space models with exact marginal likelihood. Journal of Time Series Analysis, 34(6):62—645.
- Grunwald, G. K., Raftery, A. E., and Guttorp, P. (1993). Time series of continuous proportions. Journal of the Royal Statistical Society: Series B (Methodological), 55(1):103—116.
- Santos Jr, S. V., Alves, M.B., Migon, H. S. (2024). kDGLM: a R package for Bayesian analysis of Generalized Dynamic Linear Models. https://doi.org/10.48550/arXiv.2403.13069
- Souza, M. A. d. O., Migon, H. d. S., and Pereira, J. (2016). Extended dynamic generalized linear models: The two-parameter exponential family. Computational Statistics Data Analysis.
- Terui, N., Ban, M., and Maki, T. (2010). Finding market structure by sales count dynamicsâMultivariate structural time series models with hierarchical structure for count dataâ. Annals of the Institute of Statistical Mathematics, 62(1):91—107.
- West, M., Harrison, P. J., and Migon, H. S. (1985). Dynamic generalized linear models and Bayesian forecasting. Journal of the American Statistical Association, 80(389):73—83.

Link for the package (soon it will be available on CRAN-R):

https://github.com/silvaneojunior/kDGLM

Thank you!

mariane@im.ufrj.br