Supplementary Material for: An Efficient Sequential Approach for k-parametric Dynamic Generalized Linear Models

Abstract

This paper introduces a novel sequential inferential method for Bayesian dynamic generalized linear models (DGLM), extending them to k-parametric exponential families and accommodating both univariate and multivariate responses. The method efficiently handles various response types, including multinomial, gamma, normal, and Poisson distributions, leveraging the conjugate and predictive structures of the exponential family. It enables the assignment of dynamic predictors to the k parameters that govern the response. By incorporating information geometry concepts, such as the projection theorem and Kullback-Leibler divergence, it aligns with recent variational inference advances. Applications to synthetic and real datasets highlight its computational efficiency and scalability, surpassing non-sequential methods like Markov Chain Monte Carlo. The R package kDGLM, developed in conjunction with this investigation, is available for applied researchers, enabling easy implementation of the proposed method for k-parametric dynamic generalized models.

Keywords: Bayesian analysis, Dynamic models, Information geometry, Monitoring and intervention, Sequential analysis.

S.1 A Poisson model for quarterly sales

This section presents a model for univariate, uniparametric responses. Unlike methods based on local levels, where only the level evolves dynamically, our goal is to emphasize the versatility of the proposed method, which allows the assignment of temporal dynamics for all latent components of the predictor. We highlight the lower computational cost involved in our proposal, in comparison with a local level method (Gamerman et al., 2013). Finally, we show that, in the single parameter case, our method is equivalent to the one suggested by West et al. (1985), in terms of model flexibility, computational cost and resulting inference.

For the analytical development of our proposed approach for Poisson responses, we follow the steps described in Section 3.2. Let $y|\eta \sim Po(\eta)$, $\eta > 0$. The conjugate prior is a gamma density $q(\eta|\boldsymbol{\tau}) = \frac{1}{\eta}\exp[\tau_1 \log(\eta) - \tau_0\eta - b(\tau_0, \tau_1)]$, where $b(\tau_0, \tau_1) = \log(\Gamma(\tau_1)) - \tau_1 \log(\tau_0)$. Consider the linear predictor $\lambda = \log(\eta) = \boldsymbol{F}'\boldsymbol{\theta}$ and the conjugate prior sufficient statistics vector: $\boldsymbol{H}_q(\eta) = (\log(\eta), \eta)'$. Following Algorithm 1 and using the fact that $E_q[\boldsymbol{H}_q(\eta)] = \nabla b(\tau_0, \tau_1)$ as well as approximation (4) for the digamma function $\gamma(\cdot)$, the solution of the system $E_q[\boldsymbol{H}_q(\eta)] = E_p[\boldsymbol{H}_q(\eta)]$ in Step 2.2 results in $\tau_0 = \tau_1 \exp[-(f + Q/2)]$ and $\tau_1 = (1 + \sqrt{1 + 2Q/3})/(2Q)$. The conjugate prior hyperparameters are updated to $\tau_1^* = \tau_1 + y$ and $\tau_0^* = \tau_0 + 1$. Step 4.2 is completed by solving the system $E_q[\boldsymbol{H}_q(\boldsymbol{\lambda})] = E_p[\boldsymbol{H}_q(\boldsymbol{\lambda})]$, where p is the updated distribution, obtaining: $f^* = \gamma(\tau_1^*) - \log(\tau_0^*)$ and $Q^* = \gamma'(\tau_1^*)$.

The updated moments of the states $\boldsymbol{\theta}$ are trivially obtained by applying normal distribution properties and the predictive distribution for y_t is a Poisson-Gamma$(\tau_0, \tau_1, 1)$, which can be evaluated once step 2.2 is completed.

We apply a Poisson dynamic model to predict a time series of quarterly sales with a growth trend and stochastic seasonal pattern, as detailed in West and Harrison (1997, Chap. 8, p. 268). True data are shown as solid circles in Figure S.1. We compare this with West et al. (1985), using a Poisson log-linear model with linear growth and two pairs of harmonics for a Fourier description of the seasonal pattern.

Let $y_t|\eta_t \sim Poisson(\eta_t)$; $\log(\eta_t) = \boldsymbol{F}_t'\boldsymbol{\theta}_t$; $\boldsymbol{\theta}_t = \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t$, $\boldsymbol{\omega}_t \sim N(\boldsymbol{0}, \boldsymbol{W}_t)$, with $\boldsymbol{F}_t' = [1, 0, 1, 0, 1, 0]$; $\boldsymbol{G}_t = diag[\boldsymbol{G_0}, \boldsymbol{G_1}, \boldsymbol{G_2}]$ and

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

The prior specification is as follows: $\boldsymbol{\theta}_0|D_0 \sim N(m_0 = 0, C_0 = 1)$, and we employ a diagonal block discount factor matrix for $\boldsymbol{W}_t$ covariance evolution. For the conjugate updating approach of West et al. (1985) we specify the same first- and second-order prior moments. Both methods begin with $\boldsymbol{\theta}_0|D_0$, updated to the prior $\boldsymbol{\theta}_1|D_0$, inducing a predictor's prior structure.

Figure S.1 exhibits smoothed mean responses, illustrating the proposed method's efficiency in managing predictive structures with multiple dynamic components. Our approach accommodates various stochastic patterns typical in real time series while preserving sequential analysis.

We also compare our approach with the local level method proposed by Gamerman et al. (2013), who introduced Gamma Family Dynamic Models (GFDM) integrating Smith and Miller (1986) and Harvey and Fernandes (1989). GFDM combines a local steady structure (a dynamic level) with a fixed-coefficient predictor, applicable to models such as Poisson, gamma, Weibull, and normal, as well as others outside the exponential family. Unlike our model, GFDM does not extend beyond the level dynamics. To align with our model's structure, we define regressors $x_1 = t, x_2 = \cos(wt), x_3 = \sin(wt), x_4 = \cos(2wt), x_5 = \sin(2wt)$, where $t = 1, \ldots, 35$ and $w = 2\pi/4$, applying the local level approach of Gamerman et al. (2013) to the following model: $y_t|\eta_t \sim Poisson(\eta_t)$; $\eta_t = \alpha_t \exp(\beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + \beta_5 x_{5t})$. The model is completed with a $Gamma(0.34, 0.04)$ prior for η_0 to standardize prior settings across methods and a $Uniform(0, 1)$ prior to manage the increase in the uncertainty associated with the local level approach evolution, impacting dynamic level smoothness (Gamerman et al., 2013). Parameters β_k, for $k = 1, \ldots, 5$, follow independent $Normal(0, 1)$ priors. The `NGSSEML` package (Santos et al., 2021) performed inference with MCMC steps for the static coefficients, and the `coda` package (Plummer et al., 2006) facilitated convergence diagnosis, burn-in, and thinning. Both our method and that of West et al. (1985) involving conjugate updating provided smoothed estimates and predictions in under one second. Figure S.1 illustrates the expected value of the smoothed mean response function obtained from our method and Gamerman et al. (2013), excluding West et al. (1985)'s results, which are almost identical to ours. Our method and conjugate updating do not require sampling, resulting in computational efficiency, in contrast to Gamerman et al. (2013)'s local level method, which took 818.803 seconds (effective sample

size 1,700), compared to 0.024 second for our approach and West et al. (1985) (Table S.1).

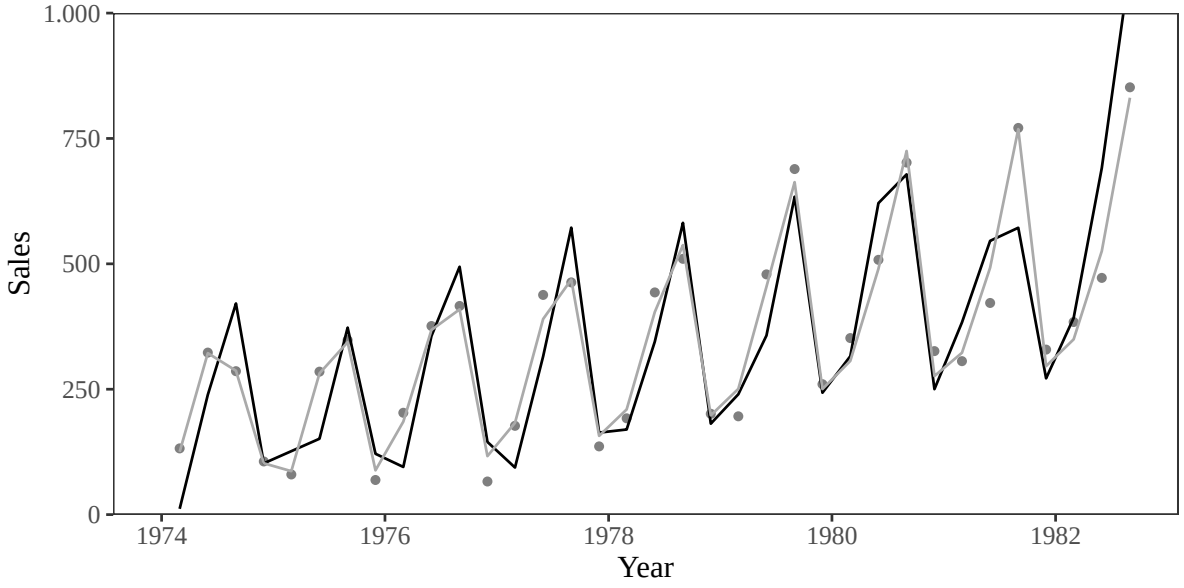

Figure S.1: Quarterly sales: observed time series (solid circles) and the expected value of the smoothed mean response function via sequential kDGLM (solid grey line) and local level method (solid black line).

Metric	Local Level	Conjugate Update	kDGLM
Mean Absolute Error	82.323	27.21166	**26.86886**
Relative Absolute Error	0.2661	0.1084	**0.1077**
Log Predictive Likelihood	-214.006	-	**-148.8231**
Computational time*	818.803s	**0.024s**	**0.024s**

Table S.1: Model comparison based on the smoothed mean response function and several metrics. *Time measured using R 4.2.0 running Windows 10 x64 (build 19044) in an Intel(R) Core™ i7-8700K CPU @ 3.70GHz with 16 GB RAM @ 2400 MHz.

According to comparison of the metrics in Table S.1, conjugate updating and our sequential approach are nearly equivalent. The local level model with deterministic seasonality, which is adjustable by the method proposed by Gamerman et al. (2013), achieved the weakest performance in terms of computational time, as well as regarding fitting and prediction, exhibiting higher uncertainty of smoothed expected values. Unlike West et al. (1985), our information geometry approach enables log predictive likelihood evaluation, which is

an inherent advantage. The local level approach lacks seasonal component dynamics, as shown in Figure S.2, where the trend fit by Gamerman et al. (2013)'s method "attempts to compensate" for seasonal inflexibility, showing that our method allows for models that capture structural components with greater precision.

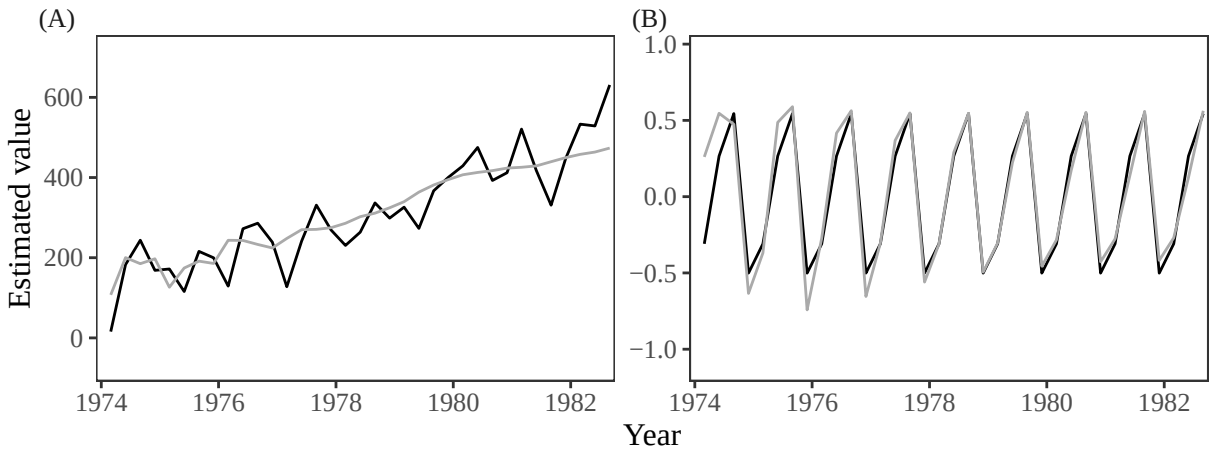

Figure S.2: Structural decomposition - smoothed mean for seasonal and trend components: kDGLM (gray) and local level (black).

S.2 Analytical development of the proposed method for Bernoulli responses

Let $\eta = \pi$ and $y|\pi \sim Ber(\pi)$, $0 < \pi < 1$. $\psi(\pi) = \log(\frac{\pi}{1-\pi})$. The conjugate prior is a beta density $q(\pi|\boldsymbol{\tau}) = exp\{\tau_1 \log(\frac{\pi}{1-\pi}) + \tau_0 \log(1-\pi) - [\log(\Gamma(\tau_1+1)) + \log(\Gamma(\tau_0-\tau_1+1)) - \log(\Gamma(\tau_0+2))]\}$, where $b(\tau_0, \tau_1) = [\log(\Gamma(\tau_1+1)) + \log(\Gamma(\tau_0-\tau_1+1)) - \log(\Gamma(\tau_0+2))]$. Consider the linear predictor $\lambda = logit(\pi) = \boldsymbol{F}'\boldsymbol{\theta}$. Following Algorithm 1 :

- Step 1: Given posterior moments for the states, obtain the prior moments of $\lambda \sim N(f, Q)$.
- Step 2.1: Obtain the conjugate prior sufficient statistics vector: $\boldsymbol{H}'_q(\pi) = (\log(\frac{\pi}{1-\pi}), \log(1-\pi))$.
- Step 2.2: Using the fact that $E_q[\mathbf{H}(\pi)] = \nabla b(\tau_0, \tau_1)$, solve the system $E_q[\boldsymbol{H}_q(\pi)] = E_p[\boldsymbol{H}_q(\pi)]$, where p is the prior density induced by the normal specification for the linear predictors:

$$\begin{cases} \gamma(\tau_1+1) - \gamma(\tau_0-\tau_1+1) & = f \\ \gamma(\tau_0-\tau_1+1) - \gamma(\tau_0+2) & \simeq \log(\frac{1}{1+e^f}) - \frac{Qe^f}{(1+e^f)^2}. \end{cases}$$

A second-order Taylor approximation was applied and the system was solved using the Newton-Raphson method, with negligible computational cost.

- Step 3: Update the hyperparameters of the conjugate specification: $\tau_1^* = \tau_1 + y, \tau_0^* = \tau_0 + 1$.
- Step 4.1: Let q denote a normal density for the linear predictor, with sufficient statistics vector: $\boldsymbol{H}_q' = (\lambda,\ \lambda^2)$.
- Step 4.2: Solve the system $E_q[\boldsymbol{H}_q(\boldsymbol{\lambda})] = E_p[\boldsymbol{H}_q(\boldsymbol{\lambda})]$, where p is the updated distribution, obtaining: $f^* \simeq \gamma(\tau_1^* + 1) - \gamma(\tau_0^* - \tau_1^* + 1)$ and $Q^* = \gamma'(\tau_1^* + 1) + \gamma(\tau_0^* - \tau_1^* + 1)$.
- Step 5: The updated moments of the states $\boldsymbol{\theta}$ are trivially obtained, applying normal distribution properties.

The predictive distribution for y_t is a binomial-beta$(\tau_0, \tau_1, 1)$, which can be evaluated once step 2.2 is completed.

S.3 Results for the simulated example of the dynamic multinomial model

In this section, we evaluate the effect of varying sample sizes T ($T = 2, 5, 20, 50$), as well as different specifications of the number of allocation categories, N, on the approximated inference produced by the sequential kDGLM proposal. We consider synthetic data simulated as described in Section 4.1 and compare the results of our proposed method to those obtained via non-sequential NUTS in Stan.

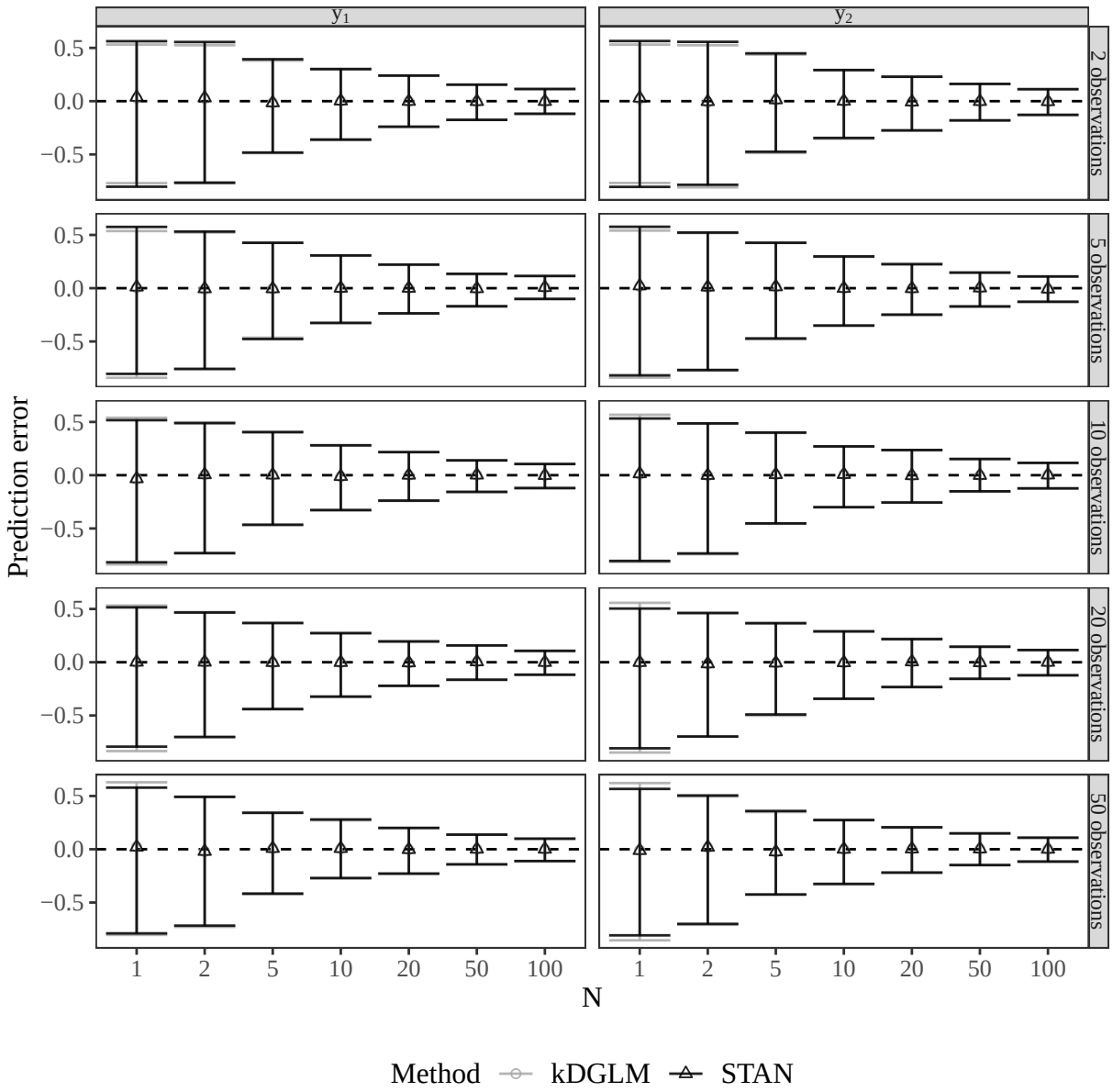

Figure S.3: The bias for the one-step-ahead prediction for $Y_{1,T+1}$ and $Y_{2,T+1}$. Gray circles and black triangles represent the average bias among all samples, while the error bars represent the 0.975 and 0.025 quantiles. To facilitate visualization, the bias is divided by N. Gray: Sequential kDGLM. Black: Stan.

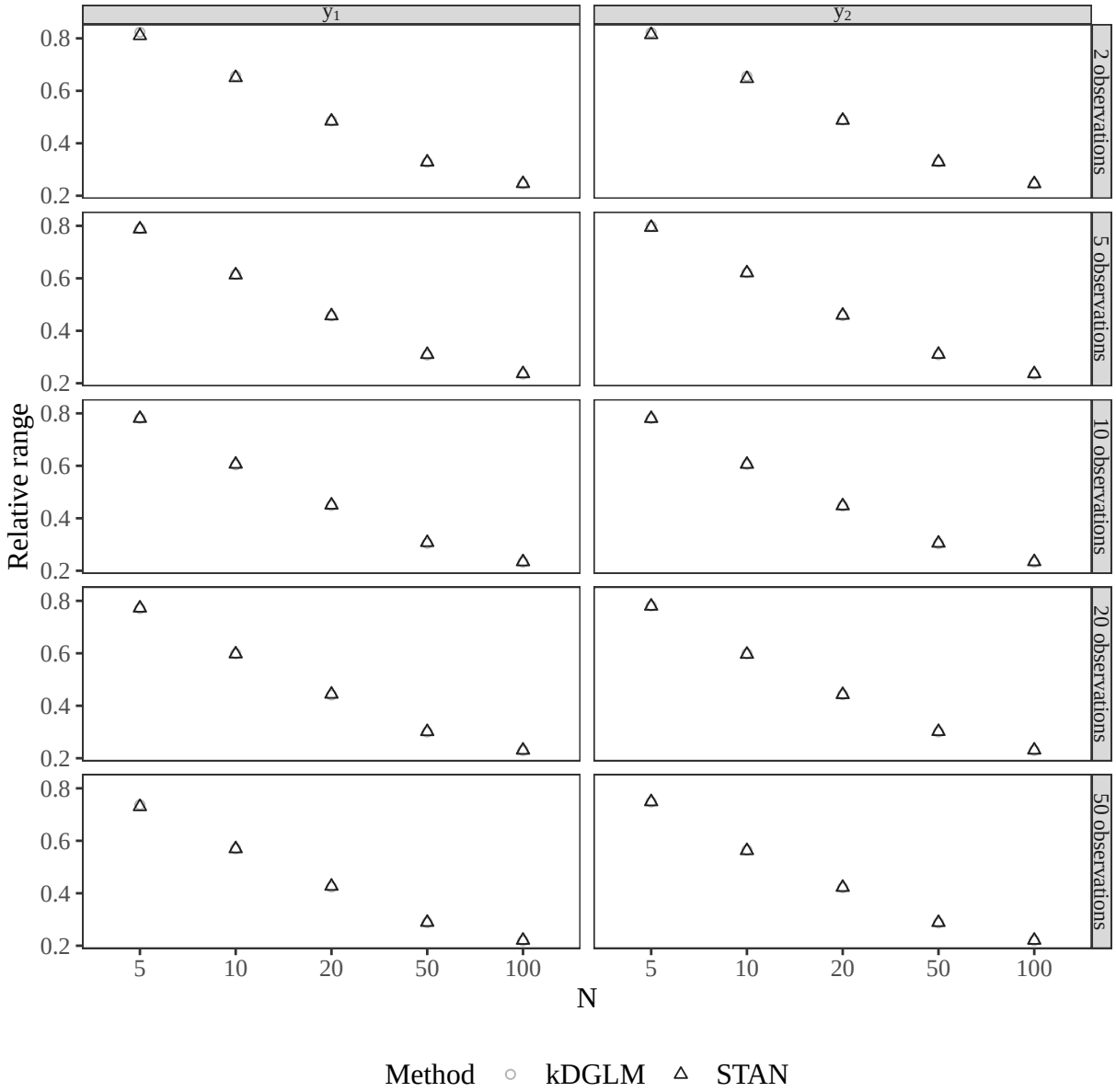

Figure S.4: The range of the 95% credibility interval for one-step-ahead prediction for $Y_{1,T+1}$ and $Y_{2,T+1}$. Gray circles: Sequential kDGLM. Black triangles: Stan.

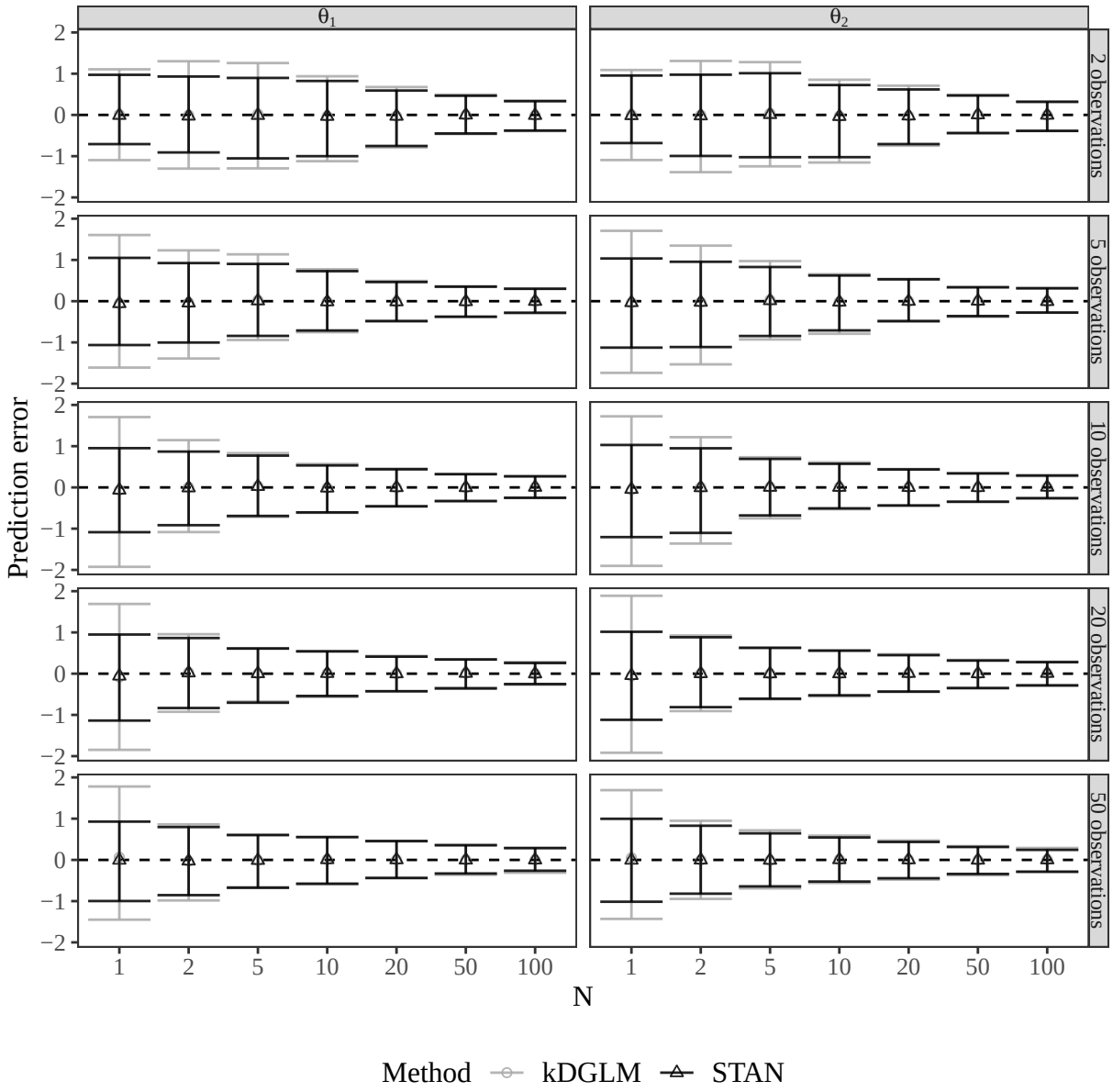

Figure S.5: The bias for the latent states at time T. Gray circles and black triangles represent the average bias among all samples, while the error bars represent the 0.975 and 0.025 quantiles. Gray: Sequential kDGLM. Black: Stan.

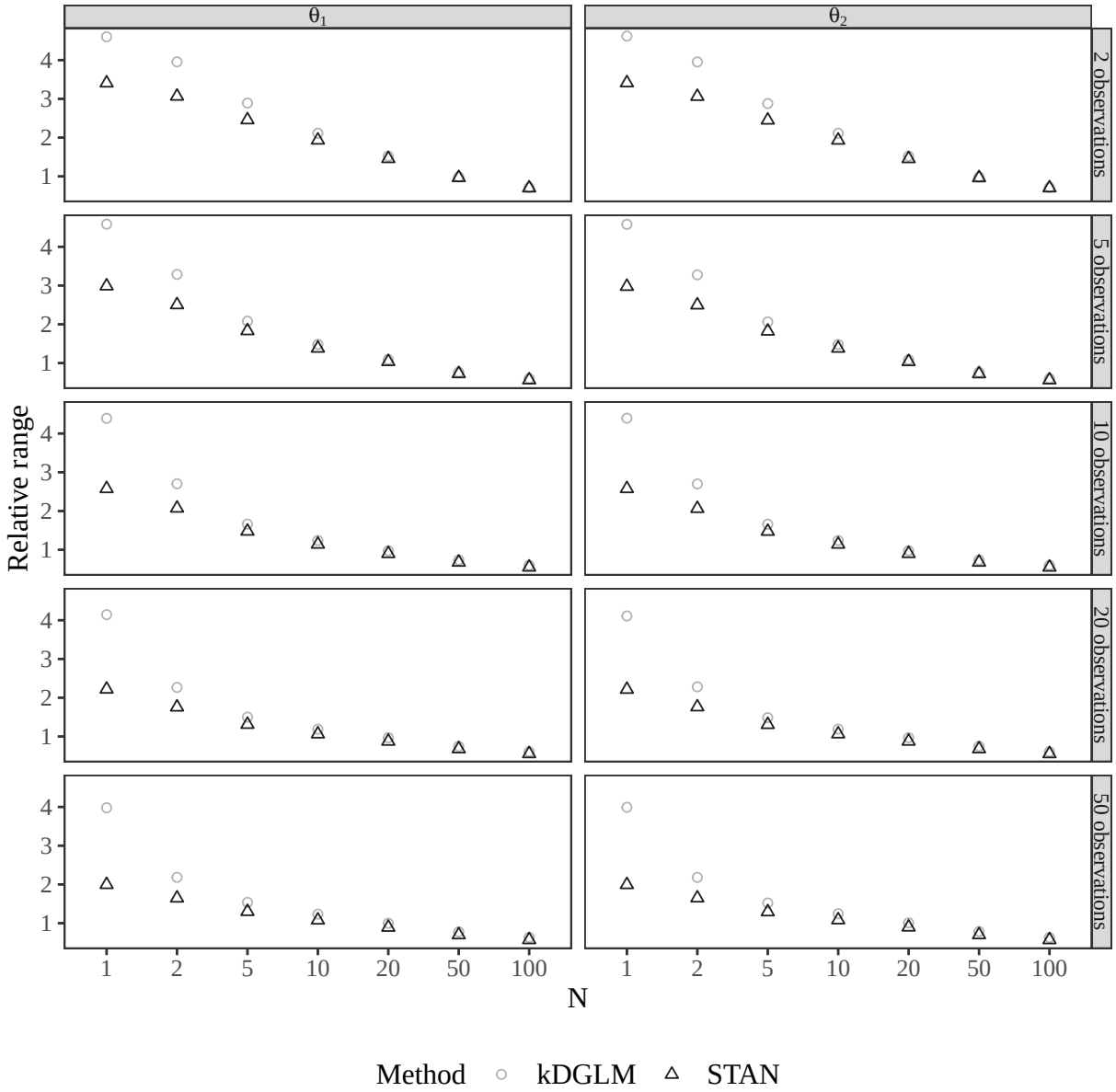

Figure S.6: The range of the 95% credibility interval for the latent states at at time T. Gray: Sequential kDGLM. Black: Stan.

S.4 Analytical development of the proposed method for normal and gamma members of the exponential family

This section provides a detailed description of the analytical developments involved in updating information for the normal and gamma distributions within the kDGLM sequential method. These are the sole instances presented in this work in which the application of the suggested methodology is not straightforward. The outlined procedures are executed at each time point t during the information updating process.

Normal distribution with unknown mean and precision: Consider the general formulation and parametrization defined in Section 4.2 for the normal model with dynamic predictive structure for the mean and precision. Since $\boldsymbol{\lambda} \sim N_2(\boldsymbol{f}, \boldsymbol{Q})$, a system of equations in τ remains to be solved, that is, $E_q[\boldsymbol{H}_q(\mu, \phi)] = E_p[\boldsymbol{H}_q(\mu, \phi)]$, with q denoting a normal-gamma prior distribution and p, the density implied by the joint normality of the linear predictors for the mean ad log-precision. Using the facts that: i) $E_q[\boldsymbol{H}_q(\mu, \phi)] = \nabla b(\tau_1, \tau_2, \tau_3, \tau_0)$; ii) $E_p[\boldsymbol{H}_q(\mu, \phi)] = E_{p(\phi)}[E_{p(\mu|\phi)}\boldsymbol{H}_q(\mu, \phi)]$ which can be obtained by normal conditioning properties; iii) under p, $\phi \sim LN(f_2, q_2)$ and $\ln \phi \sim N(f_2, q_2)$, results in the system:

$$\begin{cases} \frac{(2\tau_0+1)\tau_2^2}{2\tau_1\tau_2^2 - 8\tau_1^2\tau_3} - \frac{1}{2\tau_1} & = & \exp(f_2 + q_2/2)[(f_1 + q_{12})^2 + q_1] \\ \frac{-(2\tau_0+1)\tau_2}{\tau_2^2 - 4\tau_1\tau_3} & = & \exp(f_2 + q_2/2)(f_1 + q_{12}) \\ \frac{4\tau_1(\tau_0+1/2)}{\tau_2^2 - 4\tau_1\tau_3} & = & \exp(f_2 + q_2/2) \\ \gamma(\tau_0 + 1/2) - \ln\left(\frac{\tau_2^2}{4\tau_1} - \tau_3\right) & = & f_2 \end{cases}, \tag{S.1}$$

with $\gamma(\cdot)$ denoting the digamma function. This system can be analytically solved, as follows: applying the parametrization:

$$c_0 = -2\tau_1, \quad \mu_0 = -\frac{\tau_2}{2\tau_1}, \quad \frac{d_0}{2} = \frac{\tau_2^2}{4\tau_1} - \tau_3, \quad \frac{n_0}{2} = \tau_0 + 1/2, \tag{S.2}$$

to the third equation in (S.1), it follows that $n_0/d_0 = \exp(f_2 + q_2/2)$. Multiplying the third equation by $-\frac{2\tau_2}{4\tau_1} = \mu_0$ we obtain the second equation in (S.1) resulting in $\mu_0 = f_1 + q_{12}$. Doing some algebra in the first equation of (S.1), it follows that $c_0 = [q_1 \exp(f_2 + q_2/2)]^{-1}$. Finally, applying the approximation $\gamma(u) \approx \ln(u) - 1/(2u)$ to the fourth equation of the

system, results in $n_0 = 2/q_2$. Applying the reparametrization in (S.2) we obtain the following closed analytical expressions:

$$\begin{cases} \tau_0 &= 1/q_2 - 1/2 \\ \tau_1 &= -[2q_1 \exp(f_2 + q_2/2)]^{-1} \\ \tau_2 &= (f_1 + q_{12})[q_1 \exp(f_2 + q_2/2)]^{-1} \\ \tau_3 &= -(f_1 + q_{12})^2[2q_1 \exp(f_2 + q_2/2)]^{-1} - [q_2 \exp(f_2 + q_2/2)]^{-1}. \end{cases}$$

After observing $\boldsymbol{y}$, we obtain $\boldsymbol{\tau}^* = (\tau_1^*, \tau_2^*, \tau_3^*, \tau_0^*) = (\tau_1 - 1/2, \tau_2 + y, \tau_3 - y^2/2, \tau_0 + 1/2)$, the updated canonical parameters of a normal-gamma posterior density for (μ, ϕ), and need to evaluate the parameters $(\boldsymbol{f}^*, \boldsymbol{Q}^*)$ of the posterior distribution of the linear predictors that are compatible with them, solving: $E_q[\boldsymbol{H}_q] = E_p[\boldsymbol{H}_q]$, now considering that p is a normal-gamma density for (μ, ϕ) and q is a bivariate normal for $(\mu, \ln \phi)$. $\boldsymbol{H}_q' = (\boldsymbol{\lambda}, \boldsymbol{\lambda}\boldsymbol{\lambda}')$, so the following system is trivially solved:

$$\begin{cases} f_1^* &= -\frac{\tau_2^*}{2\tau_1^*} \\ f_2^* &= \gamma(\tau_0^* + 1/2) - \ln\left(\frac{(\tau_2^*)^2}{4\tau_1^*} - \tau_3^*\right) \\ Q_{11}^* &= \frac{\tau_2^{*2} - 4\tau_1\tau_3}{4\tau_0\tau_1 - 2\tau_1} \\ Q_{12}^* &= 0 \\ Q_{22}^* &= \gamma'(\tau_0^* + 1/2). \end{cases}$$

Notice that, no matter the values of $\tau_0^*, \tau_1^*, \tau_2^*$ and τ_3^*, $Q_{12}^* = 0$, since the parameters that minimize the KL divergence from p to q must satisfy $Cov_p[\mu, \ln(\phi)] = Cov_q[\mu, \ln(\phi)]$ and, as a property of the normal-gamma distribution, we have that (μ, ϕ) are uncorrelated (although not independent), consequently, it can be shown that $(\mu, \ln \phi)$ are also uncorrelated.

Using the reparametrization $c_0^* = -2\tau_1^*$, $\mu_0^* = -\frac{\tau_2^*}{2\tau_1^*}$, $\frac{d_0^*}{2} = \frac{\tau_2^{*2}}{4\tau_1^*} - \tau_3^*$ and $\frac{n_0^*}{2} = \tau_0^* + 1/2$, we obtain:

$$\begin{cases} f_1^* &= \mu_0^* \\ f_2^* &= \gamma(n_0^*/2) - \ln(d_0^*/2) \\ Q_{11}^* &= \frac{d_0^*/2}{c_0^*(n_0^*/2 - 1)} \\ Q_{12}^* &= 0 \\ Q_{22}^* &= \gamma'(n_0^*/2). \end{cases}$$

An important point to discuss is the equation that defines Q_{11}^*. Specifically, notice that Q_{11}^* is not defined for $n_0^* \leq 2$. Indeed, by the Projection Theorem, the parameters

that minimize the KL divergence from p to q must satisfy $Var_p[\mu] = Var_q[\mu]$, but in the conjugated posterior distribution $\mu|\phi \sim N(\mu_0^*, (c_0^*\phi)^{-1})$ and $\phi \sim \mathcal{G}(n_0^*/2, d_0^*/2)$, which implies that $\mu \sim t\left(n_0^*, \mu_0^*, \frac{d_0^*/2}{c_0^* n_0^*/2}\right)$, so that when $n_0^* \leq 2$, $Var_p[\mu] = +\infty$, and the system $E_q[\boldsymbol{H}_q] = E_p[\boldsymbol{H}_q]$ has no valid solutions. Indeed, one can always reduce the KL divergence from p to q by increasing Q_{11}^*. This situation is quite undesirable since it compromises the use of our proposed methodology. Yet, we can restrict the parameter space for the normal posterior distribution to guarantee that, inside that restricted space, we always have a minimum for the divergence from p to q. A natural restriction for the parameter space is to set $Q_{11}^* = \frac{d_0^*/2}{c_0^* n_0^*/2}$, since:

- This restriction guarantees that the scale parameter for the marginal distribution of μ is identical in both the normal and the conjugated distribution.
- for a large value of n_0^* (which we expect to have for a reasonable sample size), $\frac{d_0^*/2}{c_0^* n_0^*/2} \approx \frac{d_0^*/2}{c_0^*(n_0^*/2-1)}$, i.e., for large values of n_0^* this restriction has no significant effect, in the sense that the optimum in the restricted space will be very close to the global optimum.
- $Q_{11}^* = \frac{d_0^*/2}{c_0^* n_0^*/2}$ becomes numerically well behaved for any possible value of n_0^*, since, after updating our knowledge of μ, it is guaranteed that $n_0^* \geq 1$, avoiding a division by values close to zero.

Gamma distribution with unknown shape α: Let $\boldsymbol{\eta} = (\alpha, \mu)$ and $(y|\boldsymbol{\eta}) \sim \mathcal{G}(\alpha, \alpha/\mu)$, $\alpha, \mu > 0$. Unlike other particular cases discussed in this work, the normalising constant of the joint prior for the pair (α, μ) is unknown and the conjugate prior is $\pi(\alpha, \mu) \propto \exp\left\{\tau_0\left(\alpha \ln\left(\frac{\alpha}{\mu}\right) - \ln(\Gamma(\alpha))\right) + \tau_1\alpha - \tau_2\frac{\alpha}{\mu}\right\}$. Consider a vector of linear predictors $\boldsymbol{\lambda}' = (\lambda_1, \lambda_2) = (\ln(\mu), \ln(\phi)) = (F_1'\theta_1, F_2'\theta_2)$. Following Algorithm 1:

- Step 1: Given posterior moments for the states, obtain the prior moments of $\boldsymbol{\lambda} \sim N(\boldsymbol{f}, \boldsymbol{Q})$.
- Step 2.1: Obtain the conjugate prior sufficient statistics vector: $\boldsymbol{H}_q = \left(\alpha\ln(\mu) - \alpha\ln(\alpha) + \ln(\Gamma(\alpha)), \alpha, \frac{\alpha}{\mu}\right)'$.
- Step 2.2: Solve the system $E_q[\boldsymbol{H}_q(\boldsymbol{\eta})] = E_p[\boldsymbol{H}_q(\boldsymbol{\eta})]$, where p is the prior density induced by the normal specification for the linear predictors. It is easily seen that $\mu|\alpha \sim \mathcal{IG}(\tau_0\alpha - 1, \alpha\tau_2)$

where $\mathcal{IG}$ denotes the inverse gamma density, if $\alpha > 1/\tau_0$. Using the fact that $E_q[\boldsymbol{H}_q] = E_q[E_q[\boldsymbol{H}_q|\alpha]]$ and that $E_q[\ln(\mu)|\alpha] = -\gamma(\tau_0\alpha - 1) + \ln(\alpha\tau_2)$, it follows that:

$$\begin{aligned} E_q\left[\frac{\alpha}{\mu}\right] &= E_q\left[\alpha E_q\left[\frac{1}{\mu}|\alpha\right]\right] = E_q\left[\alpha\frac{\tau_0\alpha+1}{\alpha\tau_2}\right] = \frac{\tau_0 E_q\left[\alpha\right]-1}{\tau_2}; \\ E_q[\alpha\ln(\mu)] &= E_q[\alpha E_q[\ln(\mu)|\alpha]] = E_q\left[\alpha(-\gamma(\tau_0\alpha+1)+\ln(\alpha\tau_2))\right]. \end{aligned}$$

Assuming $E_p\left[H_q\right]$ known, and using the fact that $E_q[\alpha] = E_p[\alpha]$ (from the system to be solved), it follows that $\tau_2 = \frac{\tau_0 E_q[\alpha]-1}{E_q\left[\frac{\alpha}{\mu}\right]}$. One can trivially find that $E_q[\alpha\ln(\mu) - \alpha\ln(\alpha) + \ln(\Gamma(\alpha)] = E_q\left[\ln(\Gamma(\alpha) - \alpha\gamma(\tau_0\alpha+1)\right] + E_q\left[\alpha\right]\ln(\tau_2)$. Thus, $E_q[\boldsymbol{H}_q]$ can be expressed in terms of expected values which solely depend on the marginal prior for α: $\pi(\alpha) = \int_0^{+\infty}\pi(\alpha,\mu)d\mu \propto \frac{\alpha^{\tau_0\alpha}}{\alpha^{-\tau_0}\Gamma(\alpha+1)^{\tau_0}}\frac{\Gamma(\tau_0\alpha+1)}{\alpha^{\tau_0\alpha+1}\tau_2^{\tau_0\alpha+1}}\exp\{\tau_1\alpha\}$. Applying Stirling's formula $\Gamma(x+1) \approx \sqrt{2\pi x}\left(\frac{x}{e}\right)^x$ to this last expression results in: $\pi(\alpha) \propto \alpha^{\frac{\tau_0-1}{2}-1}\exp\{-\left[\tau_0\ln\left(\frac{\tau_2}{\tau_0}\right)-\tau_1\right]\alpha\}$, that is, α approximately follows a $\mathcal{G}\left(\frac{\tau_0+1}{2}, \tau_0\ln\left(\frac{\tau_2}{\tau_0}\right)-\tau_1\right)$ marginal conjugate prior, so that:

$$E_q[\alpha] = \frac{\tau_0+1}{2\left(\tau_0\ln\left(\frac{\tau_2}{\tau_0}\right)-\tau_1\right)}.$$

Using a second order approximation for the digamma function, $\gamma(x) \approx \ln(x) - \frac{1}{2x} - \frac{1}{12x^2}$, and the fact that $\gamma(x+1) = \gamma(x) + \frac{1}{x}$ results in:

$$\begin{aligned} -\alpha\gamma(\tau_0\alpha+1) &= -\alpha\left(\gamma(\tau_0\alpha)+\frac{1}{\tau_0\alpha}\right) \\ &\approx -\alpha\left(\ln(\tau_0\alpha) - \frac{1}{2\tau_0\alpha} - \frac{1}{12\tau_0^2\alpha^2} + \frac{1}{\tau_0\alpha}\right) \\ &\approx -\alpha\ln(\alpha) - \alpha\ln(\tau_0) + \frac{1}{12\tau_0^2\alpha} - \frac{1}{2\tau_0}. \end{aligned}$$

Once again using a second order approximation for the digamma function and the fundamental theorem of Calculus:

$$\ln(\Gamma(x)) = \int_1^x \gamma(t)dt \approx \int_1^x \ln(t) - \frac{1}{2t} - \frac{1}{12t^2}dt = x\ln(x) - x - \frac{1}{2}\ln(x) + \frac{1}{12x} + \frac{11}{12}.$$

The stated facts allow us to calculate the expected value of the remaining sufficient statistic, resulting in the following system to be solved:

$$
\begin{cases}
\mathbb{E}_p\left[(\alpha\ln(\mu)-\alpha\ln(\alpha)+\ln(\Gamma(\alpha))\right] & = & \frac{\tau_0+1}{2(\tau_0\ln(\frac{\tau_2}{\tau_0})-\tau_1)}(\ln(\tau_2/\tau_0)-1)-\frac{1}{2}\psi\left(\frac{\tau_0+1}{2}\right) \\
 & & +\frac{1}{2}\ln\left(\tau_0\ln\left(\frac{\tau_2}{\tau_0}\right)-\tau_1\right)+\frac{\tau_0+1}{\tau_0-1}\frac{\ln\left(\frac{\tau_2}{\tau_0}\right)-\frac{\tau_1}{\tau_0}}{6} \\
 & & +\frac{11}{12}-\frac{1}{2\tau_0} \\
\exp\{f_1+Q_{11}/2\} & = & \frac{\tau_0+1}{2(\tau_0\ln(\frac{\tau_2}{\tau_0})-\tau_1)} \\
\exp\{f_1-f_2+Q_{11}/2+Q_{12}+Q_{22}/2\} & = & \frac{\tau_0(\tau_0+1)-1}{\tau_2 2(\tau_0\ln(\frac{\tau_2}{\tau_0})-\tau_1)}
\end{cases}
$$

with $\mathbb{E}_p\left[(\alpha\ln(\mu)-\alpha\ln(\alpha)+\ln(\Gamma(\alpha))\right]$ evaluated by numerical integration, with negligible cost.

- Step 3: Update the hyperparameters of the conjugate specification: $\boldsymbol{\tau}^* = (\tau_0^*, \tau_1^*, \tau_2^*) = (\tau_0+1, \tau_1+\ln(y), \tau_2+y)$.

- Step 4.1: Let q denote a multivariate normal density for the vector of linear predictors, with sufficient statistics vector: $\boldsymbol{H}_q' = (\boldsymbol{\lambda},\ \boldsymbol{\lambda\lambda'})$.

- Step 4.2: Solve the system $E_q[\boldsymbol{H}_q(\boldsymbol{\lambda})] = E_p[\boldsymbol{H}_q(\boldsymbol{\lambda})]$, where p is the updated distribution, obtaining:: $f_1^* = E_p[\ln(\mu)]$, $f_2^* = E_p[\ln(\alpha)]$; $Q_{11}^* = V_p[\ln(\mu)]$, $Q_{12}^* = Cov_p[\ln(\mu),\ln(\alpha)]$, $Q_{22}^* = E_p[\ln(\alpha)]$. Using the fact that α approximately follows a $\mathcal{G}\left(\frac{\tau_0^*+1}{2}, \tau_0^*\ln\left(\frac{\tau_2^*}{\tau_0^*}\right)-\tau_1^*\right)$ conjugate marginal posterior, it follows that:

$$
\begin{aligned}
E_p[\ln(\alpha)] &\approx \gamma\left(\frac{\tau_0^*+1}{2}\right)-\ln\left(\tau_0^*\ln\left(\frac{\tau_2^*}{\tau_0^*}\right)-\tau_1^*\right) \\
V_p[\ln(\alpha)] &\approx \gamma'\left(\frac{\tau_0^*+1}{2}\right).
\end{aligned}
$$

Since, *a posteriori*, $\mu|\alpha \sim \mathcal{IG}(\tau_0^*\alpha+1, \alpha\tau_2^*)$, it follows that:

$$
\begin{aligned}
E_p[\ln(\mu)] &= E_p[E_p[\ln(\mu)|\alpha]] = -E_p\left[\gamma(\tau_0^*\alpha+1)-\ln(\alpha\tau_2^*)\right]; \\
V_p[\ln(\mu)] &= V_p[\ln(1/\mu)] = E_p[V_p[\ln(1/\mu)|\alpha]]+V_p[E_p[\ln(1/\mu)|\alpha]] \\
&= E_p[\gamma(\tau_0^*\alpha+1)]+V_p[\gamma(\tau_0^*\alpha+1)-\ln(\alpha\tau_2^*)]; \\
E_p[\ln(\alpha)\ln(\mu)] &= -E_p[\ln(\alpha)\ln(1/\mu)] = -E_p[\ln(\alpha)E_p[\ln(1/\mu)|\alpha]] \\
&= -E_p[\ln(\alpha)(\gamma(\tau_0^*\alpha+1)-\ln(\alpha\tau_2^*))].
\end{aligned}
$$

All the integrals involved in the expected values above are evaluated by numerical integration, with negligible cost.

- Step 5: The updated moments of the states $\boldsymbol{\theta}$ are trivially obtained, applying normal distribution properties.

Since the normalizing constant of the conjugate prior is unknown, we trivially obtain a sample from the predictive distribution, first sampling from the posterior $\boldsymbol{\lambda}$ and then sampling from the observational distribution of $y^*|\boldsymbol{\lambda}$. Since we assume that $\boldsymbol{\lambda}$ follows a multivariate normal distribution, the computational cost of this process is negligible.

S.5 Comparison between kDGLM and NUTS via Stan for the stochastic volatility example

Section 4.2.1 presents a comparison between the results obtained using the sequential kDGLM method proposed in this work and those from a NUTS scheme implemented via Stan. The comparison relies on synthetic, artificially generated data. As discussed, the differences observed in the inference produced by the two methods are largely attributable to the linearization employed for sequential inference for the autoregressive parameter γ in the kDGLM method.

In Figure S.7, we display the inference results under both methods, kDGLM and NUTS, with the parameter γ fixed at its true value. As shown, the two methods yield nearly identical results. Therefore, for models where the evolution matrix $\mathbf{G}$ is known, the kDGLM method is expected to produce inferences closely aligned with those generated by Markov Chain Monte Carlo schemes.

S.6 Multinomial model: obtaining the probabilities of hospital admission for each age group.

As seen in Subsection 5.1, for a multinomial response on k categories where the $k-$th category is used as a reference group, the following predictive structure is specified:

$$\lambda_{jt} = \log\left(\frac{\eta_{jt}}{\eta_{k,,t}}\right) = \boldsymbol{F}'_{jt}\boldsymbol{\theta_t}, \quad j = 1, \ldots, k-1.$$

Let E_{jt} denote the exposure of group j; $E_{1:k,t} = \sum_{j=1}^{k} E_{jt}$ denote the total exposure of all age groups, and consider the following events:

G_{jt} : allocation of an individual to age group j, $j = 1, \ldots, k-1$; at time t, t=1, 2, ...;

H_t: hospital admission of an individual at time t , t=1, 2, ...

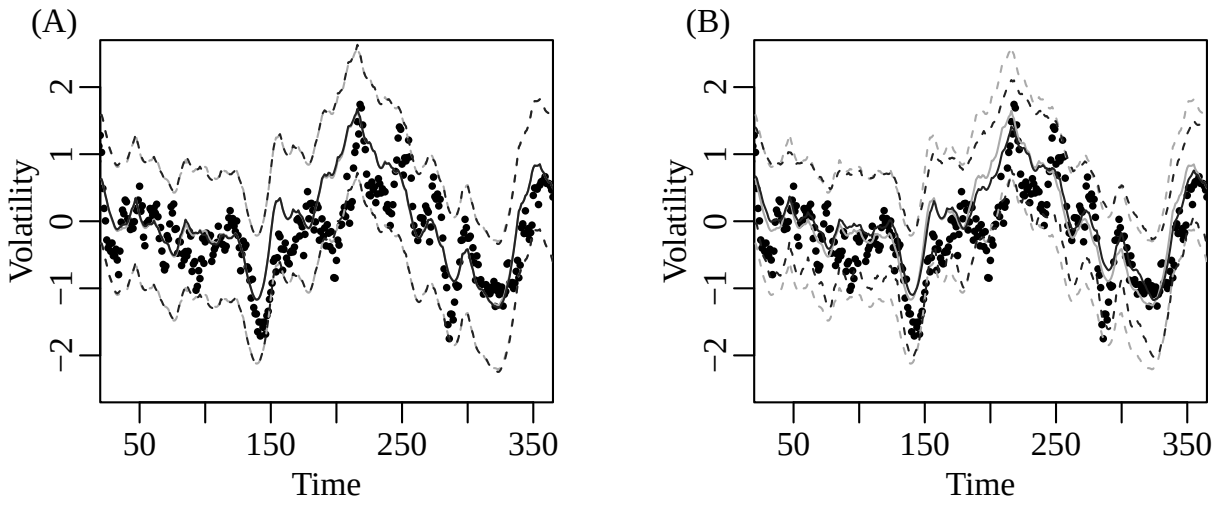

Figure S.7: Comparison of smoothed means (solid lines) and credible intervals (dashed lines) for volatility estimation from synthetic data, with fixed autoregressive parameter γ. The solid circles indicate theoretical volatilities. (A): results of the proposed sequential approach for normal (gray) and gamma (black) formulations. (B): comparison of normal formulation using the sequential approach (gray) and NUTS via Stan (black).

Note that the probability of allocation to age group j and time t, conditionally on having been hospitalized, is:

$$\eta_{jt} = \mathbb{P}(G_{jt}|H_t) = \frac{\mathbb{P}(H_t|G_{jt})\mathbb{P}(G_{jt})}{\mathbb{P}(H_t)} \Rightarrow \frac{\mathbb{P}(H_t|G_{jt})}{\mathbb{P}(H_t|G_{kt})} = \frac{\mathbb{P}(G_{jt}|H_t)}{\mathbb{P}(G_{kt}|H_t)}\frac{E_{kt}}{E_{jt}}.$$

Then it follows that:

$$\begin{aligned}\log\left\{\frac{\mathbb{P}(H_t|G_{jt})}{\mathbb{P}(H_t|G_{kt})}\right\} =& \log\left\{\frac{\mathbb{P}(G_{jt}|H_t)}{\mathbb{P}(G_{kt}|H_t)}\right\} - \log\left\{\frac{E_{jt}}{E_{kt}}\right\} \\ =& \boldsymbol{F}_{jt}'\boldsymbol{\theta}_t - \log\left\{\frac{E_{jt}}{E_{kt}}\right\}, \quad j = 1, ..., k-1.\end{aligned}$$

Thus, estimates of $\mathbb{P}(H_t|G_{jt})$ are naturally obtained if $\log\left\{\frac{E_{jt}}{E_{kt}}\right\}$ is introduced as an offset term in the dynamic predictor of category j, $j = 1, ..., k-1$.

References

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):625–645.

Harvey, A. C. and Fernandes, C. (1989). Time series models for count or qualitative observations. Journal of Business & Economic Statistics, 7(4):407–417.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). Coda: Convergence diagnosis and output analysis for mcmc. R News, 6(1):7–11.

Santos, T. R., Franco, G. C., and Gamerman, D. (2021). NGSSEML: Non-Gaussian state space with exact marginal likelihood. R Journal, 13:111.

Smith, R. and Miller, J. (1986). A non-Gaussian state space model and application to prediction of records. Journal of the Royal Statistical Society. Series B (Methodological), 48(1):79–88.

West, M. and Harrison, P. J. (1997). Bayesian Forecasting and Dynamic Models. Springer, New York, NY.

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.