Received: 28 July 2022 | Revised: 7 January 2023 | Accepted: 22 February 2023
DOI: 10.1002/asmb.2756

PRACTITIONER PAPER

WILEY

A review of Bayesian dynamic forecasting models: Applications in marketing

Helio S. Migon[1] | **Mariane Branco Alves**[1] | **André F. B. Menezes**[2] | **Eduardo G. Pinheiro**[2]

[1]Instituto de Matemática, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

[2]Murabei Data Science, São Paulo, Brazil

Correspondence
Mariane Branco Alves, Instituto de Matemática, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil.
Email: mariane@im.ufrj.br

Funding information
Fundação de Amparo à Pesquisa do Estado de São Paulo, Grant/Award Number: 2018/04654; Universidade do Estado do Rio de Janeiro, Grant/Award Number: E-26/007/10667/2019

Abstract
We briefly review the main developments of Bayesian dynamic models. The emphasis is on marketing applications. Typical examples in this area are discussed. The concepts of monitoring and intervention are carefully explained with illustrative examples and open source computational routines. We avoid algebraic developments and instead use graphical examples to illustrate theoretical aspects. Two real-world problems using Bayesian dynamic models are discussed. Finally, we describe recent developments and alternative proposals to formally address the dependence when dealing with the modeling of multiple time series.

KEYWORDS
missing data, modeling multivariate responses, monitoring and intervention, product development, sales promotion and advertising, what-happened (retrospective) analysis

1 | INTRODUCTION

Although *dynamic models and Bayesian forecasting* were introduced by Jeff Harrison and collaborators more than 50 years ago (Harrison and Stevens[1]), there are still a few applied areas that use this method regularly. Bayesian dynamic models naturally account for time-varying parameters and accommodate real-time monitoring and interventions on time series, establishing a powerful method which promptly react to pattern changes in the observed trajectory of the data. Predicting outcomes for future times, such as sales and returns, in order to plan marketing actions, is a common goal that is also performed in a fairly intuitive way in this class of models, accommodating the uncertainty involved in estimating unobservable quantities. Sequential inference schemes in Bayesian dynamic models allow the learning system to quickly update and correct forecasts in regime change scenarios. As highlighted by Leeflang et al.,[2] dynamic models allow for a single-stage analysis of long-term phenomena, and the Bayesian nature also allows for the inclusion of subjective information.

This article presents a didactic and concise but comprehensive description of the models developed since the end of the last century (partially described in West and Harrison's seminal book[3]) for researchers and professionals in the field of marketing. We highlight the main developments of the method through examples that are illustrated by graphs, avoiding tedious formulas, as much as possible. The emphasis is on aspects of smoothing, monitoring and subjective or automatic interventions, based on decision processes under uncertainty. Many advances have taken place since the models were initially developed in the 1960s, reflecting the practical experiences of the authors in industry. We highlight modeling through multiple discount factors, the use of variance laws, natural treatment of missing data, models for multiple data-generating processes (mixtures of distributions), nonlinear and non-normal models, etc. The basic model includes polynomial trend and seasonality components (including cycles), regression and transfer functions, as well as autoregressive components.

Sales forecasting (Johnston and Harrison[4]), TV advertising and the evolution of its impact on consumer perception (Migon and Harrison[5]), modeling based on microeconometric foundations (Migon[6]), and optimal portfolio development (Migon et al.,[7] Polson and Tew[8]) are some of the marketing applications that have appeared in the literature. Migon et al.,[9] is a review paper worth highlighting. With the proliferation of massive data and thus in order to make analysis computationally efficient and scalable,[10] several methodological advances exploring the ideas of *decouple/recouple* have been recently proposed. Yanchenko et al.,[11] focuses on multivariate revenue forecasting across collections of supermarkets and product categories, adopting hierarchical dynamic models.

The Bayesian literature in marketing also includes some other situations. We highlight the works of Rossi and Allenby[12] and the excellent book by Rossi et al.[13] Econometric time series modeling reviews are presented by Pauwels,[14] who use repeated measures, as well as by Dekimpe and Hanssens,[15] who review marketing time series models. Chandukala et al.[16] discuss the use of sequential hierarchical models to model the relationship between two specific quantities when measured by others, such as in advertising-sales relationship models intermediated by effects of *cognition, affect and experience*. This sort of application relates to dynamic factor models (Lopes[17]).

The remainder of the paper is organized as follows. In Section 2, we discuss marketing motivation, including dynamic decouple/recouple models, new product releases, advertising effects, and marketing structure. Section 3 provides an introduction to Bayesian Forecasting analysis with focus on Dynamic Linear Model and its practical aspects applied in sales time series data. Section 4 briefly describes non-normal dynamic models. In Section 5, two challenges developed by Murabei Data Science[1] are described, illustrating the usefulness of dynamic linear models to leverage decision making in real-world problems. Recent research advances of dynamic models are addressed in Section 6 with focus on multivariate models. The paper closes with some concluding remarks in Section 7.

2 | MOTIVATION

This section describes some marketing-related applications. In general, these applications include high-dimensional data sets requiring the use of efficiently inferential methods. Models of dynamic factors, aggregation of seasonal structure, release of new products, advertising effect, and the hierarchical structure of a market (top-down models) are some of the examples addressed as motivation.

Bayesian modeling's sequential nature combines harmoniously with chronologically available observations. The inferences about the involved probability distributions are continuously updated, and forecasts are naturally made, enabling regular monitoring of forecasts, and as a result the quality of the developed models can be improved. Furthermore, the predictive distribution of observable quantities serves as a practical instrument for evaluating the quality of adjusted models, since it allows observed values for these quantities to be compared to predictions, assisting the decision-making process under uncertainty.

2.1 | Modeling multivariate responses: Decoupling/recoupling

Dekimpe and Hanssens[15] highlight that data sets from marketing applications have been increasing into different directions: from various measured variables, such as performance metrics, prices, sales, and advertising to different temporal aggregation (hours, week, days), across geographical places (countries or even continents) and over long time spans. Dealing with high-dimensional time series data requires the development of efficient models and computational procedures that are scalable. However, commonly used models in marketing, as discussed by Leeflang et al.[2]—such as vector-autoregressive (VAR) and the state-space models when using intense simulation (for instance Markov Chain Monte Carlo methods[18-20]) - are not scalable and lose the sequential aspect of inference.

The concept of decoupling and recoupling, as seen in West,[21] is one major advance that can improve the computational efficiency in marketing applications, decomposing high-dimensional multivariate analyses into smaller blocks which are processed in low computational time and can be recoupled by means of some factor aiming at recovering dependence structures among blocks. In this sense, the dynamic latent factor modeling proposed by Lavine et al.[22]

[1] https://www.murabei.com/.

explores the sequential inference and computational speed of univariate dynamic models that are linked via dynamic factor processes that share the information across series, preserving scalability as it allows partial parallel inference over the time series.

The modeling of daily sales of various products at a supermarket is an emblematic example considered by Berry and West.[10] There are a large number of items, multiple time periods, and forecasts are needed for several periods in the future. It is reasonable to assume that these patterns are due to a common factor, for example customer flow to stores, but this is often estimated with great inaccuracy due to noise that is present in the most disaggregated levels of the data. To overcome these difficulties, the authors propose the use of a common seasonal factor estimated from a convenient aggregation of the data and then incorporating this factor to the models of each item. Alternatively, the multiparametric/multivariate exponential family models proposed by Alves et al.[23] can be applied, dealing with counts on multiple categories in a dynamic multinomial framework which is efficiently updated, preserving sequential processing.

Nonnegative counts time series can arise in marketing settings from underlying complex phenomena, for instance, forecast product demand at different levels of time or space aggregation is crucial for planning marketing decision. This type of data usually presents general challenges such as over-dispersion and zero-inflation that are ignored in commonly used marketing models. The class of Dynamic Count Mixture Models (DCMMs) proposed by Berry and West[10] can be applied to deal with these challenges. Furthermore, this class of model can also be applied for modeling of multivariate count time series data throughout dynamic factor and copula approaches proposed by Lavine et al.[22]

2.2 | Product development

The four stages that are often featured in the product development process are opportunity identification, product design, sales forecasting, and commercialization (see chapter 3 of Fan et al.[24]). One could be interested in modeling both the quantity of new ideas generated by someone as well as whether or not a new concept is implemented in the context of opportunity identification. Both situations offer potential applications for dynamic generalized models. The works of Bass[25] and Schmittlein and Mahajan[26] describe stochastic modeling for the problem of diffusion/adoption of new ideas or products within social systems. Individuals of a certain society are classified as innovators or followers. Innovators exercise their “shopping” independently of others in the social system, while “followers” are influenced by the former and other factors inherent in the social system itself. After some algebraic efforts, Bass et al.[27] obtained a logistic-type growth model, incorporating some strategic variables of marketing, namely effect of price variation, advertising, promotions and so forth.

The application of this class of models goes beyond marketing problems and can describe the evolution of an epidemic, retail services, industrial technologies, consumption of durable goods, and so forth. Bayesian estimation of this class of models can be seen in Ramírez-Hassan and Montoya-Blandón.[28] The logistic model is slightly modified by Bass et al.,[27] including marketing components as the price and advertising effects, the coefficients of innovation and imitation, and the involved population. The dynamic version of Bass's generalized models (Bass et al.,[27]) can be easily handled as a member of the wide class of dynamic growth curve models, that also includes the logistic, Gompertz, and modified exponential models (Migon and Gamerman[29]).

2.3 | Sales promotions and advertising effect

Sales promotions take up a big portion of the companies' marketing budgets. Promotions have long-term effects on brands and categories, in addition to the positive and immediate effects they frequently have on sales (Blattberg et al.[30] and Chib et al.[31]). Modeling these dynamic effects frequently uses Almon's lag models as a foundation (Seetharaman[32]). The class of Bayesian transfer functions (see Section 3.2), which can parsimoniously explain these impacts, is an efficient alternative for this purpose. For instance, it is well known how promotions can affect brand preference, category incidence, and purchase volume. These effects can be modeled using dynamic models with responses in the exponential family, such as the multinomial and Poisson models.

In the 1980s, a dynamic model for memorization of advertising on TV was developed by Migon and Harrison.[5] The memorization of an ad, broadcast on a certain television channel, was evaluated by interviewing consumers at the door of a supermarket. A card was presented with competing brands and consumers were asked: “which of these brands' TV ads do you remember seeing last week?” The ability of an ad to generate memory was simply the proportion of consumers who memorized the brand's ad (y_t). These responses were related to a standardized audience index (x_t) of the TV program on which the ad was aired (*TVR - television rate*). This example contains a lot of interesting features. First, the introduction of the memorization measure avoids the discussion of the direction of causality between sales and advertising. The method of inference should be sequential, consider missing observations (because for a few weeks the survey was not performed due to weather issues) and external information about changes in the advertising "message'. It should also model proportion data. The relationship between the proportion of those who remembered the advertisement and TVR, audience rates where the message was conveyed, is nonlinear requiring proper approximations.

Another interesting application related to advertising was developed by Fernandez et al.,[33] where a new media optimization system based on the Markovitz mean-variance model was presented. These models are relevant to media planning because they optimally choose resource allocations in different media, that is, how many inserts for each distinct medium and when to allocate. The inserts can be the advertising space in a magazine or newspaper, or ad time on a particular radio or TV program. An audience index is constructed from a sample of individuals from the target audience of the advertising campaign under review. In this way, the assets are the inserts and the returns are the ratings. Therefore, the goal is to allocate the available resources in a number of inserts in order to maximize the expected audience index, while simultaneously minimizing the associated risk. The quantities involved in this optimization problem are time series of past experiences. The idea is to make media allocations that are efficient for future allocations. Therefore, it is necessary to predict the audience index within the established planning horizon.

2.4 | Modeling of the market structure

The modeling of the market structure is relevant at least at two moments: when starting a new business or when modifying the strategy in an existing market. Several references cited in Terui et al.[34] describe alternative static models to address this problem of structured or hierarchical modeling in marketing. The basis proposal, however, involves the dynamic modeling of the involved components.

Suppose one has count data describing sales of m distinct brands. The total market is also observed over time, characterizing a time series consisting of the numbers of the various products sales. An example with three levels (total market, submarkets and brands within each submarket) is illustrated in Figure 1.

Modeling is based on well-known properties of probability theory. For each brand, a count time series is modeled by a Poisson distribution with parameters varying in time. Assuming conditional independence between the products and using the fact that the sum of independent Poisson distributions is also a Poisson, one obtains a structure of competition between the brands (*statistical dependence*), transforming the distribution of each brand into its conditional distribution given the total market (sum over brands). It is known from probability theory that the joint distribution of trademarks, conditional on total, will be a multinomial distribution. Thus, one recovers the competition between brands.

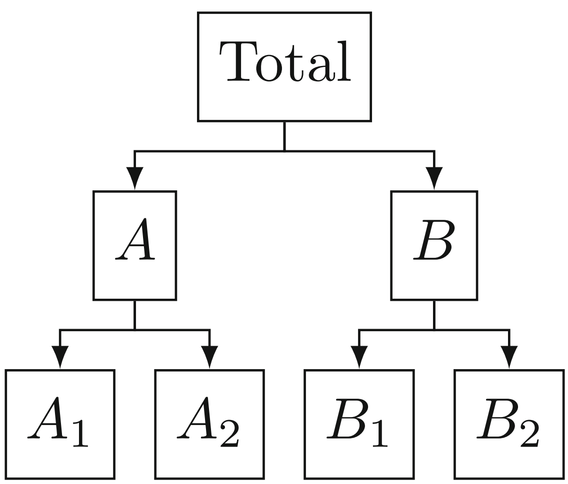

FIGURE 1 Hierarchical structure composed by the total market, submarkets, A and B, and brands ($A_i, B_i,\ i = 1, 2$).

What are the lessons of this simple example? First, that implicitly one is assuming that competition between brands can vary along the planning horizon, which is obviously not captured by traditional statistical models. On the other hand, this is yet another lesson in how one can achieve computational efficiency with massive data. Each univariate brand's sales is modeled to obtain its joint distribution or the *competition between brands*.

Recent developments by Alves et al.[23] for dynamic models in the k-parametric exponential family allow the sequential modeling of multinomial and compositional data with appropriate approximations and reduced computational time, which allows, in addition to the aspects already mentioned, testing different structures for the model, in a feasible time frame.

3 | INTRODUCTION TO DYNAMIC MODELS

We begin this section by reviewing the fundamentals in Bayesian inference and the paradigm adopted for updating information in dynamic linear models (DLM). The stages of the inference process in such models are described with emphasis on the sequential updating, which allows the execution of monitoring and intervention in the system in a timely manner, when strategically necessary. Transformations in the data are discussed to accommodate them to possibly restrictive hypotheses of the DLM class. We also discuss modifications to the models' variance law in order to accommodate the data with its original scale, which facilitates communication with decision-makers. The main concern of this paper is to present sophisticated strategies like intervention and monitoring that help marketing analysts make better decisions.

All the illustrations shown in the remainder of the paper were performed using the open source packages in R and Python, namely `RBATS`,[35] `PyBATS`,[36] and `PyBATS-detection`.[37] The codes are available through the Github repository at https://github.com/Murabei-OpenSource-Codes/dynamic_models_marketing.

3.1 | Basic concepts of Bayesian inference

Let y be a random variable whose generating process is described by a probability distribution. Its density (continuous case) or probability (discrete case) function is denoted generically by $p(y|\boldsymbol{\theta})$, $y \in \mathcal{Y}$, and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where y is observable and $\boldsymbol{\theta}$ is not observable (parameter). Following the observation of y, the likelihood function of $\boldsymbol{\theta}$ is calculated as follows: $\ell(\boldsymbol{\theta}|y) = p(y|\boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta}$, whose analytic expression is the probabilistic model, but the function argument will now be the unknown quantity $\boldsymbol{\theta}$ with the observable y provided. This function describes the relevance of each value of $\boldsymbol{\theta}$, based on the observed value of y. Note that the likelihood function of a sequence of independently and identically distributed observations, a random sample, will be the product of the marginal distributions.

In Bayesian inference, an initial distribution over the parameter space, called a prior distribution, is introduced. This distribution probabilistically describes what is known about the parameter initially and is subjectively evaluated. Its probability density is denoted by $p(\boldsymbol{\theta}|\psi), \boldsymbol{\theta} \in \boldsymbol{\Theta}$, where ψ denotes hyperparameters. Bayes' theorem allows obtaining a posterior distribution:

$$p(\boldsymbol{\theta}|y, \psi) \propto \ell(\boldsymbol{\theta}|y)p(\boldsymbol{\theta}|\psi). \tag{1}$$

A numerical summary of posterior distributions is desirable. Based on fundamentals of decision theory, we can choose the average, the mode or the median of the posterior distribution as a point estimate, depending on the loss function chosen by the decision maker. We can use the variance of the posterior distribution or the curvature around the mode as a measure of uncertainty . Finally we can build intervals of highest posterior density. These concepts can be seen in detail in Migon et al.[38]

The Bayesian paradigm naturally produces predictive distributions, that is, the distribution of future observations based on observed data. We need, again, to solve an integral, possibly of high dimensionality:

$$p(y_f|y) = \int p(y_f|\boldsymbol{\theta})\ p(\boldsymbol{\theta}|y)\ \mathrm{d}\boldsymbol{\theta} = E_{\theta|y}\left[p(y_f|\boldsymbol{\theta})\right], \tag{2}$$

where y denotes the given observed data and y_f denotes future data. Predictive distributions play a crucial role in assessing the quality of a model: since they are distributions of future observable quantities, once these are available it is possible to compare the predictions made to the actual observed values. In the context of time series, predictive distribution is not infrequently one of the main objectives of the analysis. Examples of Bayesian updating based on normal and Poisson models are presented in Appendix A and could be beneficial for marketing researchers unfamiliar with Bayesian modeling.

3.2 | Dynamic linear models

A broad class of time series models was introduced by Jeff Harrison in works published in the 1970s.[1,39] The models are formulated in state space and probabilistically describe both the observations and the unobservable parameters, or states. Inference is based on Bayes' theorem and is processed sequentially as each new observation becomes available. The class includes several facilities for dealing with applied problems, especially in marketing, and includes the author's actual expertise in the *Imperial Chemical Industry* (ICI). These models go far beyond the equally relevant and well-known Kalman filter, which was created in the field of electrical engineering in the 1960s.

The class of parametric dynamic linear models (DLM) is a natural extension of regression models in which all the parameters are assumed to vary with time. The following are the observational and parameter evolution (system) equations:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{F}_t' \,\boldsymbol{\theta}_t + \epsilon_t, & \epsilon_t &\sim N_m[0, \mathbf{V}_t], \\ \boldsymbol{\theta}_t &= \mathbf{G}_t \,\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, & \boldsymbol{\omega}_t &\sim N_p[0, \mathbf{W}_t], \end{aligned} \tag{3}$$

where $\mathbf{y}_t$ is an $m-$dimensional time series of observations that are conditionally independent given $\boldsymbol{\theta}_t$ ($p \times 1$ vector) and normally distributed with covariance matrix $\mathbf{V}_t$. A Markovian rule guides the evolution of the states $\boldsymbol{\theta}_t$, and conditional on $\mathbf{G}_t$ and $\boldsymbol{\theta}_{t-1}$ they are normally distributed with a covariance matrix $\mathbf{W}_t$. In summary, the above-mentioned class of models is defined by the quadruple $\{\mathbf{F}, \mathbf{G}, \mathbf{V}, \mathbf{W}\}_t$, where $\mathbf{F}_t$ is a $p \times m$ matrix of regressors and $\mathbf{G}_t$ is a $p \times p$ matrix describing state evolution. For univariate time series ($m = 1$), there are two broad model classes to consider: time series models, where $\mathbf{F}_t = \mathbf{F}$ and $\mathbf{G}_t = \mathbf{G}, \forall t$, and dynamic regression models, when $\mathbf{F}_t' = (X_1, \ldots, X_p)_t'$, $\mathbf{G}_t = \mathbf{I}_p$, where $\mathbf{I}_p$ denotes the identity matrix of order p. Let $D_t = \{D_0, y_1, \cdots, y_t, I_t\}$ the available information up to time t, where I_t denotes any subjective information external to the data. If no extra information is incorporated into the observed time series at each time t, that is, $D_t = \{D_{t-1}, y_t\}$, the model is said to be closed to external information and static if $\mathbf{W}_t = 0$.

In several temporal contexts, the main focus are h steps ahead predictions, given information D_t up to time t. For each t and $h \geq 0$, the forecast distribution is

$$\mathbf{y}_{t+h} \sim N_m[\mathbf{f}_t(h), \mathbf{Q}_t(h)], \tag{4}$$

with $\mathbf{f}_t(h) = E[\mathbf{y}_{t+h}|D_t] = \mathbf{F}_t'\mathbf{a}_t(h)$; $\mathbf{Q}_t(h) = V[\mathbf{y}_{t+h}|D_t] = \mathbf{F}_t'\mathbf{R}_t(h)\mathbf{F}_t + \mathbf{V}_{t+h}$, where $\mathbf{a}_t(h) = \mathbf{G}_{t+h}\mathbf{a}_t(h-1)$; $\mathbf{R}_t(h) = \mathbf{G}_{t+h}\mathbf{R}_t(h-1)\mathbf{G}_{t+h}' + \mathbf{W}_{t+h}$; with initial values: $\mathbf{a}_t(0) = \mathbf{m}_t = E[\boldsymbol{\theta}_t|D_t]$; $\mathbf{R}_t(0) = \mathbf{C}_t = V[\boldsymbol{\theta}_t|D_t]$.

In some applications, as seen in Subsection 5.1, interest lies in the distribution of $(\mathbf{X}_t(h)|D_t)$, called h-steps lead-time distribution, where $\mathbf{X}_t(h) = \mathbf{Y}_{t+1} + \mathbf{Y}_{t+2} + \ldots + \mathbf{Y}_{t+h}$. The lead-time distributions for any time t and $h > 0$ are naturally obtained from the forecasting h-steps ahead distributions at time t, as seen in West and Harrison[3(pp 38-39, 138,139)]. Following that, we will show some univariate examples of how the model's components are defined ($m = 1$).

(a) First-order polynomial model

This model corresponds to the specification $\{1, 1, V, W\}_t$. As usual, $y_t \sim N[\mu_t, V_t]$ and

$$\mu_t = \mu_{t-1} + \omega_t, \quad \omega_t \sim N[0, W_t]. \tag{5}$$

Some relevant remarks are: (i) this is the simplest DLM and corresponds to a time-varying mean model; (ii) to implement it, we need to know the values of V_t and W_t, which are not known in practical applications (these topics will be addressed ahead); (iii) the point prediction function $f_t(h) = E[y_{t+h}|D_t]$ of a first-order polynomial model is constant for whatever $h = 1, 2, \ldots$, resulting in forecasts that do not appropriately account for trend and seasonal patterns, for example. As a consequence, it is a good model for making very short-term forecasts. The estimation of $V_t = V$ will be detailed in Section 3.5, as well as the specification of W_t using discount factors.

(b) Second-order polynomial model (linear growth)

By inserting a linear growth factor in the first-order polynomial model, a slightly more sophisticated second-order model is generated. Then $\mathbf{F} = (1, 0)'$ and $\mathbf{G} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ will be obtained. This model has an increasing or decreasing linear

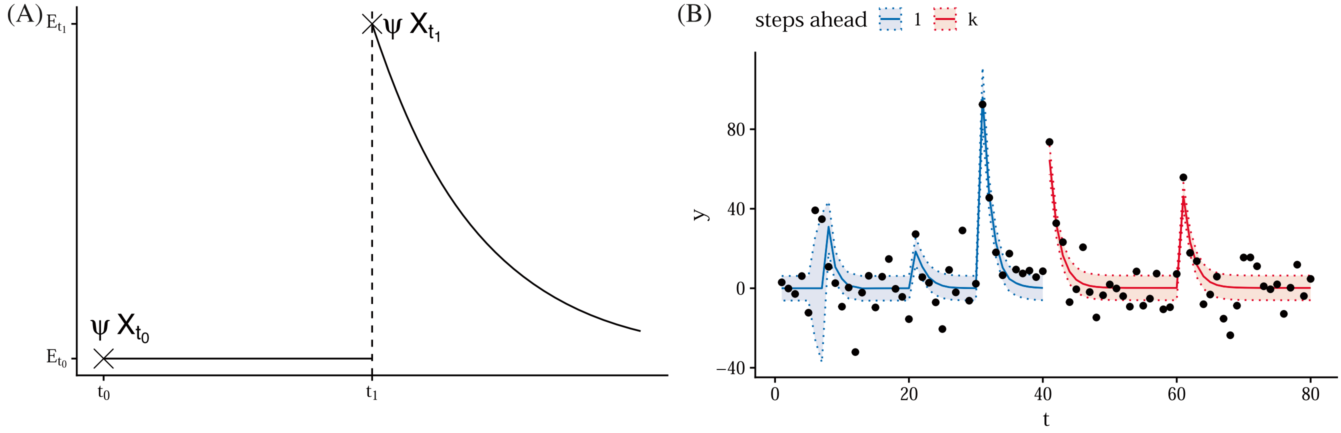

FIGURE 2 (A) First-order transfer function with instantaneous stochastic gain factor $\psi \sim N[\mu_\psi, \sigma^2_\psi]$ and decay factor $0 < \lambda < 1$. (B) 1(k)-step ahead forecast in blue (red) for DLM with first-order transfer function in a simulated example.

prediction function, $f_t(h) = E[y_{t+h}|D_t]$, $h = 1, 2, \ldots$, which is suitable for medium-term prediction of time series with trend patterns. In general, a kth order polynomial model generates a polynomial prediction function of order $k - 1$, so to capture linear trends it suffices to specify a second-order polynomial model.

(c) Seasonal model

Seasonal models can be specified in terms of seasonal factors. For a seasonal pattern of period p, $\mathbf{F} = E_p = (1, 0, \ldots, 0)'$, with evolution dictated by the permutation matrix $\mathbf{G} = \mathbf{P} = \begin{pmatrix} \mathbf{0} & \mathbf{I}_{p-1} \\ 1 & \mathbf{0}' \end{pmatrix}$. Alternatively, one can parsimoniously specify seasonal patterns, using harmonics. The kth seasonal harmonic component is given by $F = E_2 = (1, 0)'$, $\mathbf{G}(k, \omega) = \begin{pmatrix} \cos(k\omega) & \sin(k\omega) \\ -\sin(k\omega) & \cos(k\omega) \end{pmatrix}$, $\omega = 2\pi/p$, $k = 1, 2, \ldots, h$, $h = p/2$, if p is even, and $h = (p-1)/2$ if p is odd.

(d) Dynamic regression model

Suppose the pair of values (y_t, x_t), $t = 1, \cdots, T$ is observed. $\{\mathbf{F}, I_2, V, \mathbf{W}\}_t$ is the dynamic regression model, with $\mathbf{F}'_t = (1, x_t)$. The regressor x_t could be the price of a product, the existence or absence of a promotion, a holiday indicator and so on. The dynamic regression model approximates a true nonlinear relationship between x_t and y_t at a local level. This approach explicitly acknowledges that the impact of a regressor on the response y_t might change over time.

(e) Transfer function models

These are parsimonious representations of a regression with lagged regressors. Let x_t be an observable input variable, and E_t denote the state component describing its effect. When the effect of the regressor on the response is not only instantaneous, but also spread over time, as detailed in the works by Alves et al.[40] and Ravines et al.,[41] this type of structure is effective. Let ψ denote an instantaneous effect per unit variation of x_t, and $\lambda \in (0, 1)$ denote an exponential decay parameter. A very basic first-order transfer function example is given by $E_t = \lambda E_{t-1} + \psi x_t$. Consider the observational model $y_t = \mu_t + E_t + \epsilon_t$, where $\epsilon_t \sim N[0, V_t]$.

An example of first-order dynamics is the relationship between advertising investments and their dynamic effect on recall (y_t). Note that if λ and ψ are not known, the model will be nonlinear. For example, $\psi \sim N[\mu_\psi, \sigma^2_\psi]$ represents an instantaneous stochastic gain. We can increase the vector of states to sequentially estimate the quantities (λ, ψ) (Figure 2).

(f) Superposition of structural blocks

In practice, it is advisable to develop a component-to-component model through the superposition principle, to take into account, in different structural blocks, the perceptible components of trend, seasonality, regression and so forth. The ith model is denoted by $\{\mathbf{F}_i, \mathbf{G}_i, 0, \mathbf{W}_i\}$, $i = 1, \ldots, r$ and by $\{0, 0, V_i, 0\}$ a model for observational noise. Define $\mathbf{F}_t = (\mathbf{F}'_{1t}, \ldots, \mathbf{F}'_{rt})'$, $\mathbf{G} = \text{diag}(\mathbf{G}_1, \ldots, \mathbf{G}_r)$ and $\mathbf{W}_t = \text{diag}(\mathbf{W}_{1t}, \cdots, \mathbf{W}_{rt})$. Consider the monthly observed data, where $r = 3$.

One possible structure for the DLM is composed of a linear growth block (b), $\mathbf{F}_{1t} = \mathbf{F}_1 = (1, 0)'$ and $\mathbf{G}_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$; a seasonal component (c) with annual and half-yearly period specified in terms of harmonics and given by $\mathbf{F}_{2t} = \mathbf{F}_2 = (1, 0, 1, 0)'$ and $\mathbf{G}_2 = \begin{pmatrix} \mathbf{G}(1, \omega) & \mathbf{0} \\ \mathbf{0} & \mathbf{G}(2, \omega) \end{pmatrix}$, $\omega = 2\pi/12$; and a dynamic regression component (d) $F_{3t} = x_t$, $G_3 = 1$.

3.3 Inference in DLMs

As stated above, the set of available information up to time t, in general, equispaced observations, is denoted by $D_t = \{D_0, y_1, \ldots, y_t\}$. The main inferential aspect in the DLMs is to obtain the prior, predictive, and posterior distribution, which are defined, respectively, by:

$$p(\boldsymbol{\theta}_t|D_{t-1}) = \int p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, D_{t-1})\; p(\boldsymbol{\theta}_{t-1}|D_{t-1})\; \mathrm{d}\boldsymbol{\theta}_{t-1}, \tag{6}$$

$$p(y_t|D_{t-1}) = \int p(y_t|\boldsymbol{\theta}_t, D_{t-1})\; p(\boldsymbol{\theta}_t|D_{t-1})\; \mathrm{d}\boldsymbol{\theta}_t, \tag{7}$$

$$p(\boldsymbol{\theta}_t|D_t) \propto p(\boldsymbol{\theta}_t|D_{t-1})\; p(y_t|\boldsymbol{\theta}_t, D_{t-1}), \tag{8}$$

where, technically, some hypotheses of conditional independence are used.

These expressions can be found in chapter 4 of West and Harrison,[3] where the normal posterior (mean and covariance) parameters at time $t-1$ are denoted by $\mathbf{m}_{t-1}, \mathbf{C}_{t-1}$, the prior parameters for time t, $\mathbf{a}_t$, $\mathbf{R}_t$, and the mean and variance for the predictive distribution for time t, f_t, Q_t, respectively. For normal models with known variances, Figure 3 illustrates the evolution and updating of the distributions of the random quantities involved, as well as the parameters defining it.

3.4 Some examples of inference in DLM

(i) The components of a time series

Our first illustration is based on the well-known monthly number of passengers carried in the United Kingdom from 1949 to 1960, as seen in Figure 4A. This time series presents multiplicative seasonality and linear growth trend. As an approximation for a multiplicative model, a dynamic linear growth model (Example (c)) with the superposition of two harmonic components describing the seasonal effects (see West and Harrison[3(p. 97)]) was sequentially adjusted to the data. As will be seen later, “default” discount factors describe the evolution of the states: 0.95 for the trend block and 0.98 for the seasonal block, respectively. All the analyses in this article use vague normal priors with null mean and large variance. We

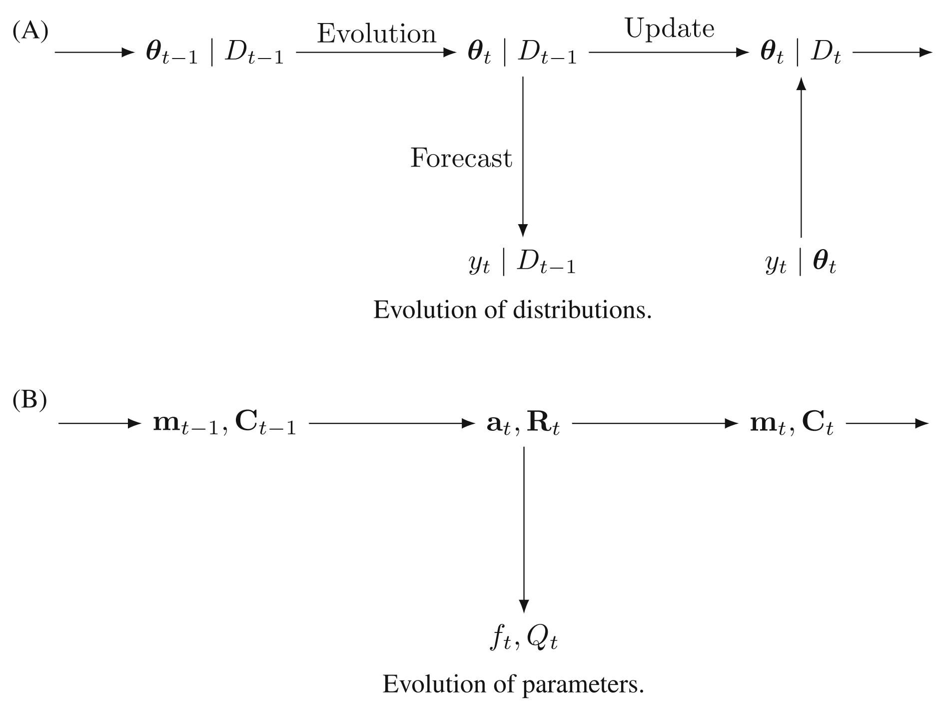

FIGURE 3 Evolution and updating of the distributions involved (A) and notation of the parameters that characterize these distributions (B).

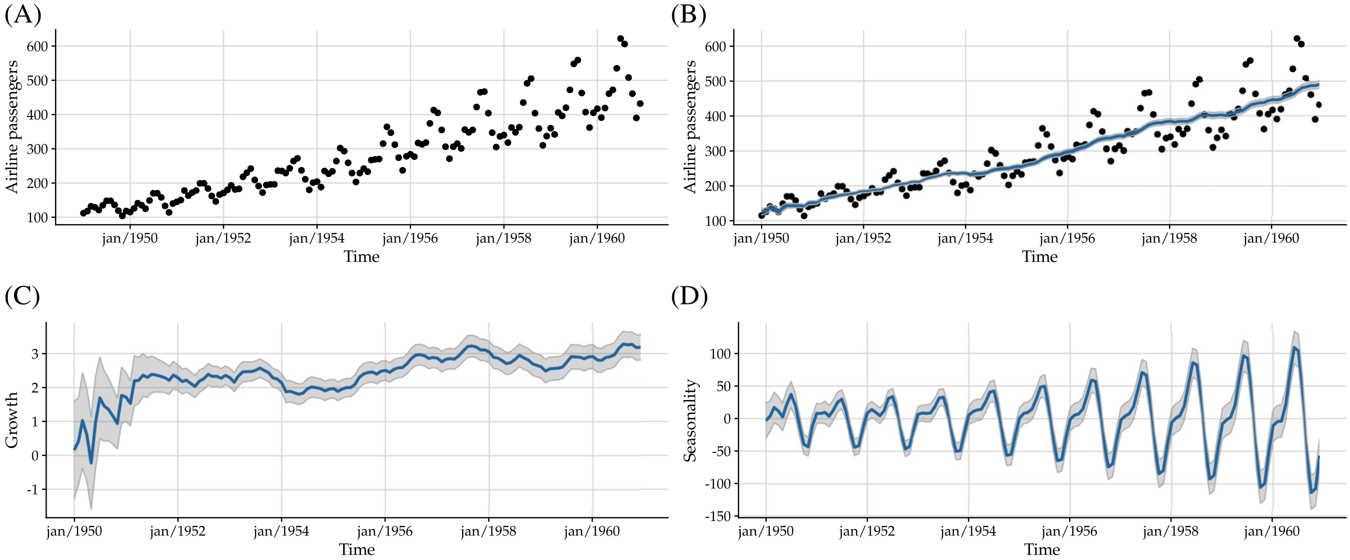

FIGURE 4 (A) Observed time series and $E[\boldsymbol{\theta}_t|D_t]$, filtered expected values of the model components . (B) Current level. (C) Growth factor. (D) Seasonal component.

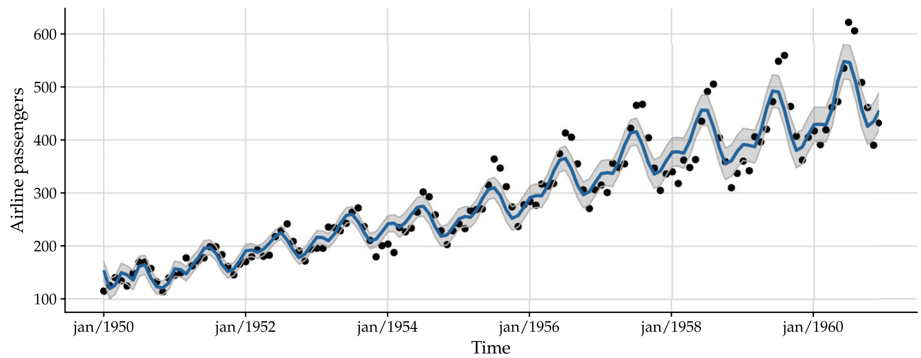

FIGURE 5 Expected value of forecasts one step ahead ($E[y_t|D_{t-1}]$) with 95 % credibility intervals.

use a conjugate inverse-gamma prior for the observational variance, which is assumed to be unknown, but constant. The conjugation beta-gamma (random walk approach) is used when the observational variance is time-varying. In addition, the first $2\,p$ observations are omitted in all of the following graphs, where p represents the dimension of the state vector.

The expected value of the posterior distributions of the filtered states ($E[\theta_t|D_t]$) and the corresponding 95% credibility intervals are shown in Figure 4: the observed time series in (A); the current level in (B); the growth factor in (C); and finally, seasonality in (D). This DLM is a good “approximation” of the observed data, although its generating process is clearly nonlinear.

Figure 5 exhibits the expected value of the predictive distribution $E[y_t|D_{t-1}]$ and the corresponding 95% credibility interval.

(ii) Sequential nature of Bayesian inference

A relevant aspect relates to obtaining the mean of the predictive distribution for a certain h horizon, called the $h-$horizon forecast function. This function is useful for defining the components of the dynamic model. For time series models closed to external information, this function reduces to $f_t(h) = E\,[y_{t+h}|D_t] = E[\mathbf{F}\mathbf{G}^h\ \theta_t|D_t] = \mathbf{F}\ \mathbf{G}^h\ \mathbf{m}_t$, $h = 1, \ldots, H$.

Figure 6 shows 3 year forecasts beginning in December 1955 and one year ahead starting in December 1957, respectively. When including data from 1956 and 1957, the forecasts for the next year become substantially more accurate. This example demonstrates the method’s sequential nature as well as the seasonal additive model’s ability to approximate the current situation, where seasonality is clearly multiplicative. After receiving the additional observations, it was possible to significantly improve the forecasts for the following year. This is primarily due to the method’s sequential structure.

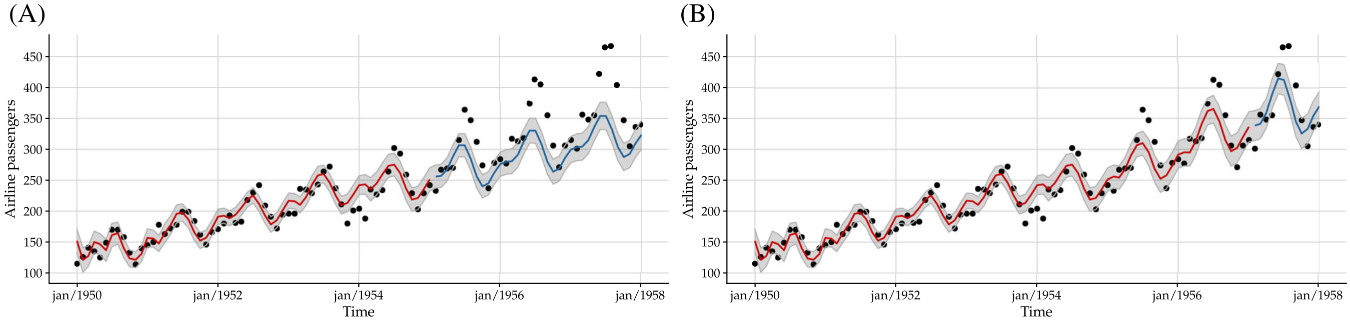

FIGURE 6 Example of the sequential nature of dynamic Bayesian models. (A) Forecasts after December 1955. (B) Forecasts after December 1957 ($E\left[y_{t+h}|D_t\right], h = 1, \dots, H$). Red line: forecast 1 month ahead. Blue line: forecast 36 and 12 months ahead, respectively.

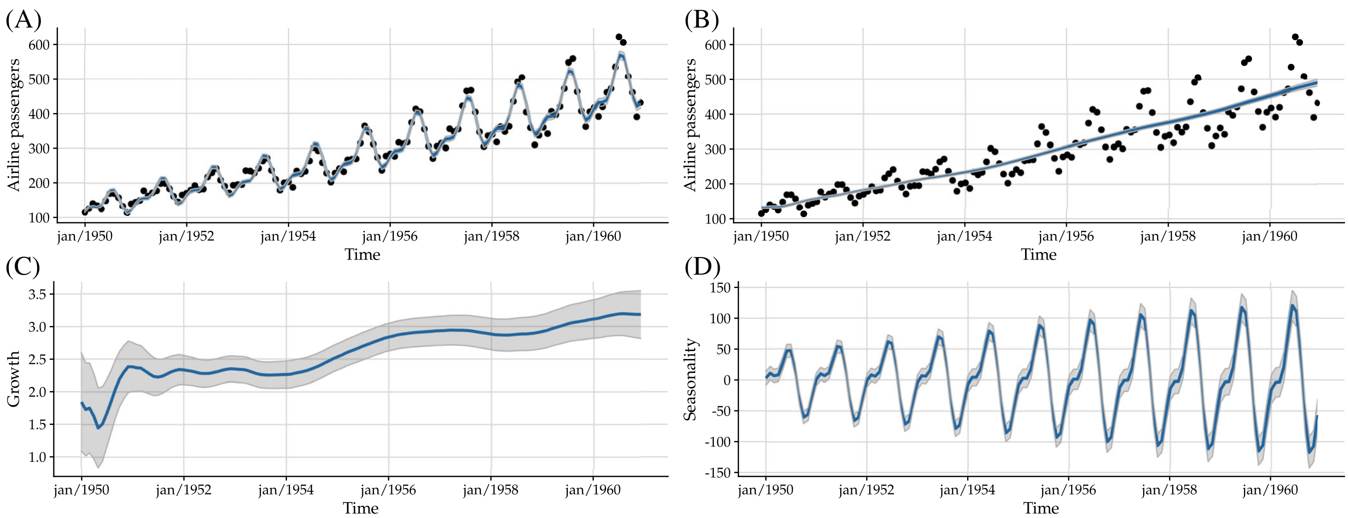

FIGURE 7 Smoothed expected values of model components, $E\left[\boldsymbol{\theta}_t|D_T\right]$, with 95% credibility intervals. (A) Predictive distribution. (B) Current level. (C) Growth factor. (D) Seasonal component.

(iii) Retrospective analysis or smoothing

The structure of the upgrading process makes it natural to obtain smoothed (or retrospective) distributions, denoted by $p(\boldsymbol{\theta}_t|D_{t+k}),\ k > 0$, in addition to the ones listed before.

The smoothed distribution, which corresponds to reviewing all past states based on all the observations may be of particular interest (see details in West and Harrison[3(p. 116)]). For example, an advertising campaign launched at time t_0 can result in effects that dynamically fade over time. It will be beneficial to plan future campaigns if these effects can be more precisely established.

Figure 7 illustrates the smoothed expected values ($E[\boldsymbol{\theta}_t|D_T]$) and the corresponding 95% credibility intervals for the previous example of the time series of passengers carried in the UK.

As can be seen, retrospective analysis has no predictive value because the nature of the prediction process follows the chronological order. Nevertheless, it is useful for revising latent components of the model in light of all information available up to time T, giving the analyst a new perspective on the structure under investigation and possibly deepening the analyst's understanding of the process.

3.5 | Practical aspects of Bayesian forecasting models

(i) Normal model with transformed data

Before applying the normal dynamic models, the Box-Cox family (1964) can be used to transform the data in order to stabilize the variance of the observations. Its simplest form corresponds to the power transformation y_t^λ if $\lambda \neq 0$

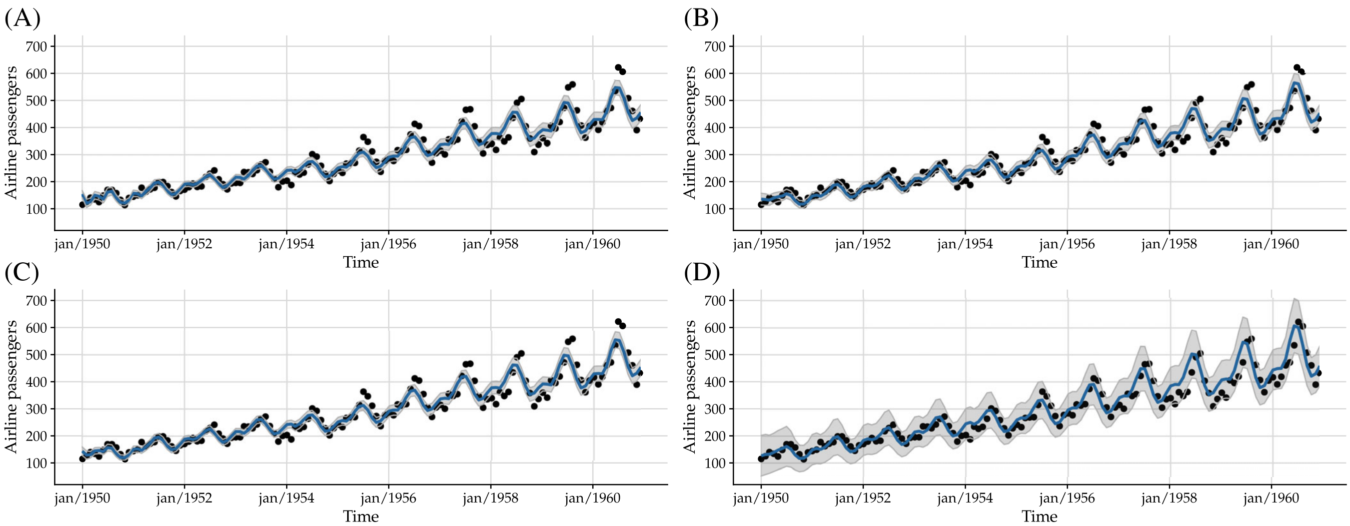

FIGURE 8 One-step-ahead predictions with 95% credibility interval, with different data transformations: (A) original scale; (B) square root; (C) power 3/4 and (D) logarithm. The MAPEs are 0.083, 0.041, 0.061, and 0.016, respectively.

TABLE 1 Summary statistics of the predictive distribution for different values of λ.

λ	ℓ^*	MAPE	MSPE
1	−767.049	0.082	0.009
1/2	−693.857	0.041	0.002
3/4	−740.243	0.061	0.005
log	**−656.838**	**0.015**	**<0.001**

and $\log(y_t)$ if $\lambda \to 0$. The transformed series' likelihood function will be equal to the original series' likelihood multiplied by the transformation Jacobian, $J = \lambda^n \prod_{i=1}^{n} y_i^{\lambda-1}$, which is critical for choosing the best transformation (see Pole et al.[42(pp 95–100)]) .

We once more use the series of passengers transported in the United Kingdom to demonstrate the effect of time series transformations. Thus, using Box-Cox transformations with $\lambda = 1/2$ and $3/4$, as well as logarithmic transformation, the previously specified DLM was adjusted to the transformed series. The values of the predictive log-likelihood adjusted by the Jacobian (ℓ^*), the mean absolute percent error (MAPE), and the mean square percent error (MSPE) for the different models are reported in Table 1. The logarithmic transformation resulted in the lowest MAPE and MSPE, and the highest predictive log-likelihood values. The effect of each of these transformations on predictions can be seen in Figure 8, which shows one-step-ahead forecasts for the four models, where second-order approximations were employed to restore the original scale. Because the logarithmic transformation transforms the model into an additive one, it is the most appropriate choice based on all the adopted criteria.

(ii) Variance laws

As seen above, it is customary in statistical applications to seek transformations of the data in order to stabilize variance and/or induce symmetry in the distribution. In such cases, inferences, including forecasts, should be communicated to the decision maker on the transformed scale, which can be inconvenient. An alternative, avoiding data transformation, is to use a variance law, $V_t = V\ k(\mu_t)$, where $k(\mu_t) = \mu_t^{\alpha_1}\ (1-\mu_t)^{\alpha_2}$, and $\mu_t = \mathbf{F}_t'\ \theta_t$ represents the mean of the data generating process.

Two special cases are: (i) binomial variance law if $\alpha_1 = \alpha_2 = 1$, and (ii) Poisson variance law if $\alpha_1 = 1,\ \alpha_2 = 0$. An interpretation of the Poisson variance law in a business environment would be the following: the number of orders of a product follows a Poisson distribution, but the number of items varies by order so that the number of items per order follows a composite Poisson process.

The constants α_1, α_2 can be chosen to correspond to the Box-Cox family of transformations. For instance, $\alpha_1 = -2$, $\alpha_2 = 0$ corresponds the logarithmic transformation often used in economic data modeling. Theorem 4.3 in West and

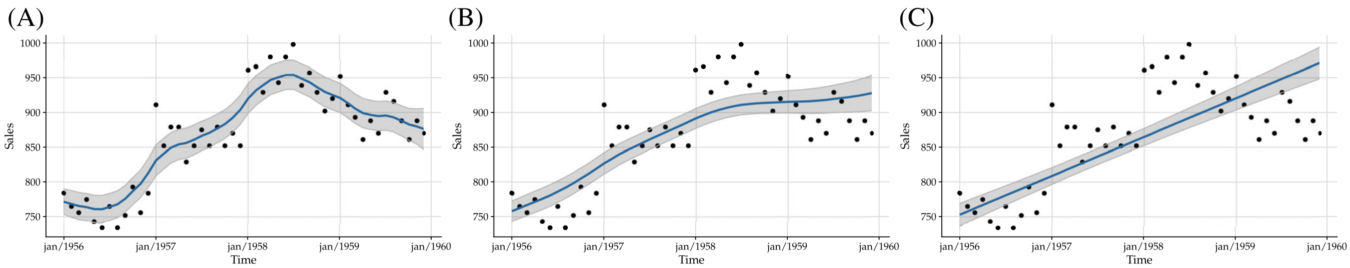

FIGURE 9 Expected values and 95 % credibility interval of the smoothed level for different discount factors. (A) $\delta = 0.80$. (B) $\delta = 0.95$. (C) $\delta = 1.0$. The MAPEs are 0.057, 0.071, and 0.079, respectively.

Harrison[3] establishes that the constant, V, can be estimated sequentially. In summary, we have $V|D_t \sim \text{Gamma}\left[\frac{n_t}{2}, \frac{d_t}{2}\right]$ where $n_t = n_{t-1} + 1$ and $d_t = d_{t-1} + \frac{e_t^2}{Q_t}$. Note that n_t is the number of degrees of freedom and d_t is the sum of the squares of the standardized prediction errors.

(iii) Discount factors and missing data

The use of discount factors to specify the variance of the evolution of the states (rather than estimating it), and the treatment of missing data are two important issues in dynamic models.

The sequential updating of DLMs depends on the specification of covariance matrices for the states evolution. Thus, the structure and magnitude specification of the $\mathbf{W}_t$ matrices are critical for modeling and forecasting using DLMs. In this regard, an intuitive and computational efficient solution is to use discount factors.

The discount factors, $0 < \delta \leq 1$, are constants that can be used in place of $\mathbf{W}_t$ when evaluating the states' prior variances, $\mathbf{R}_t$. Denoting $\mathbf{P}_t = \mathbf{G}\ \mathbf{C}_{t-1}\ \mathbf{G}'$, where $\mathbf{C}_{t-1}$ is the posterior variance at time $t-1$, gives $\mathbf{R}_t = \mathbf{P}_{t-1} + \mathbf{W}_t$. As can be seen, the role of $\mathbf{W}_t$ is to add uncertainty when passing from the posterior at time $t-1$ to the prior at time t. Alternatively, one can specify $\mathbf{R}_t = \mathbf{P}_{t-1}/\delta,\quad \delta \in (0, 1)$, as both increase the structural part of the previous variance, as desired. Both specifications are equivalent (one step ahead) if $\mathbf{W}_t = (1/\delta - 1)\mathbf{P}_{t-1}$. The discount factor can be chosen subjectively or through the optimization of an objective function. An educated guess about the discount factor can be calculated using the information's half-life (N) as follows: $\delta = \frac{3N-1}{3N+1}$, where N is the number of periods required for the information of an observation to decay to half its original value. It is important to remember that if $\delta = 1$, the model is static.

As discussed in Section 3.2, DLMs are usually designed in terms of structural blocks with the system matrix $\mathbf{G}$ assuming a block-diagonal form, with submatrices reflecting the contribution of individual component models. Consider a DLM model resulting from the superposition of r submodels M_i: $\{\mathbf{F}_{it}, \mathbf{G}_{it}, V_{it}, \mathbf{W}_{it}\}$, so that: $\mathbf{F}_t = (\mathbf{F}'_{1t}, \dots, \mathbf{F}'_{rt})'$, $\mathbf{G}_t = \text{diag}(\mathbf{G}_{1t}, \dots, \mathbf{G}_{rt})$ and $\mathbf{W}_t = \text{diag}(\mathbf{W}_{1t}, \cdots, \mathbf{W}_{rt})$. One can be interested in modeling the decay of current information at different rates for each structural component, leading to the adoption of different discounts $\delta_1, \dots, \delta_r$. Thus $\mathbf{W}_{it} = \frac{1-\delta_i}{\delta_i}\mathbf{P}_{it}$, which is the construction that West and Harrison[3] refer to as component discount DLM. Under this block discounting strategy, component covariances are left unchanged and block diagonal elements are divided by their respective discount factors, resulting in $\mathbf{R}_{it} = (1/\delta_i)\mathbf{P}_{it}, i = 1, \dots, r$.

The effect of different discount factors is illustrated through the analysis of a monthly tobacco time series in the UK (CP6).[3] Figure 9 shows the expected values and 95% credibility interval of the smoothed level ($E[\mu_t|D_T],\ t = 1, \dots, T$) with the following discount factors: $\delta = 0.80, 0.95$, and 1.0, for the CP6 series. When $\delta = 1$, notice the excessive smoothing and the attempt to better adapt to the data when the discount is set to 0.8, which corresponds to a three-period half-life. This time series will be re-analyzed in the next sections to illustrate subjective interventions.

One can also consider weighted observations, through the adoption of an observational variance divisor $V_t = V/k_t$. Thus an observation with little or no weight ($k_t \to 0$) has infinite variance (equivalently, null precision). Missing observations can be directly and simply handled assuming that the observation precision, at that specific moment, is null ($V_t^{-1} = k_t/V = 0$). As a result, the posterior at time t (the missing observation moment) will equal the prior at this time instant. As a consequence, its mean and variance will be $\mathbf{m}_t = \mathbf{a}_t$ and $\mathbf{C}_t = \mathbf{R}_t$.

(iv) Intervention analysis

External data can be included in the DLM analysis as long as it is probabilistically described. Therefore, marketing experts should be trained to use probability distributions to describe their subjective information or have the assistance of an analyst who can play that role (Garthwaite et al.[43]).

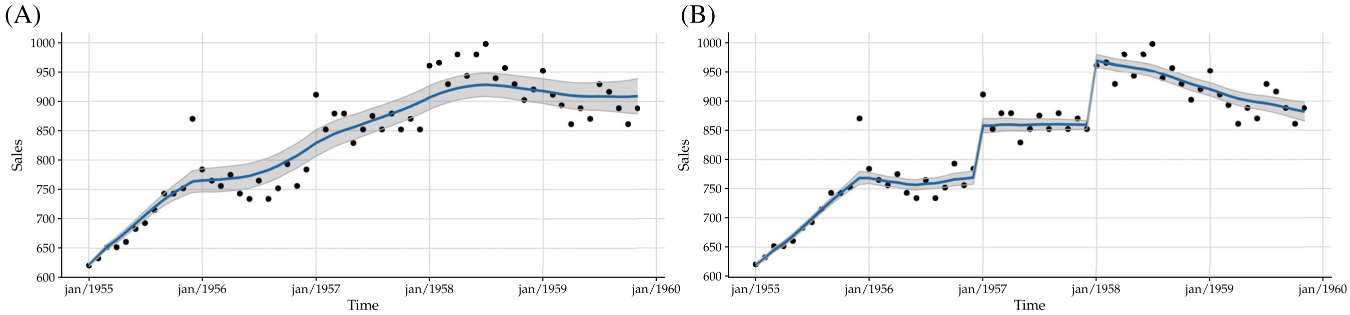

FIGURE 10 Expected values and 95% credibility intervals of the smooth level. (A) Without and (B) with intervention. The MAPEs are 0.029 and 0.019, respectively.

We highlight transitory and persistent interventions among the types of intervention (West and Harrison[3]). The first consists of adjusting the variance of the observations (V_t) to reduce or eliminate an outlying observation. The second is intended to incorporate the effect of regime changes and consists of introducing the component $\boldsymbol{\psi}_t \sim N[\mathbf{h}_t, \mathbf{H}_t]$, where $\mathbf{h}_t$ represents a subjective assessment of change in the states prior mean, at time t, and $\mathbf{H}_t$ is the uncertainty associated with this change.

Changes in level and growth components are of particular interest. Suppose that ψ_t is uncorrelated with $(\theta_{t-1}|D_{t-1})$ and with ω_t, the states will as follows: $\boldsymbol{\theta}_t = \mathbf{G}_t\ \boldsymbol{\theta}_{t-1} + \boldsymbol{\psi}_t + \boldsymbol{\omega}_t$. simplest type of intervention is just to change $\mathbf{W}_t$ arbitrarily, leaving the system in alert because some unknown event is about to occur.

To illustrate the intervention analysis, we again use the monthly tobacco sales series. It can be seen in Figure 10 that sales increased rapidly in 1955, but then decreased towards the end of the year. Early in 1957 and 1958, two level adjustments with higher volatility were observed. The growth factor of total sales changed in the last two years of the series, 1958 and 1959. The intervention analysis was carried out using a dynamic linear growth model as follows:

1. Observation y_{12} was ignored (*outlier*) and $\mathbf{H}_{12}$ was set appropriately to anticipate the change in the growth factor;
2. In Jan/1957 ($t = 25$), an elicited subjective intervention was undertaken through modifications in $\mathbf{h}_{25}$ and $\mathbf{H}_{25}$, in addition to treating y_{25} as an outlier, to anticipate the change in sales level;
3. In Jan/1958 ($t = 37$), changes occurred in both the level and the growth factor. A subjective intervention is performed by changing the values of the prior parameters.

Figure 10 exhibits the expected values of the smoothed level and the corresponding 95% credibility interval without (A) and with (B) intervention.

(v) Bayesian monitoring and interventions

In this section, we present a sequential and automatic monitoring method (West and Harrison[3]). This consists of verifying whether a future observation is consistent with forecasts made with the current model (M_0) or whether it favors some alternative model (M_1). The monitoring will be based on the most recent observations at time t. The concept of the local Bayes factor is defined as follows: $H_t(k) = \frac{p_0(y_t,\cdots\ y_{t-k+1})}{p_1(y_t,\cdots\ y_{t-k+1})}$, where p_0 and p_1 are the current model's and an alternative model's predictive distributions, respectively. For each $t > 1$, the Bayes factor accumulates multiplicatively: $H_t(k) = H_t H_{t-1}(k-1), k = 2, \ldots, t$. For each k, $H_t(k)$ evaluates the predictive performance of M_0 against M_1 for the k most recent observations. A small value of $H_t = H_t(1)$ signs for a possible outlier or the onset of a change in the time series, at time t. A small $H_t(k)$ for $k > 1$ indicates that, possibly, changes have taken place at least k instants back. Let $L_t = \min_{1 \le k \le t} H_t(k)$, with $L_1 = H_1$. This quantity can be calculated recursively using $L_t = H_t\ \min\{1, L_{t-1}\}$ and $l_t = 1 + l_{t-1}\, I_{(-\infty,1)}(L_{t-1})$, where $H_t = H_t(1)$ and l_t denotes the time the minimum occurs. This procedure depends on a threshold τ, $0 < \tau < 1$ and $k_{\max}$. Small values of $L_t < \tau$ or the accumulation of a maximum number of installments $l_t > k_{\max}$ indicate possible inadequacy of M_0. West and Harrison[3(pp 392–397)] provide a detailed discussion on model monitoring via Bayes factors.

The application of Bayesian monitoring is illustrated using two examples, one based on simulated data and the other on a real-time series. The monitoring parameters were set as $k_{\max} = 3$ and $\tau = 0.135$. The data in the first application is simulated from a normal generating process $N[\mu, \sigma^2]$, with $\mu = 100{,}104{,}98.$ and $\sigma^2 = 0.8\ , 0.5\ , 0.5$ respectively, for the first 40 observations, for the next 20 observations, and for the last 40.

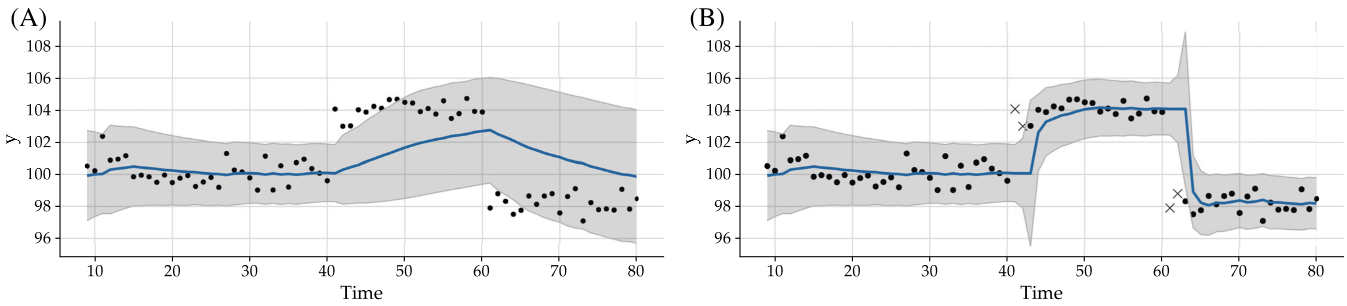

FIGURE 11 One-step-ahead forecasts with 95% credibility intervals for the simulated data, (A) without monitoring, (B) with monitoring. The MAPEs are 0.010 and 0.008, respectively.

TABLE 2 Summary of monitor detection.

t	Time	Detection type	e_t	l_t	L_t	H_t
48	Dec/1965	Upper outlier	2.51	1	0.13	0.13
60	Dec/1966	Upper outlier	3.18	1	0.01	0.01
72	Dec/1967	Upper outlier	2.65	1	0.08	0.08
77	May/1968	Lower outlier	−3.18	1	0.01	0.01
80	Aug/1968	Lower parametric change	−2.06	3	<0.001	0.78
84	Dec/1968	Upper outlier	2.54	1	0.11	0.11
97	Jan/1970	Upper parametric change	1.07	3	<0.001	42.05
115	Jul/1971	Lower outlier	−3.80	1	<0.001	<0.001
120	Dec/1971	Upper outlier	2.69	1	0.06	0.06
138	Jun/1973	Lower outlier	−2.53	1	0.12	0.12
140	Aug/1973	Lower outlier	−3.06	1	0.01	0.01
141	Sep/1973	Lower outlier	−4.64	1	<0.001	<0.001
146	Feb/1974	Lower outlier	−5.80	1	<0.001	<0.001
147	Mar/1974	Lower outlier	−11.82	1	<0.001	<0.001
148	Apr/1974	Lower outlier	−4.79	1	<0.001	<0.001

Evidence against model M_0 was detected at $t = 41$ and $t = 61$, with $L_{41} = 6.85\,e^{-6}$ and $L_{61} = 2.23\,e^{-4}$, both with $l_t = 1$, indicating a potential outlier. With the arrival of the following observations, a regime change is recognized by the monitor. The interventions performed can be observed in Figure 11, which shows that the monitored learning system quickly adapts to regime changes (B), compared to the learning system without monitoring (A).

The second illustration is based on the monthly average number of phone calls in Cincinnati, USA (Pankratz[44]). The series features three level changes. The first was in early 1968, with three months of impact; the second was in the middle of 1973, less significant; and the third was in early 1974, more lasting. Table 2 exhibits the monitor's detection.

Figures 11 and 12 show that the monitor sometimes abdicates model M_0 in favor of a more adaptive model. It is noteworthy that there are peaks of uncertainty at these moments, making it clear that the alternative model has inflated variance components, readily accommodating new patterns in moments of structural changes.

4 NON-NORMAL DYNAMIC MODELS

In the 1980s, developments in the area of Bayesian forecasting models were very intense. We highlight the use of discount factors by model component, applications to advertising data, models for sales data in a certain beverage industry, and the

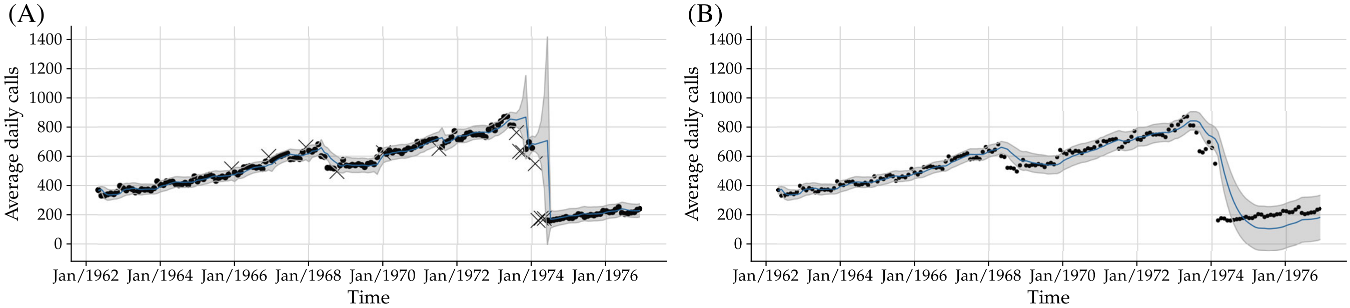

FIGURE 12 One-step-ahead forecasts with 95% credibility intervals for the phone calls series. (A) With monitoring. (B) Without monitoring. Observations represented by × indicate instants with intervention. The MAPEs are 0.208 and 0.076, respectively.

extension of the standard DLM to dynamic generalized linear models (DGLM) as well as nonlinear models (West et al.[45]). The DGLM class presented by West et al.[45] includes models for counting data and for continuous scale data. Observations are described by probabilistic models in the uniparametric exponential family.

4.1 | Dynamic generalized linear models-DGLM

Let EF(η_t) denote a distribution in the uniparametric exponential family, parameterized by η_t. The class of dynamic generalized models is defined as:

$$\begin{aligned} y_t|\eta_t &\sim \mathrm{EF}(\eta_t), \\ g(\eta_t) &= \lambda_t = \mathbf{F}_t' \, \boldsymbol{\theta}_t, \\ \boldsymbol{\theta}_t &= \mathbf{G}_t \, \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim [0, W_t]. \end{aligned} \tag{9}$$

Note that now the observed data, conditionally independent, given η_t, are members of the exponential family, and the evolution of the states follows a Markovian rule, but with partially specified probabilistic structure, where $[A, B]$ denotes a distribution partially specified in terms first and second moments. In addition, we now have to deal with the monotone differentiable link function $g(.)$, which relates the exponential family parameter η_t to the linear predictor λ_t.

The exponential family is characterized by probability or density functions in the form:

$$p(y_t|\eta_t, \phi_t) = b(y_t, \phi_t) \;\; exp[\phi_t^{-1}\{y_t\eta_t - a(\eta_t)\}], \quad y_t \in \mathcal{Y}, \tag{10}$$

where $\eta_t \in \mathcal{H} \subset \mathcal{R}$ is the natural parameter, $\phi_t > 0$ is a scale parameter, $a(\eta)$ is a twice-differentiable function from which the expected value and variance can be derived: $E[y_t|\eta] = \frac{d}{d\eta}\, a(\eta)$ and $\mathrm{Var}[y_t|\eta] = \phi_t \, \frac{d^2}{d^2\eta}\, a(\eta)$. The conjugate distribution for η_t is $p(\eta_t|D_{t-1}) = c(r_t, s_t)\;\, \exp[r_t\;\eta_t - s_t\; a(\eta)]$, where r_t and s_t are known functions of D_{t-1}.

Once y_t is observed, the conjugate posterior distribution can be easily obtained by updating $r_t^* = r_t + \frac{y_t}{\phi}$ and $s_t^* = s_t + \frac{1}{\phi}$. The predictive distribution is also easily available: $p(y_t|D_{t-1}) = \frac{c(r_t,s_t)\; b(y_t,\phi_t)}{c(r_t+\phi_t^{-1}y_t,\; s_t+\phi_t^{-1})}$.

The sequential inference approach for observations in a single parametric exponential family relies solely on partially specified prior knowledge, in terms of first and second moments, and on the Linear Bayes method. Three steps are necessary to complete the inference (see details in West et al.[45]), since we are also interested in the structure of the linear predictor:

(i) Make prior distributions compatible, which implies relating (r_t, s_t) to (f_t, q_t) through the link function g, where $f_t = E[\lambda_t|D_{t-1}]$, $q_t = Var[\lambda_t|D_{t-1}]$.
(ii) After updating the exponential family parameters, obtain $f_t^* = E[\lambda_t|D_t]$ and $q_t^* = \mathrm{Var}[\lambda_t|D_t]$ compatible with r_t^*, s_t^*.
(iii) Finally, use the linear Bayes method to obtain the first and second moments of the posterior of the state parameters.

4.2 | Sequential inference

In this class of models sequential inference demands some approximations. The first consists of reconciling, at each time, the parameters of the conjugate prior distribution (r_t, s_t) with those of the linear predictor (f_t, q_t), which in turn depend on the first and second moments of the prior distribution of the states $(\mathbf{a}_t, \mathbf{R}_t)$. If on the one hand the linearity of the linear predictor facilitates the computations, on the other the link function must be linearized. Remember that states and natural parameters of the exponential family assume values in different spaces.

The graph in Figure 13 exhibits the different stages of the inference procedure in DGLMs (see West and Harrison[3] for details).

4.3 | Illustration: A Poisson model for quarterly sales

The time series of total quarterly sales of turkeys in Eire during the years 1974 to 1982 is used to exemplify a Poisson log-linear dynamic model composed of a second order polynomial trend block (level and growth factor) and another structural block representing quarterly seasonality, with 2 harmonics. The discount factors for trend and seasonality blocks are respectively 0.95 and 0.98, reflecting the usual fact that level and trend vary more than seasonal patterns. Vague priors are used to specify the first two moments.

Figure 14 shows one-step-ahead predictive modes (blue) and 95% credibility intervals (gray area). Note that in this model, the exact predictive distribution is available: $y_y|D_{t-1} \sim \text{BNeg}[\alpha_t, \beta_t/(1-\beta_t)]$. The model presented was able to identify the changes in the seasonal patterns (Figure 15A,D) induced by marketing campaigns.

Other exponential family models that could be useful in marketing applications are Binomial, for instance accounting for the number of consumers aware of an advertised product; multinomial, for multiple counts

$$
\begin{tikzcd}
{} \arrow[r] & (\boldsymbol{\theta}_t \mid D_{t-1}) \sim [\mathbf{a}_t, \mathbf{R}_t] \arrow[d, "\lambda_t = \mathbf{F}_t^\top \boldsymbol{\theta}_t"] & & (\boldsymbol{\theta}_t \mid D_t) \sim [\mathbf{m}_t, \mathbf{C}_t] \arrow[r] & {} \\
& (\lambda_t \mid D_{t-1}) \sim [f_t, q_t] \arrow[d, "g^{-1}(\lambda_t)"] & & (\lambda_t \mid D_t) \sim [f_t^*, q_t^*] \arrow[u, "\text{LB}"] & \\
& (\eta_t \mid D_{t-1}) \sim \text{CD}(\cdot,\cdot) \arrow[rr, "(y_t \mid \eta_t) \sim \text{CD}(\cdot,\cdot)"] & & (\eta_t \mid D_t) \sim \text{CD}(\cdot,\cdot) \arrow[u, "g(\lambda_t)"] &
\end{tikzcd}
$$

FIGURE 13 Inference graph-DAG.

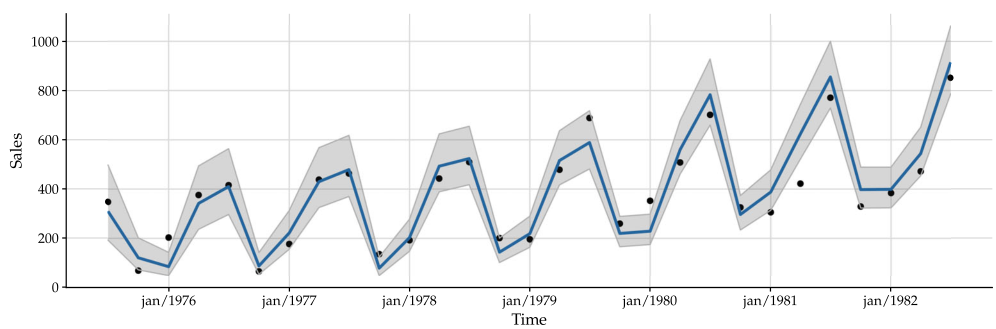

FIGURE 14 Mode of the predictive distribution and 95% credible interval. The MAPE is 0.1922.

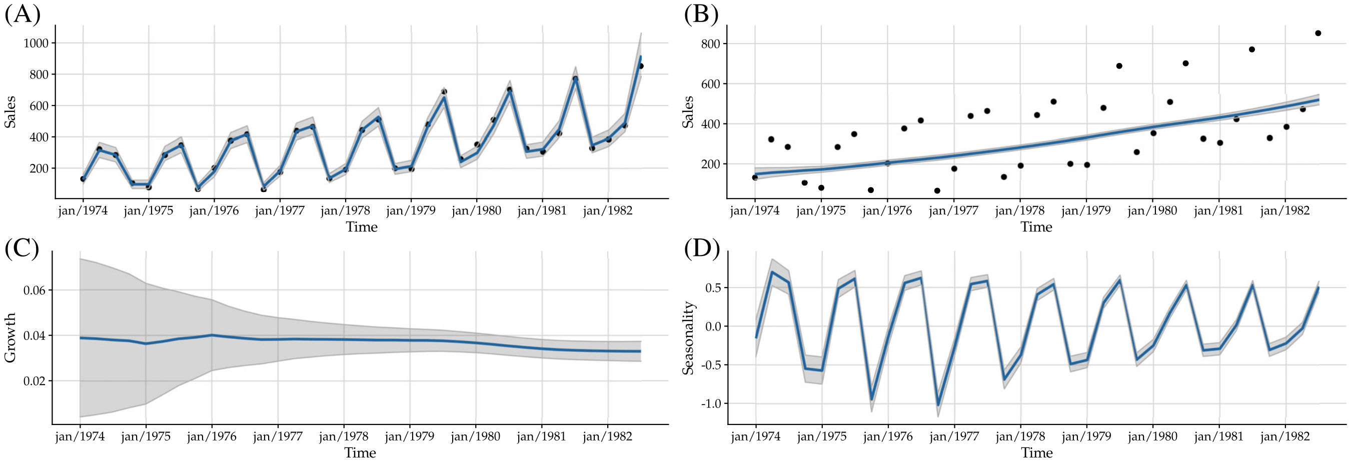

FIGURE 15 Smoothed expected values with the corresponding ranges of about 2 standard deviations from the model components. (A) Predictive. (B) Current level. (C) Growth factor. (D) Seasonal component.

in marketing structure etc. Additional comprehensive examples in the exponential family are provided by Triantafyllopoulos.[46]

5 | REAL-WORLD CHALLENGES

This section discuss the model formulation of two real-world problems solved using Bayesian dynamic models by two of the authors of this paper during their work at Murabei Data Science.

5.1 | Marketing planning decision of returnable containers

A real world challenge solved by two of the authors while working for Murabei Data Science using dynamic latent factors and the copula approach to represent the dependence relationship between multivariate time series is discussed here.

This challenge addresses forecasting the availability of returnable containers in production plants. This topic is import for different sectors, such as alcoholic and soft drinks, water and cooking gas, among others. The use of returnable packaging has gained attention in recent years, prompted by various laws and regulations on reverse logistics, which specify responsibility for the full life cycle of products, based on the circular economy idea, reflecting the increased concerns of consumers about the social and environmental impacts of waste disposal (Narayana[47]). Additionally, having a forecast probability distribution for available containers can leverage marketing campaigns that encourage customers to return their empty containers.

The random variable Y_t denotes available containers at time t and is modeled using a Bayesian dynamic linear model with dynamic level and growth components. The Bayesian dynamic model can be used to derive a measure of uncertainty about the factory's lack of containers, giving a more accurate picture of the risk of production stoppage than a point forecast. The h-step lead time forecast distribution is of particular interest in this context and is defined as the distribution of the aggregate:

$$X_t(h) = Y_{t+1} + Y_{t+2} + \ldots + Y_{t+h}, \quad h > 0.$$

Given the available information D_t up to time t, it is possible to obtain an estimate of the probability of container shortage for the next h months using h-step lead time forecast distribution, as:

$$\pi_h = \Pr\left[X_t(h) < N_o | D_t\right],$$

where N_o is a fixed and known number related to the operational need of the factory. Decisions based on the model's results can lead to cost savings as well as better production and delivery scheduling.

5.2 Pre-launch forecast

Another common and difficult task in marketing is to forecast a new product's initial few months of sales. Obtaining a medium to long-term pre-launch forecast is critical for planning purposes, such as production, distribution and advertising campaigns. This is challenging because accurate forecasts require a large amount of past data.

The diffusion model (Schmittlein and Mahajan[26]), more specifically the Bass model (Bass[25]), is one of the models that has been extensively investigated for new product sales forecasts (Bass[25]). Because of its interpretation, the Bass model is particularly attractive.

The previous model was extended by Bass et al.,[27] allowing it to include factors such as price growth rate and advertisement. As already mentioned, Ramírez-Hassan and Montoya-Blandón[28] recently proposed a Bayesian inference approach for the generalized Bass model. Although the Bayesian approach has several fundamental aspects to guide marketers, such as prior specifications by marketing experts and predictive distributions to support the decision making, taking into account the uncertainty, there are some drawbacks to this model that can be easily solved by the dynamic generalized exponential growth models (DGEGM) introduced by Gamerman and Migon[48] and Migon and Gamerman.[29] Specifically, the Bayesian inference proposed by Ramírez-Hassan and Montoya-Blandón[28] is computationally expensive, since it requires Markov chain Monte Carlo (MCMC) methods, while inference in DGEGMs is performed sequentially and analytically.

The class of DGEGMs assumes that y_t is modeled by a probability distribution on the exponential family with mean response function $\mu_t = \mathbb{E}(y_t|\boldsymbol{\theta}_t)$, where $\boldsymbol{\theta}_t$ denote the state vector parameters. Application of a Box-Cox link function that maps the mean of y_t, μ_t, to the state parameters vector, that is,

$$h(\mu_t) = \begin{cases} \mu_t^{\lambda}, & \text{if } \lambda \neq 0, \\ \log \mu_t, & \text{if } \lambda = 0, \end{cases} \tag{11}$$

where λ is a known value, leads to well-known models such as the modified exponential ($\lambda = 1$), logistic ($\lambda = -1$), and Gompertz ($\lambda = 0$).

Using linearization techniques based on Taylor approximations, the DGEGM can be approximate as a dynamic linear model and all inferences discussed are valid. The DGEGM is defined as follows

$$\begin{aligned} y_t &= F_t(\boldsymbol{\theta}_t) + \nu_t, \\ \boldsymbol{\theta}_t &= \mathbf{g}_t(\boldsymbol{\theta}_{t-1}) + \boldsymbol{\omega}_t, \end{aligned}$$

where $F_t(\cdot)$ is a known, nonlinear function mapping the vector $\boldsymbol{\theta}_t$ to the real line, $\mathbf{g}_t(\cdot)$ is a known nonlinear vector evolution function, and ν_t and $\boldsymbol{\omega}_t$ are error terms subject to the usual DLM assumptions.

To complete the model specifications $F_t(\cdot)$ and $\mathbf{g}_t(\cdot)$ are defined as:

$$F_t(\boldsymbol{\theta}_t) = h^{-1}(\theta_{1t}) \quad \text{and} \quad \mathbf{g}_t(\boldsymbol{\theta}_{t-1}) = \begin{pmatrix} \theta_1 + \theta_2 \\ \theta_2\ \theta_3 \\ \theta_3 \end{pmatrix},$$

where $h^{-1}(\cdot)$ denotes the inverse of (11).

Further details concerning the inference in DGEGMs can be found in Migon and Gamerman[29] and chapter 13 of West and Harrison.[3] It should be mentioned that the R package `RBATS` provides computational implementation for this class of models through the class `dgegm`.[35]

6 RECENT ADVANCES

This section is devoted to discussing relevant recent research advances in the field of Bayesian dynamic models. A recent overview is presented by West[21] with main focus on the "decouple/recouple" concept, applied to large-scale data. Alves et al.[23] recently extended the DGLMs by taking into account the k-parametric family of distributions, which can include, for example, the multinomial distribution. A distinct class of models, that was just recently developed by Berry and West,[49] combines binary and conditionally Poisson DGLMs to handle zero-inflated time series data. In the sequel, we present two alternative versions that explore particular specifications of the dependence structure on multiple observed quantities.

6.1 | Linear matrix-variable dynamic models

Matrix-varied models have been used to facilitate the estimation of the covariance matrix of m time series. Modeling of wind intensity and direction is a recent example (Garcia et al.[50]), in which case $m = 2$.

$$\begin{aligned} y_{t,j} &= F_t'\theta_{t,j} + \epsilon_{t,j}, \quad \epsilon_{t,j} \sim N[0, V_t\sigma_j^2], \\ \theta_{t,j} &= G_t\theta_{t-1,j} + \omega_{t,j}, \quad \omega_{t,j} \sim N[0, W_t\sigma_j^2], \end{aligned} \tag{12}$$

where $j = 1, \ldots, m$, F_t, G_t, and Ω_t are common to the m series, V_t is a scale factor common to all series but $\theta_{t,j}$ are distinct, reflecting the fact that the structural components can have different impacts on each of the m responses. In matrix form, we have:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{F}_t'\boldsymbol{\theta}_t + \epsilon_t', \\ \boldsymbol{\theta}_t &= \mathbf{G}\ \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \end{aligned}$$

where $\mathbf{y}_t = (y_{1,t}, \cdots, y_{m,t})'$, $\boldsymbol{\theta}_t = (\theta_{t,1}, \cdots, \theta_{1,q})$ is a $q \times m$ matrix, $\boldsymbol{\epsilon}_t \sim N[0, V_t\Sigma]$ and $\boldsymbol{\omega}_t \sim MN[0, \mathbf{W}_t, \Sigma]$.

It is worth noting that Barbosa[51] created a broader framework for multivariate DLMs in which the model can include various components. Extensions of those concepts can be found in Triantafyllopoulos.[52]

6.2 | Dynamic hierarchical models

The dynamic hierarchical model introduced by Gamerman and Migon[53] is very useful for describing multivariate structures in a simple manner. The following is an example of a two-level hierarchical structure: at the most aggregated level, we have a dynamic regression equation (observation equation); at the second level, the regression coefficients vary between groups (structural equation). Finally, the effects shared by the various groups change over time in an efficient way. Therefore we have:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{F}_{1,t}\ \boldsymbol{\theta}_{1,t} + \epsilon_{1,t}, \\ \boldsymbol{\theta}_{1,t} &= \mathbf{F}_{2,t}\ \boldsymbol{\theta}_t + \epsilon_{2,t}, \\ \boldsymbol{\theta}_t &= \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \end{aligned} \tag{13}$$

where the disturbance terms $\boldsymbol{\epsilon}_{1,t}$, $\boldsymbol{\omega}_t$, $\boldsymbol{\epsilon}_{2,t}$ are independent and $\mathbf{F}_{1,t}$, $\mathbf{F}_{2,t}$ are known matrices.

The inference in this class of models follows the usual evolution/update schemes (Gamerman and Migon[53]) as in normal dynamic linear models, being trivial when variance components are known. It is easy to verify that the implicit multivariate structure in this class of models corresponds to a special decomposition of variance. Marginalizing the previous model, that is, replacing the structural equation in the equation of observations, we have $\mathbf{y}_t = \mathbf{x}_t\ \boldsymbol{\theta}_{2,t} + \mathbf{a}_t$, where $\mathbf{x}_t = \mathbf{F}_{1,t}\ \mathbf{F}_{2,t}$ and $\mathbf{a}_t = \mathbf{F}_{1,t}\epsilon_{2,t} + \epsilon_{1,t}$. Thus, $\text{Var}[\mathbf{a}_t] = \boldsymbol{\Sigma}_t = \mathbf{F}_{1,t}\mathbf{V}_{2,t}\mathbf{F}_{1,t}' + \mathbf{V}_{1,t}$.

7 | CONCLUDING REMARKS

In this article, we have sought to provide an overview of the Bayesian dynamic model class, which, although first presented more than five decades ago, is still relatively little explored in several application areas, including marketing. The class, presented under the Bayesian inference paradigm, is shown to be a formulation that naturally accommodates the dynamics of temporally observed processes.

Some aspects such as monitoring and intervention were formally discussed, allowing the analyst to perform interventions at the level and uncertainty of the machine learning latent system, quickly accommodating pattern changes in the time series, enabling accurate forecasting that accommodates new standards in a timely manner, allowing efficient management of resources and real-time action planning.

We also discussed the modeling of lagged impacts of regressors, through transfer functions, exemplified by the effect of advertising campaigns on consumers recall of a product. There can be several input variables in a system, each of which have an effect on the response by undergoing some form of propagation in time. The treatment of the analytical form of such propagation and the quantification of these effects can be a valuable tool, for example, to plan strategic moments for the launch of marketing campaigns, aiming to maintain sales or memory of a product above desirable thresholds.

As seen, the Bayesian sequential inferential updating of the class of Gaussian dynamic linear models is amenable to analytical solution, by adopting conjugated formulations allied to discount factor strategies for indirect specification of evolutional variances. There are, however, several responses of practical interest for which the Gaussian assumption is inadequate, such as due to asymmetric behavior or discrete nature, as in the case of low count data or binary responses. An alternative discussed was the transformation of the original data to induce normality. However, not every temporally observed variable can be normalized and although the induction of normality via transformations generates satisfactory results, there is a price to be paid from the point of view of interpretability of the obtained results. The maintenance of the data in its original scale of observation is desirable, at the cost of not having an available analytical formulation for updating the Bayesian information system. Approximate solutions, for example through Monte Carlo Markov chain methods, can result in high computational cost and are not capable of preserving the sequential aspect of the analysis, invalidating the possibility of real-time monitoring and intervention. The article reviewed some alternatives to perform sequential inference in reduced computational time, in the class of dynamic generalized linear models, for responses belonging to the uni or multiparametric exponential family.

Another relevant aspect that was discussed concerns methods that seek computational efficiency in the approach to multivariate dynamic models, via multivariate or hierarchical formulations in their genesis, or even through strategies of decoupling/ recoupling, in which series are marginally modeled and then recoupled through copulas, common factors or conditioning on some aggregate response.

We aimed to present an overview of the class of dynamic models and highlight their importance as a valuable tool for monitoring, intervention, forecasting and efficient management of temporal observed processes in real time. We believe that from discussions, examples and references presented here, readers can explore the aspects of their interest.

ACKNOWLEDGMENTS

This research was funded in part by Fapesp Grants 2018/04654 and PAPD/UERJ, E-26/007/10667/2019 (Helio S. Migon). *Murabei Data Science* sponsored these developments, which the authors are grateful for.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings were all obtained from books which are cited in the paper.

ORCID

Helio S. Migon https://orcid.org/0000-0003-4913-3210
Mariane Branco Alves https://orcid.org/0000-0002-2489-9300

REFERENCES

1. Harrison PJ, Stevens CF. A Bayesian approach to short-Term forecasting. *Operat Res Quart (1970-1977)*. 1971;22(4):341-362.
2. Leeflang Peter SH, Bijmolt Tammo HA, van Doorn J, et al. Creating lift versus building the base: Current trends in marketing dynamics. *Int J Res Mark*. 2009;26(1):13-20. doi:10.1016/j.ijresmar.2008.06.006
3. West M, Harrison PJ. *Bayesian Forecasting and Dynamic Models*. 2nd ed. Springer; 1997.
4. Johnston FR, Harrison PJ. An application of forecasting in the alcoholic drinks industry. *J Oper Res Soc*. 1980;31(8):699-709.

5. Migon HS, Harrison PJ. An application of non-linear Bayesian forecasting to television advertising. *Bayesian Stat*. 1985;2:681-696.
6. Migon HS. The prediction of Brazilian exports using Bayesian forecasting. *Investigat Operat*. 2000;9:95-106.
7. Migon HS, Pereira BB, Suyama E. *Dynamic Bayesian Modelling and Optimal Portfolio Choice*. Universidade Federal do Rio de Janeiro –Departamento de Matemática e Estatística; 1990.
8. Polson NG, Tew BV. Bayesian Portfolio Selection: An Empirical Analysis of the S&P 500 Index 1970–1996. *J Bus Econ Stat*. 2000;18(2):164-173.
9. Migon HS, Gamerman D, Lopes HF, Ferreira MAR. Dynamic models. *Handbook Stat*. 2005;25:557-592.
10. Berry LR, Mike W. Bayesian Forecasting of Many Count-Valued Time Series. *J Bus Econ Stat*. 2020;38(4):872-887.
11. Yanchenko AK, Graham T, Joseph L, Christoph H, Andrew C, Mike W. Multivariate dynamic modeling for Bayesian forecasting of business revenue. *Appl Stoch Model Bus Ind*. 2022:1-18.
12. Rossi PE, Allenby GM. Bayesian Statistics and Marketing. *Mark Sci*. 2003;22(3):304-328.
13. Rossi PE, Allemby GM, McCulloch R. *Bayesian Statistics and Marketing*. Wiley; 2005.
14. Pauwels KH. Modeling dynamic relations among marketing and performance metrics. *Found Trends Market*. 2018;11(4):215-301.
15. Dekimpe MG, Hanssens DM. Time-Series Models in Marketing: Some Recent Developments. *Marketing JRM*. 2010;1:24-29.
16. Chandukala SR, Dotson JP, Brazell JD, Allenby GM. Bayesian Analysis of Hierarchical Effects. *Mark Sci*. 2011;30(1):123-133.
17. Lopes HF. *Modern Bayesian Factor Analysis*. John Wiley & Sons; 2014:115-153.
18. Frühwirth-Schnatter S. Data augmentation and dynamic linear models. *J Time Ser Anal*. 1994;15(2):183-202.
19. Carter Chris K, Robert K. On Gibbs sampling for state space models. *Biometrika*. 1994;81(3):541-553.
20. Gamerman D. Markov chain Monte Carlo for dynamic generalised linear models. *Biometrika*. 1998;85(1):215-227.
21. West M. Bayesian forecasting of multivariate time series: scalability, structure uncertainty and decisions. *Ann Inst Stat Math*. 2020; 72:1-31.
22. Lavine I. *Bayesian Computation for Variable Selection and Multivariate Forecasting in Dynamic Models*PhD thesis. Duke University; 2020.
23. Alves MB, Migon HS, Marotta R, Santos Jr SV. *k*-parametric dynamic generalized linear models: A sequential approach via Information Geometry. arXiv preprint, arXiv:2201.05387, 2023.
24. Fan T, Golder N, Lehmann DR. Innovation and new products research: a state-of-the-art review, models for managerial decision making, and future research directions. In: Wierenga B, Lans R, eds. *Handbook of Marketing Decision Models*. International Series in Operations Research & Management Science. Vol 254. Springer; 2017:79-116.
25. Bass FM. A new product growth for model consumer durables. *Manag Sci*. 1969;15(5):215-227.
26. Schmittlein DC, Vijay M. Maximum Likelihood Estimation for an Innovation Diffusion Model of New Product Acceptance. *Mark Sci*. 1982;1(1):57-78.
27. Bass FM, Krishnan TV, Jain DC. Why the Bass Model Fits without Decision Variables. *Mark Sci*. 1994;13(3):203-223.
28. Ramírez-Hassan A, Montoya-Blandón S. Forecasting from others' experience: Bayesian estimation of the generalized Bass model. *Int J Forecast*. 2020;36(2, (4/6)):442-465.
29. Migon HS, Dani G. Generalized exponential growth models a Bayesian approach. *J Forecast*. 1993;12(7):573-584.
30. Blattberg RC, Briesch RA, Fox EJ. How promotions work. 1995.
31. Siddhartha C, Padma S, Andrei S. Model of Brand Choice with a No-Purchase Option Calibrated to Scanner-Panel Data. *J Mark Res*. 2004;41:184-196.
32. Padma S. Modeling Multiple Sources of State Dependence in Random Utility Models: A Distributed Lag Approach. *Mark Sci*. 2004;23:263-271.
33. Fernandez PJ, Lauretto MS, Pereira CAB, Stern JM. A new media optimizer based on the mean-variance model. *Pesquisa Operacional*. 2007;27(3):427-456.
34. Terui N, Ban M, Maki T. Finding market structure by sales count dynamics – Multivariate structural time series models with hierarchical structure for count data. *Ann Inst Stat Math*. 2010;62:91-107.
35. Menezes AFB. RBATS: Bayesian Dynamic Models2021. https://github.com/AndreMenezes/RBATS.
36. Lavine I, Cron A. PyBATS: Bayesian time series modeling and forecasting. 2021. https://pypi.org/project/pybats/.
37. Menezes A, Pinheiro EG. pyBATS-detection: A python package for outlier and structural changes detection in time series analysis. 2022. https://pypi.org/project/pyBATS-detection/.
38. Migon HS, Gamerman D, Louzada F. *Statistical Inference: An Integrated Approach*. 2nd ed. Chapman and Hall; 2015.
39. Harrison PJ, Stevens CF. Bayesian forecasting. *J R Stat Soc B Methodol*. 1976;38(3):205-247.
40. Alves MB, Gamerman D, Ferreira MAR. Transfer function in dynamic generalized linear models. *Stat Model*. 2010;10:3-40.
41. Ravines RR, Schmidt AM, Migon HS. Revisiting distributed lag models through a Bayesian perspective. *Appl Stoch Model Bus Ind*. 2006;22:193-210.
42. Pole A, West M, Harrison PJ. *Applied Bayesian Forecasting and Time Series Analysis*. Chapman & Hall/CRC; 1994.
43. Garthwaite PH, Kadane JB, O'Hagan A. Statistical Methods for Eliciting Probability Distributions. *J Am Stat Assoc*. 2005;100(470): 680-701.
44. Pankratz A. *Forecasting with Dynamic Regression Models*. Willey; 1992.
45. West M, Harrison PJ, Migon HS. Dynamic generalized linear models and Bayesian forecasting. *J Am Stat Assoc*. 1985;80(389): 73-978.
46. Triantafyllopoulos K. *Dynamic Generalized Linear Models for Non-Gaussian Time Series Forecasting*; 2008. arXiv.

47. Narayana SA, Pati RK, Padhi SS. Market dynamics and reverse logistics for sustainability in the Indian pharmaceuticals industry. *J Clean Prod*. 2019;208:968-987.
48. Gamerman D, Migon HS. Forecasting the Number of Aids Cases in Brazil. *J Royal Stat Soc D*. 1991;40(4):427-442.
49. Berry LR, West M. Bayesian Forecasting of Many Count-Valued Time Series. *J Bus Econ Stat*. 2019;38(4):1-34.
50. García I, Huo S, Prado R, Bravo L Dynamic Bayesian temporal modeling and forecasting of short-term wind measurements. *Renew Energy*. 2020;161:55-64.
51. Barbosa EP. *Dynamic Bayesian Models for Vector Time Series Analysis & Forecasting*, PhD thesis. University of Warwick; 1989.
52. Triantafyllopoulos K. *On Observational Variance Learning for Multivariate Bayesian Time Series and Related Models*, PhD thesis. University of Warwick; 2002.
53. Gamerman D, Migon HS. Dynamic Hierarchical Models. *J R Stat Soc Ser B*. 1993;55(3):629-642.

How to cite this article: Migon HS, Alves MB, Menezes AFB, Pinheiro EG. A review of Bayesian dynamic forecasting models: Applications in marketing. *Appl Stochastic Models Bus Ind*. 2023;1-23. doi: 10.1002/asmb.2756

APPENDIX . BAYESIAN CONJUGATE ANALYSIS IN THE EXPONENTIAL FAMILY: SOME EXAMPLES

A.1 Normal model with unknown mean and given variance

This model is useful for describing continuous, symmetrically distributed data. The density function is $p(y|\mu) \propto \exp\left[-\frac{1}{2\sigma^2}(y-\mu)^2\right], y \in \mathfrak{R}, \mu \in \mathfrak{R}$, where $\propto$ means proportional to, $\mu = E\left[y|\mu,\sigma^2\right]$ and $\sigma^2 = \text{Var}[y|\mu,\sigma^2]$. The likelihood function is $\ell(\mu|y) \propto \exp\left[-\frac{1}{2\sigma^2}(\mu-y)^2\right]$.

One can adopt conjugated priors, that is, priors which, when combined with the likelihood, produce a posterior in the same parametric family. According to Bayes' theorem, $p(\mu|y) \propto \exp\left[-\frac{1}{2\sigma^2}(y-\mu)^2\right] \cdot \exp\left[-\frac{1}{2\tau_0^2}(\mu-\mu_0)^2\right]$, where the conjugated prior distribution is $N[\mu_0, \tau_0^2]$, with $\psi = (\mu_0, \tau_0^2)$ specified by the analysts in order to describe their beliefs, prior to the observation of sample information, about the value of μ. After some algebraic operations, we have: $\mu|y, \psi_1 \sim N[\mu_1, \tau_1^2]$, where $\mu_1 = \omega\mu_0 + (1-\omega)y$, $\omega = \frac{\tau_0^{-2}}{\tau_0^{-2}+\sigma^{-2}}$ and $\tau_1^{-2} = \tau_0^{-2} + \sigma^{-2}$. Note that the posterior mean will be a linear combination of the prior mean and the sample observation, with weights given by the relative precision of each of these components.

A.2 Poisson model

This is a model for count data. Its probability function is $p(y|\lambda) = \frac{1}{y!}\lambda^y \exp(-\lambda), y \in \{0, 1, 2, \cdots\}, \lambda \in \mathfrak{R}_+$ and $E\left[y|\lambda\right] = \text{Var}[y|\lambda] = \lambda$. Its likelihood function is $\ell(\lambda|y) \propto \lambda^y \exp(-\lambda)$.

Since $\lambda > 0$, it is reasonable to use a family of priors that places probability mass on the positive reals. The gamma family, whose density is: $p(\lambda|a,b) = \frac{b^a}{\Gamma(a)}\lambda^{a-1} \exp[-b\lambda], a, b > 0$, where $\psi = (a, b)$, satisfies this condition. In this family we have $E[\lambda|a,b] = \frac{a}{b}$ and $\text{Var}[\lambda|a,b] = \frac{a}{b^2}$, where a, b should be subjectively chosen in order to portray prior knowledge about λ. From Bayes' theorem, we have: $(\lambda|y,a,b) \sim \text{Gamma}(a_1, b_1), a_1, b_1 > 0$, where $a_1 = a + y$ and $b_1 = b + 1$. The posterior mean is $E[\lambda|y,a,b] = \frac{a_1}{b_1} = \omega\mu_0 + (1-\omega)y$, $\omega = \frac{b}{b+1}$ and $\text{Var}[\lambda|y,a,b] = \frac{a_1}{b_1^2}$. It can be shown that the predictive distribution of a future observation, in this case, will be a Negative Binomial, $\text{BNeg}(a_f, b_f)$, $a_f = a_1$, $b_f = b_1 + 1$. Its mean and variance are given by $E[y_f|y] = E_{\lambda|y}\left\{E[y_f|\lambda]\right\} = \frac{a_1}{b_1} = \frac{a+y}{b+1}$ and $\text{Var}[y_f|y] = E_{\lambda|y}\left\{\text{Var}[y_f|\lambda]\right\} + \text{Var}_{\lambda|y}\left\{E[y_f|\lambda]\right\} = \frac{a_1}{b_1}\,(1 + \frac{1}{b_1}) \geq \text{Var}(\lambda|y)$.

A.3 Likelihood, prior and posterior: Normal and Poisson models

In Figure A1 we present the prior distribution (blue), the likelihood function (black) and the posterior distribution (red) for the examples of the normal data model (left) and the Poisson data (right). The values used in the elaboration of this figure were: (i) prior distribution $\mu \sim N[2, 6]$, observational distribution $y|\mu \sim N[\mu, 3]$ and the observed value $y = 10$; (ii) prior distribution $\lambda \sim \text{Gamma}(2, 0.25)$, observational distribution $y|\lambda \sim \text{Poisson}[\lambda]$ and observed value $y = 12$.

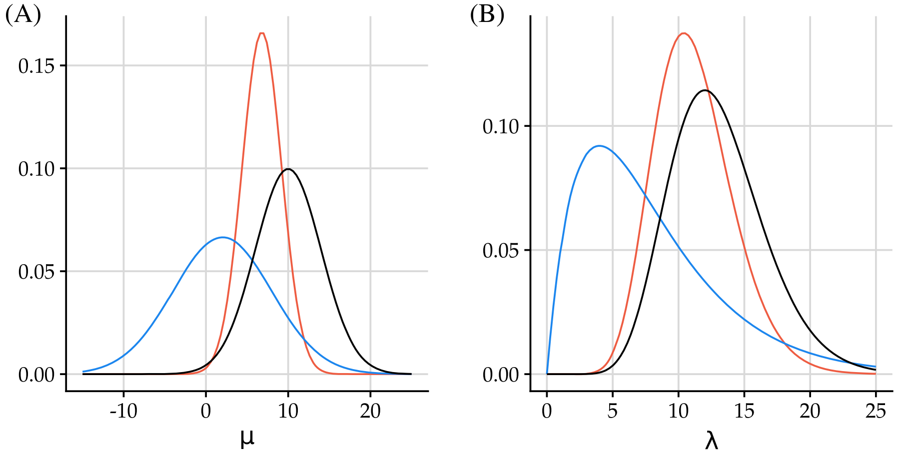

FIGURE A1 Prior distribution (blue), likelihood function (black) and posterior distribution (red). (A) Normal model (μ). (B) Poisson model (λ).

The above results are based on hypotheses of conditional independence and extend naturally to the case of random samples, that is, independent and identically distributed observations.