Dynamic Generalized Linear Model via Product Partition Model

Jéssica da Assunção Almeida

Departamento de Estatística - ICEx - UFMG

Fevereiro de 2016

Dynamic Generalized Linear Model via Product Partition Model

Jéssica da Assunção Almeida

Orientador: Fábio Nogueira Demarqui
Co-orientador: Thiago Rezende dos Santos

Dissertação apresentada ao programa de Pós-graduação em Estatística da Universidade Federal de Minas Gerais.

Departamento de Estatística
Instituto de Ciências Exatas
Universidade Federal de Minas Gerais

Belo Horizonte, MG - Brasil
Fevereiro de 2016

Aos meus pais.

“Por vezes sentimos que aquilo que fazemos não é senão uma gota de água no mar. Mas o mar seria menor se lhe faltasse uma gota”.

Agradecimentos

Ao meus pais, Maria José e Antônio, pelo incentivo e apoio incondicional de sempre.

Ao Pedro, pela compreensão, carinho e paciência.

Ao meu orientador, professor Fábio Demarqui, e co-orientador, professor Thiago Santos, pela dedicação e orientação, essenciais a este trabalho.

Aos professores Hélio Migon e Rosangela Loschi, pela participação na banca de defesa e pelas sugestões.

Aos amigos de mestrado, em especial Edson, Erick, Guilherme e Natália pelo companheirismo ao longo do curso.

Ao CNPq, pelo apoio financeiro que me possibilitou dedicação integral a este trabalho.

Resumo

Métodos Bayesianos aplicados a séries temporais começaram a se destacar quando a classe de modelos lineares dinâmicos (MLD's) foi definida. A necessidade de trabalhar com séries temporais não Gaussianas atraiu muito interesse ao longo dos anos. A classe de modelos lineares generalizados dinâmicos (MLGD's) é uma extensão atrativa do MLD para observações na família exponencial e, ao mesmo tempo, corresponde a uma extensão do modelo linear generalizado, permitindo que os parâmetros variem no tempo. Recorrentemente, essas séries temporais podem ser afetadas por eventos externos que mudam a estrutura da série, resultando em problemas de ponto de mudança. O modelo partição produto (MPP) aparece como uma alternativa atrativa para identificar múltiplos pontos de mudança. Neste trabalho, propusemos estender a classe de modelos lineares generalizados dinâmicos utilizando o modelo partição produto para acomodar séries temporais na família exponencial com problemas de ponto de mudança. Nesse contexto, o MPP promove uma estrutura de agrupamento para os dados, e a inferência tradicional do MLGD é feita, mas ao invés de ser feita para cada observação, a inferência é feita por blocos de observações, implicando que observações no mesmo bloco terão um parâmetro de estado comum. A nova classe proposta é uma classe ainda mais ampla, uma vez que garante a flexibilidade do MLGD conjuntamente com a habilidade de detectar pontos de mudança através da metodologia do MPP. Nesse trabalho, analisamos bancos de dados reais objetivando ilustrar a aplicabilidade do modelo proposto.

Palavras-chave: Modelos dinâmicos, Modelo partição produto, Pontos de mudança, Família exponencial.

Abstract

Bayesian methods applied to time series have begun receiving highlighted when the class of dynamic linear models (DLM's) was defined. The need of working with non-normal time series have attracted a lot of interest during the years. The class of dynamic generalized linear models (DGLM's) is an atrractive extension of the DLM for observations in the exponential family and, in turn, corresponds to an extension of generalized linear model allowing the parameters to be time-varying. Recurrently, these time series may be affect by external events that can change the structure of the series, resulting in a change point problem. The product partition model (PPM) appears as an attractive alternative to identify multiple change points. In this work, we proposed to extend the class of dynamic generalized linear models by using the product partition model in order to accommodate time series in the exponential family within the change point problem. In this fashion, the PPM provides a blocking structure for the data, and the traditional inference of the DGLM is performed, but instead of making inference for each observation, the inference takes place by blocks of observations, implying that observations in the same block will have a common state parameter. The new class proposed is a wider class, since it guarantees all the flexibility of the DGLM along with the ability to detect change point problems through the PPM framework. In this work, we analyzed real data aiming to illustrate the usefulness of the proposed model.

Keywords: Dynamic models, Product partition model, Change points, Exponential family.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

A time series is a set of observations ordered in time wherein the observation's ordination implies in a correlation structure of the data which, in general, can not be neglected. We can enumerate many different fields where this type of data may occur such as economy, with weekly data related to stock exchange and interest rates, monthly sales and prices indexes, climatology, when observing daily temperature, monthly rainfall, tides records and an extensive list of areas as agriculture, business, geophysics, engineering, quality control and others, as cited by Shumway and Stoffer (2006). These series can be non-stationary and have non-observable components such as seasonality, cycle and trend. Besides, its state space may be continuous or discrete, allowing the series to assume either continuous or discrete distributions.

There is an extensive literature of time series analysis focused on developing appropriate models. A very popular class of models is the autoregressive integrated moving average (ARIMA), proposed by Box and Jenkins (1970). Bayesian methods applied to time series have begun receiving highlighted after the work of Harrison and Stevens (1976), where the class of dynamic linear models (DLM's) was defined. According to West et al. (1985) the fundamental idea of the DLM is that at any time t, the process under study is viewed in terms of a meaningful parameter $\boldsymbol{\theta}$, whose values are allowed to change as time passes. Although dealing only with Gaussian time series, Pole et al. (1994) emphasizes that the DLM class is very large and flexible indeed.

The need of working with non-normal and non-linear time series have attracted a lot of interest during the years and, consequently, extensions of the DLM were proposed. West (1981) proposed a robust sequential approximated Bayesian estimation which provided a procedure applicable to symmetric and uni-modal errors distributions. Meinhold and Singpurwalla (1989) provided a robustification of the Kalman-Filter. The solution proposed by Carlin et al. (1992) is a Monte Carlo approach by using the Gibbs sampling algorithm. These extensions were still restrictive in the sense that the distributions allowed were still related to the normal distribution.

Durbin and Koopman (2000) provided a treatment in both classical and Bayesian framework, by using importance sampling simulation and antithetic variables for the analysis of non-normal time series by using state-space models. In an alternative to escape from the Gaussian assumption Gamerman et al. (2013) introduced a wider class on non-Gaussian distributions, which retains analytical availability of the marginal likelihood function, and provided both Bayesian and classical approaches for the state-space models considering that class.

An attractive extension of the DLM, which shall be considered in this work, was defined by West et al. (1985), the class of dynamic generalized linear models (DGLM's), which is based on the theory of generalized linear models (GLM's), proposed by Nelder and Wedderburn (1972). Thus, the DGLM is an extension of DLM for observations in the exponential family and, in turn, corresponds to an extension of GLM allowing the parameters to be time-varying.

The DGLM class is rather broad since it treats any time series which distribution is a member of the exponential family without the stationary assumption, property which receives an special attention since it is usually an important characteristic for both modeling and forecasting, and others non-observable components may be include in the model through a system component.

According to Santos et al. (2010) time series may be affect by external events, called interventions, which effects on a given response variable are discussed by Box and Tiao (1975). These interventions can change the structure of the series, resulting in a change point problem. The change point problem has been intensively studied, since the subject found application in many different areas, implying in a really voluminous literature.

Usually, during the inference, its necessary to detect or estimate these change points. Perron (2005) provides a recent review of the methodological issues for models involving structural breaks. Among the several tools available in the literature to treat structural breaks, the product partition model (PPM) appears as an attractive alternative to identify multiple change points.

The PPM was first defined for general partitions by Hartigan (1990), which allowed the data to weight the partitions likely to hold and assumed that observations in different components of random partitions of the data are independent given the partition. In a more specific work Barry and Hartigan (1992) defined the PPM in terms of a partition of a set of observations into contiguous subsequences (or blocks), wherein the partition has a prior product distribution, and given the partition, the parameters in different blocks have independent prior distributions. Barry and Hartigan (1993) developed the PPM to identify multiple change points in normally distributed data, then proposed an approach to estimate change points and parameters through a Gibbs sampling scheme and provided a comparison with other approaches already used in the literature. Loschi and Cruz (2002) proposed a competitive and easy-to-implement modification of the Gibbs sampling scheme (Barry and Hartigan, 1993) for the estimation of normal means and variances.

Important extensions of the PPM began to appear in different areas. Loschi et al. (2005) used the PPM to estimate the mean and the volatility for time series data with structural changes that include both jumps and heteroskedasticity. Demarqui et al. (2012) applied the PPM in survival analysis to estimate the time grid in piecewise exponential model considering a class of correlated Gamma prior distributions for the failure rates, obtained via the dynamic generalized modeling. Muller et al. (2011) proposed the PPM in the presence of covariates which are included by a a new factor in the cohesion function. Monteiro et al. (2011) provided an extension of the PPM by assuming that observations within the same block have their distributions indexed by different parameters. In that approach, it was used a Gibbs distribution as a prior specification for the canonical parameter, implying that the parameters of the observations in the same block were also correlated. Ferreira et al. (2014) studied the identification of multiple change points using the PPM and including dependence between blocks through the prior distribution of the parameters. The reversible jump Markov chain Monte Carlo (MCMC) algorithm was used in that work to sample from the posterior distributions.

Fearnhead (2006) provided an approach related to work on product partition model and demonstrate how to perform direct simulation from the posterior distribution of a class of multiple changepoint models where the number of change points is unknown. This approach is useful even when the independence assumptions do not hold.

In this work, we propose to extend the class of dynamic generalized linear models by using the product partition model in order to accommodate time series in the exponential family within the change point problem. In this fashion, the PPM provides a blocking structure for the data, and the traditional inference of the DGLM is performed, but instead of making inference for each observation, the inference takes place by blocks of observations.

The inference of the proposed class will imply that observations in the same block will have a common state parameter. Although having a common state parameter, the observations in the same block are not independent and identically distributed, as in the usual PPM, since each distribution parameter depends on a set of covariates, that determines the hyperparameters of the prior distribution for the canonical parameter. Besides, the dynamic modeling implies that parameters in different blocks are correlated due to the evolution system.

The main advantage of working with the DGLM class is that it provides a closed expression for the one-step ahead forecast and, consequently, for the marginal likelihood. Besides, the evolution covariance matrix may be specified by the aid of the discount factor, and its estimation is no longer required. If the DGLM is specified via MCMC or particle filtering instead of using the linear Bayes approach and the aid of the discount factor, the marginal likelihood would have to be estimated numerically, and the computational time, which is already high due to the partition estimation, would be even higher.

The new class proposed, which we shall refer here to as the dynamic generalized linear model via product partition model (DGLM via PPM) is a wider class, since it guarantees all the flexibility of the DGLM along with the ability to detect change point problems through the PPM framework.

This work is organized as follows. In Chapter 2 we present the definition of the uniparametric exponential family of distributions and some important properties related to this family, and we briefly describe the generalized linear model theory. The dynamic models are introduced in Chapter 3, by first defining the dynamic linear model and its inference, then defining the dynamic generalized linear model and its inference. In Chapter 4 the product partition model introduced by Barry and Hartigan (1992) is reviewed, as well as a Gibbs sampling scheme used to obtain inferences about the partition. We extend the use of the product partition model in the class of dynamic generalized linear models in Chapter 5, by proposing the dynamic generalized linear model via product partition model. At last, in Chapter 6 we illustrate the usefulness of the proposed model by using two real time series data series, and comparing the inference with the traditional DGLM inference.

Chapter 2

Generalized Linear Models

The relationship between a response variable and a set of independent explanatory variables has been traditionally described by the classic linear model, defined as follows

$$Y_t = \boldsymbol{F}_t'\boldsymbol{\theta} + v_t, \tag{2.1}$$

where Y_t, $t = 1, ..., n$, is the response variable, $\boldsymbol{F}_t'$ is a $1 \times p$ vector of explanatory variables, $\boldsymbol{\theta}$ is a $p \times 1$ vector of unknown parameters and v_t is an error term.

The errors v_t are assumed to be independent and identically distributed as the Gaussian distribution, such that

$$v_t \overset{iid}{\sim} \mathrm{N}\,[0, V]. \tag{2.2}$$

As a consequence, the response variable is distributed such as the Gaussian distribution

$$Y_t|\boldsymbol{\theta} \sim \mathrm{N}\,[\boldsymbol{F}_t'\boldsymbol{\theta}, V]. \tag{2.3}$$

Despite being a very popular model, it finds some restriction when its assumptions are not satisfied. This may happen, for example, when it is not convenient to assume normality for the response variable, this being binary, as in the binomial distribution, being a count, as in the Poisson distribution, or being asymmetric, as in the gamma distribution. Some kind of transformation on the response variable may be done in order to solve the violation of normality assumption, but it will not be effective in most cases for many different reasons, specially, due to the data nature.

Some of the distributions used in statistics may be united in a family of parametric distributions, known as exponential family of distributions. This class of distributions holds some important properties in statistical analysis and enabled the development of the so-called generalized linear models.

2.1 Exponential Family of Distributions

Suppose that the probability density function of a random variable Y (if Y is continuous) or the probability function (if Y is discrete) can be written in the form

$$P(Y|\eta, V) \;=\; \exp\{V^{-1}[Y\eta \;-\; a(\eta)]\}b(Y, V), \tag{2.4}$$

where η is the canonical parameter, $V > 0$ is a dispersion parameter, and $a(.)$ e $b(.)$ are known functions. Besides, $\phi \;=\; V^{-1}$ is called the precision parameter. Then Y is said to belong to the uniparametric exponential family of distributions.

The function $a(.)$ is assumed twice differentiable in η. Then, it can be shown that

$$\mu \;=\; E[Y|\eta, V] \;=\; \frac{da(\eta)}{d\eta} \;=\; \dot{a}(\eta),$$

and

$$\text{Var}[Y|\eta, V] \;=\; \ddot{a}(\eta)/\phi.$$

In the following examples it is presented two distributions in the uniparametric exponential family.

Example 2.1 *The Poisson model*
Consider Y being distributed as the Poisson with mean μ, hence the pf is

$$\begin{aligned} p(Y|\mu) &= \frac{e^{\mu}\mu^{Y}}{Y!} \\ &= \frac{1}{Y!}\exp\{Y\ln\mu - \mu\},\;\; \mu > 0,\;\; Y = 0, 1, ... \end{aligned}$$

Note that, this distribution is a special case of the exponential family of distributions, where $V^{-1} = 1$, $b(Y, V) = \frac{1}{Y!}$, $\eta = \ln\mu$ and $a(\eta) = e^{\eta}$.

Example 2.2 *The Gaussian model*
Let Y be distributed as the Gaussian distribution with mean μ and variance V, hence the pdf is

$$f(Y|\mu, V) = \frac{1}{\sqrt{2\pi V}} \exp\left\{-\frac{(Y-\mu)^2}{2V}\right\}$$
$$= \frac{1}{\sqrt{2\pi V}} \exp\left\{-\frac{Y^2}{2V}\right\} \exp\left\{V^{-1}\left(Y\mu - \frac{\mu^2}{2}\right)\right\}, \ -\infty < Y, \ \mu < \infty, \ V > 0.$$

Therefore, this is a special case of the exponential family of distributions with $V^{-1} = \phi$, $b(Y, V) = \frac{1}{\sqrt{2\pi V}} \exp\left\{-\frac{Y^2}{2V}\right\}$, $\eta = \mu$ *and* $a(\eta) = \frac{\eta^2}{2}$.

Some of the most popular distributions such as the Binomial, Poisson, negative Binomial, in the discrete case, and the Gaussian, Gamma, Beta, in the continuous case, that belongs to exponential family are presented in Table 2.1, with the canonical parameter η, precision parameter V^{-1}, and the functions $a(.)$ and $b(.,)$ being identified (see McCullagh and Nelder (1989) for details).

Distribution	V^{-1}	η	$a(\eta)$	$b(Y, V)$
Poisson: $\mathrm{P}(\mu)$	1	$\ln \mu$	e^{η}	$\frac{1}{Y!}$
Binomial: $\mathrm{B}(m, \mu)$	$\frac{1}{m}$	$\ln\left(\frac{\mu}{1-\mu}\right)$	$\ln(1 + e^{\eta})$	$\binom{m}{\mu}$
Negative Binomial: $\mathrm{NB}(\pi, m)$	1	$\ln(1-\pi)$	$-m\ln(1 - e^{\eta})$	$\binom{Y+m-1}{m-1}$
Normal: $\mathrm{N}(\mu, V)$	V	μ	$\frac{\eta^2}{2}$	$\frac{1}{\sqrt{2\pi V}} \exp\left(-\frac{Y^2}{2V}\right)$
Gamma: $\mathrm{G}(\alpha, \beta)$	1	$-\beta$	$-\alpha \ln \beta$	$\frac{Y^{\alpha-1}}{\Gamma(\alpha)}$
Beta: $\mathrm{B}(\lambda)$	1	$-\lambda$	$-\ln(-\eta)$	$1/Y$

Table 2.1: Some distributions in the uniparametric exponential family

The exponential family of distributions holds some important properties in statistical analysis. Particularly, in the Bayesian framework, the conjugate analysis property. Using this proper facilitates the analysis in the sense that the prior and the posterior distributions belongs to the same class of distributions, and the updating involves only changes in the hyperparameters. This implies that required distributions can be analytically calculated as shown below.

Assuming that Y is a random variable member of the exponential family, the prior density from the conjugate family has the form

$$p(\eta|r,s) \;=\; c(r,s)\exp[r\eta - sa(\eta)], \tag{2.5}$$

where η is the canonical parameter and $a(.)$ is a known function, both obtained from (2.4), r and s are known quantities, and $c(r,s)$ is a normalizing constant that can be found by

$$c(r,s) \;=\; \left(\int \exp\{r\eta - sa(\eta)\}\; d\eta\right)^{-1}. \tag{2.6}$$

Once the prior distribution has been established, the predictive distribution may be calculated as follows

$$p(Y^*|Y) \;=\; \frac{c(r,s)b(Y^*,V)}{c(r+\phi Y,\;\; s+\phi)}, \tag{2.7}$$

where $b(.)$ is a known function obtained from (2.4) and $\phi = V^{-1}$.

Moreover the posterior distribution can be obtained by

$$p(\eta|Y) \;=\; c(r+\phi Y,\;\; s+\phi)\exp[(r+\phi Y)\eta - (s+\phi)a(\eta)]. \tag{2.8}$$

Example 2.3 *The Poisson Model*
In the Poisson model, the conjugate prior for $\eta = \ln\mu$ is given by

$$p(\eta|r,s) \;=\; c(r,s)\exp\{r\eta - se^{\eta}\},$$

where the normalizing constant is

$$c(r,s) = \left(\int \exp\{r\eta - se^{\eta}\}\; d\eta\right)^{-1} \;=\; \frac{s^r}{\Gamma(r)}.$$

That is, $\eta \;\sim\;$ Log-Gamma(r,s).

2.2 The Generalized Linear Models

Nelder and Wedderburn (1972) proposed an attractive extension of the linear regression models, the class of Generalized Linear Models (GLMs), this being a form to explain the relationship between a response variable, now, relaxed the normality assumption, called the random component, and the linear predictor, through a link function. The components of the GLM are detailed as follows.

- The random component of the model is the response variable, Y, whose distribution must be a member of the exponential family in (2.4), relaxing the assumption of normality in the regression linear model, and including others popular distributions such as the Poisson, binomial, gamma, among others. That way we can model response variables in the form of proportions, counts or rates, and others.

- The linear predictor is the linear function of the explanatory variables, being similar to the structure of the linear model, that is

$$\lambda_t = \boldsymbol{F}_t'\boldsymbol{\theta}. \tag{2.9}$$

- The link function is a continuous, monotonic and differentiable function, denoted by g(.), responsible to link the mean of the random component to the linear predictor:

$$\lambda_t = g(\mu_t). \tag{2.10}$$

  The choice of the link function is arbitrary. There are many commonly used depending on the data, as enumerated by McCullagh and Nelder (1989).

Note that we are not modeling the mean μ_t as before, but a function of it.

Example 2.4 *The Poisson model*
Suppose a generalized linear model in which the random component Y has a Poisson distribution. The Poisson distribution appears, frequently, associated to count data, and plays an important role in data analysis. One of the most important cases of GLM is defined as follow

$$\log(\mu_t) = \lambda_t = \boldsymbol{F}_t'\boldsymbol{\theta}$$

This models is known as the log-linear model and has a large application in contingency table data.

Although the generalized linear model represents a large advance in the statistics modeling field, it finds some restriction in dealing with response variables that are time dependent, as in the case of time series data, since the assumption of independence can not be relaxed in this class of models.

Chapter 3

Dynamic Linear Models

The linear models discussed in Chapter 2 are static models, in the sense that parameters values are fixed across all experiment, and assume that the ordination of data is irrelevant. According to Pole et al. (1994) the order property is crucial when dealing with time series data and, besides that, time itself involves circumstantial changes that alter the structure of the series, bringing the need to work with dynamic models. The dynamic models are formulated such that changes during time are allowed in the parameters. The main idea is to build a linear model considering the parameters are now time-varying and stochastically related through an evolution equation.

3.1 The Normal Dynamic Linear Model

The normal dynamic linear model, refer only as DLM, is one of the most popular subclass of dynamic linear models given its large applicability when dealing with real data. Its analytical structure is presented as follows.

For time t defines:

- $\boldsymbol{F}_t$ is a known p-dimensional vector, being the design vector of the values of independent variables (possibly time-varying);
- $\boldsymbol{\theta}_t$ is the state, or system p-dimensional column vector;
- v_t is the observational error, having zero mean and a known variance V_t;
- $\boldsymbol{G}_t$ is the known evolution, system, transfer or state matrix $(p \times p)$;
- $\boldsymbol{w}_t$ is the evolution, or system error, having zero mean and a $(p \times p)$ known evolution variance matrix $\boldsymbol{W}_t$.

The formal definition of the normal DLM is given as follows.

Definition 3.1 *The general univariate dynamic linear models is written as:*

$$\text{Observation equation:} \quad Y_t = \boldsymbol{F}_t'\boldsymbol{\theta}_t + v_t, \qquad v_t \sim N[0, V_t],$$

$$\text{Evolution equation:} \quad \boldsymbol{\theta}_t = \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + w_t, \qquad w_t \sim N[0, \boldsymbol{W}_t],$$

$$\text{Initial information:} \quad \boldsymbol{\theta}_0|\boldsymbol{D}_0 \sim N[\boldsymbol{m}_0, \boldsymbol{C}_0],$$

for some prior moments $\boldsymbol{m}_0$ *and* $\boldsymbol{C}_0$*. The observational and evolution errors are assumed to be internally and mutually independent.*

West and Harrison (1997) suggest a Bayesian analysis for DLM, established through the mechanism presented in Theorem 3.1 and emphasizes that the central characteristic of this model is that at any time, existing information about the system is represented and sufficiently summarized by the posterior distribution for the current state vector.

Theorem 3.1 *In the univariate DLM, one-step forecast and posterior distributions are given, for each* t*, as follows:*

1. *Posterior at* $t-1$*:*

$$(\boldsymbol{\theta}_{t-1}|\boldsymbol{D}_{t-1}) \sim \mathrm{N}[\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1}].$$

   where $\boldsymbol{m}_{t-1}$ *and* $\boldsymbol{C}_{t-1}$ *are the posterior moments at time* $t-1$*.*

2. *Prior at* t*:*

$$(\boldsymbol{\theta}_t|D_{t-1}) \sim \mathrm{N}[a_t, R_t],$$

   where

$$\boldsymbol{a}_t = \boldsymbol{G}_t\boldsymbol{m}_{t-1} \quad \text{and} \quad \boldsymbol{R}_t = \boldsymbol{G}_t\boldsymbol{C}_{t-1}\boldsymbol{G}_t' + \boldsymbol{W}_t$$

3. *One-step forecast:*

$$(Y_t|\boldsymbol{D}_{t-1}) \sim \mathrm{N}[f_t, q_t],$$

   where

$$f_t = \boldsymbol{F}_t'\boldsymbol{a}_t \quad \text{and} \quad q_t = \boldsymbol{F}_t'\boldsymbol{R}_t\boldsymbol{F}_t + V_t.$$

4. *Posterior at* t*:*

$$(\boldsymbol{\theta}_t|\boldsymbol{D}_t) \sim \mathrm{N}[\boldsymbol{m}_t, \boldsymbol{C}_t],$$

with

$$\boldsymbol{m}_t = \boldsymbol{a}_t + A_t e_t \quad \text{and} \quad \boldsymbol{C}_t = \boldsymbol{R}_t - A_t q_t A_t',$$

where

$$A_t = \boldsymbol{R}_t \boldsymbol{F}_t q_t^{-1} \quad \text{and} \quad e_t = Y_t - f_t.$$

The resulting algorithm of Theorem 3.1 is known as Kalman filter (Migon et al. (2005)) and its proof is obtained by induction using the multivariate normal distribution theory (see West and Harrison (1997)).

There is, commonly, a great interest in looking back in time and make inference about past state vectors, that is, a interest in the retrospective marginal distribution $(\boldsymbol{\theta}_{t-k}|\boldsymbol{D}_t)$. This distribution is called the k-step filtered distribution for the state vector at time t, and may be derived using the Bayes Theorem as presented by West and Harrison (1997) in Theorem 3.2.

Theorem 3.2 *In the univariate DLM, for all* t*, define*

$$\boldsymbol{B}_t \;=\; \boldsymbol{C}_t \boldsymbol{G}_{t+1}' \boldsymbol{R}_{t+1}^{-1}.$$

For all k*,* $(1 \leq k < t)$*, the filtered marginal distributions are*

$$(\boldsymbol{\theta}_{t-k}|\boldsymbol{D}_t) \;\sim\; \mathrm{N}[\boldsymbol{a}_t(-k), \boldsymbol{R}_t(-k)],$$

where

$$\boldsymbol{a}_t(-k) \;=\; \boldsymbol{m}_{t-k} \;+\; \boldsymbol{B}_{t-k}[\boldsymbol{a}_t(-k+1) - \boldsymbol{a}_{t-k+1}]$$

and

$$\boldsymbol{R}_t(-k) \;=\; \boldsymbol{C}_{t-k} \;+\; \boldsymbol{B}_{t-k}[\boldsymbol{R}_t(-k+1) - \boldsymbol{R}_{t-k+1}]\boldsymbol{B}_{t-k}',$$

with starting values

$$\boldsymbol{a}_t(0) = \boldsymbol{m}_t \quad \text{and} \quad \boldsymbol{R}_t(0) = \boldsymbol{C}_t,$$

and,

$$\boldsymbol{a}_{t-k}(1) = \boldsymbol{a}_{t-k+1} \quad \text{and} \quad \boldsymbol{R}_{t-k}(1) = \boldsymbol{R}_{t-k+1}.$$

The performance of the DLM inference depends heavily of the specification of the evolution variance W_t. An alternative is the discount factor as an aid for choosing W_t. By definition, we have that

$$\mathrm{Var}(\boldsymbol{\theta}_t|\boldsymbol{D}_{t-1}) \;=\; \mathrm{Var}(G_t\;\boldsymbol{\theta}_{t-1} + w_t|\boldsymbol{D}_{t-1})$$

$$= G_t' C_{t-1} G_t + W_t.$$

From that, we may observe that W_t is a fixed proportion of C_{t-1}, that is, the addition of the error w_t leads to an incrementation of the uncertain of C_{t-1}. Then, it is natural to think about a value δ, $0 < \delta \leq 1$, known as discount factor, so we can rewrite the following variance as

$$\text{Var}(\boldsymbol{\theta}_t|\boldsymbol{D}_{t-1}) = \frac{1}{\delta} \text{Var}(\boldsymbol{\theta}_{t-1}|\boldsymbol{D}_{t-1}),$$

so that, the specification of W_t with the aid of the discount factor will be given by

$$W_t = G_t' C_{t-1} G_t (1-\delta)/\delta.$$

The discount factor may be interpreted as the amount of information which is allowed to pass from time $t-1$ to time t. As closer to 1 is its value, more information is allowed to pass.

3.1.1 Linear Bayes' Optimality

Section 3.1 presented in Theorem 3.1 the dynamic linear model updating which is provided by using the multivariate normal distribution theory as a consequence of the normality assumption for the observational and evolution errors. West and Harrison (1997) pointed out that the updating for $\boldsymbol{m}_t$ and $\boldsymbol{C}_t$ may also be derived using approaches that do not invoke the normality assumption due to strong optimality properties that are derived when the distributions are only specified in terms of means and variance.

Replacing the normality assumption of the observational and evolutions error by the first and second-order moment the DLM equations become

$$\begin{aligned} Y_t &= \boldsymbol{F}_t' \boldsymbol{\theta}_t + v_t, \qquad v_t \sim [0, V_t], \\ \boldsymbol{\theta}_t &= \boldsymbol{G}_t \boldsymbol{\theta}_{t-1} + w_t, \qquad w_t \sim [0, \boldsymbol{W}_t], \\ &\boldsymbol{\theta}_0 | \boldsymbol{D}_0 \sim [\boldsymbol{m}_0, \boldsymbol{C}_0]. \end{aligned}$$

Suppose that the joint distribution of $\boldsymbol{\theta}_t$ and Y_t is partially specified in terms of the mean and the variance matrix, that is,

$$\begin{pmatrix} \boldsymbol{\theta}_t \\ Y_t \end{pmatrix} \sim \left[\begin{pmatrix} \boldsymbol{a}_t \\ \boldsymbol{f}_t \end{pmatrix}, \begin{pmatrix} \boldsymbol{R}_t & \boldsymbol{A}_t \boldsymbol{Q}_t \\ \boldsymbol{Q}_t \boldsymbol{A}_t' & \boldsymbol{Q}_t \end{pmatrix} \right] \tag{3.1}$$

In the remaining section we describe the decision theoretically based linear Bayes' estimation procedure in the the DLM framework. According to West and Harrison (1997) the idea of this procedure is to model a function of the parameter $\boldsymbol{\theta}_t$ and the observation Y_t, $\phi(\boldsymbol{\theta}_t, Y_t)$, independent of Y_t, thus observed the value $Y_t = y$, the posterior distribution of the function $\phi(\boldsymbol{\theta}_t, y)$ is identical to the prior distribution $\phi(\boldsymbol{\theta}_t, Y_t)$.

Let $\boldsymbol{d}_t$ be any estimate of $\boldsymbol{\theta}$, and the accuracy in estimation is measured by a loss function given by

$$L(\boldsymbol{\theta}_t, \boldsymbol{d}_t) \;=\; (\boldsymbol{\theta}_t - \boldsymbol{d}_t)'(\boldsymbol{\theta}_t - \boldsymbol{d}_t) \;=\; tr(\boldsymbol{\theta}_t - \boldsymbol{d}_t)(\boldsymbol{\theta}_t - \boldsymbol{d}_t)'.$$

where $tr(A)$ denotes the trace of matrix A.

Hence, the estimate $\boldsymbol{d}_t = \boldsymbol{m}_t = \boldsymbol{m}_t(Y_t)$ is optimal with respect to the loss function if the function $r(\boldsymbol{d}_t) = E[L(\boldsymbol{\theta}_t, \boldsymbol{d}_t)|Y_t]$ is minimized as function of $\boldsymbol{d}_t$ when $\boldsymbol{d}_t = \boldsymbol{m}_t$.

Definition 3.2 *A linear Bayes' estimation (LBE) of $\boldsymbol{\theta}$ is a linear form*

$$\boldsymbol{d}(Y_t) = \boldsymbol{h}_t + \boldsymbol{H}_t Y_t,$$

for some $n \times 1$ vector $\boldsymbol{h}_t$ and $n \times p$ matrix $\boldsymbol{H}_t$, that is optimal in the sense of minimizing the overall risk

$$r(\boldsymbol{d}_t) = trace\ E[(\boldsymbol{\theta}_t - \boldsymbol{d}_t)(\boldsymbol{\theta}_t - \boldsymbol{d}_t)'].$$

According to West and Harrison (1997) the above definition provide as main result the following theorem.

Theorem 3.3 *In the above framework, the unique LBE of is*

$$\boldsymbol{m}_t = \boldsymbol{a}_t + \boldsymbol{A}_t(Y_t - \boldsymbol{f}_t.)$$

The associated risk matrix is given by

$$\boldsymbol{C}_t = \boldsymbol{R}_t - \boldsymbol{A}_t \boldsymbol{Q}_t \boldsymbol{A}_t',$$

and is the value of $E[(\boldsymbol{\theta}_t - m_t)(\boldsymbol{\theta}_t - m_t)']$, so that the minimum risk is simply $r(m_t) = trace(\boldsymbol{C}_t)$.

3.2 The Dynamic Generalized Linear Model

Despite being a very useful and popular model in time series analysis, the dynamic linear model finds some restriction when it is not reasonable to suppose that the time series Y_t

is normally distributed. West et al. (1985) proposed a more general model, in the class of dynamic models, the dynamic generalized linear model (DGLM). This new subclass is a generalization of the DLM, in the sense that any distribution which have the form of the uniparametric exponential family form presented in Chapter 2.1 is accept to model the time series in study. There is a reasonable gain in this new model, given that relaxed the normality assumption, the model is able to work with asymmetric distributions such as the exponential and gamma, or even discrete distributions such as the Poisson, binomial. Besides, the DGLM is a generalization of the GLM, but now, allowing the parameters to be time varying.

Suppose that the time series Y_t is generated from a distribution member of the uniparametric exponential family, that is

$$P(Y_t|\eta_t, V_t) \;=\; \exp\{V_t^{-1}[Y_t\eta_t \;-\; a(\eta_t)]\}b(Y_t, V_t), \tag{3.2}$$

where η_t is the canonical parameter, $V_t^{-1} > 0$ is a precision parameter, and $a(.)$ e $b(.)$ are known functions.

The idea of generalized linear modeling is to use a non-linear function $g(.)$, known as link function, which maps $\mu_t = E[Y_t|\eta_t, V_t]$ to a linear predictor λ_t.

We have the formal definition of DGLM presented bellow.

Definition 3.3 *The dynamic generalized linear model (DGLM) for the time series Y_t, $t = 1, ..., n$, is defined by the following components:*
Observational model:

$$p(Y_t|\eta_t) \qquad e \qquad g(\eta_t) = \lambda_t = \boldsymbol{F}_t' \,\boldsymbol{\theta}_t;$$

Evolution equation:

$$\theta_t \;=\; \boldsymbol{G}_t \,\boldsymbol{\theta}_{t-1} \;+\; \boldsymbol{w}_t \;, \qquad \boldsymbol{w}_t \;\sim\; [0, \boldsymbol{W}_t]; \tag{3.3}$$

where,

- $\boldsymbol{\theta}_t$ *is an p-dimensional state vector at time t;*
- $\boldsymbol{F}_t$ *is a known p-dimensional regression vector;*
- $\boldsymbol{G}_t$ *is a known $p \times p$ evolution matrix;*
- $\boldsymbol{w}_t$ *is an p-vector of evolution errors, where $\boldsymbol{w}_t \;\sim\; [0, \boldsymbol{W}_\mathbf{t}]$;*

- $\lambda_t = \boldsymbol{F}_t' \, \boldsymbol{\theta}_t$ *is a linear function of the state vector parameters;*

- $g(\eta_t)$ *is a known, continuous and monotonic function mapping* η_t *to the real line.*

As the DLM is a particular case of the DGLM, we can motivate the components of the DGLM analysis by, first, providing a reformulation of the Theorem 3.1 used in the DLM analysis.

Notice, at first, that the observational component of the DLM is given by

$$
\begin{aligned}
(Y_t|\eta_t) \;&\sim\; N[\mu_t, V_t], \\
\mu_t \;=\; \eta_t \;&=\; \lambda_t \;=\; \boldsymbol{F}_t'\boldsymbol{\theta}_t.
\end{aligned}
$$

The model specification will be completed by the posterior distribution at $t-1$, as usual, that is

$$(\boldsymbol{\theta}_{t-1}|\boldsymbol{D}_{t-1}) \;\sim\; \mathrm{N}[\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1}].$$

Using the evolution equation in (3.3), the prior distribution at t is obtained

$$(\boldsymbol{\theta}_t|\boldsymbol{D}_{t-1}) \;\sim\; \mathrm{N}[\boldsymbol{a}_t, \boldsymbol{R}_t],$$

where

$$\boldsymbol{a}_t = \boldsymbol{G}_t\boldsymbol{m}_{t-1} \quad \text{and} \quad \boldsymbol{R}_t = \boldsymbol{G}_t\boldsymbol{C}_{t-1}\boldsymbol{G}_t' + \boldsymbol{W}_t.$$

That way, we can proceed our analysis, in an alternative form, through the steps presented bellow.

Step 1: Joint prior distribution for μ_t and $\boldsymbol{\theta}_t$

Notice that $\lambda_t = \mu_t$, that is, μ_t is a function of the vector $\boldsymbol{\theta}_t$. So, under the prior distribution for $\boldsymbol{\theta}_t$ defined previously, the joint prior distribution for μ_t and $\boldsymbol{\theta}_t$ is given by

$$\left(\begin{array}{c|c} \mu_t & \\ \boldsymbol{\theta}_t & \boldsymbol{D}_{t-1} \end{array}\right) \sim \mathrm{N}\left[\begin{pmatrix} f_t \\ a_t \end{pmatrix}, \begin{pmatrix} q_t & \boldsymbol{F}_t'\boldsymbol{R}_t \\ \boldsymbol{R}_t\boldsymbol{F}_t & \boldsymbol{R}_t \end{pmatrix}\right] \tag{3.4}$$

where

$$f_t = \boldsymbol{F}_t'\boldsymbol{a}_t \quad \text{and} \quad q_t = \boldsymbol{F}_t'\boldsymbol{R}_t\boldsymbol{F}_t.$$

Step 2: One-step ahead forecasting

The only relevant information to predict Y_t is totally summarized by the marginal prior $(\mu_t|\boldsymbol{D}_{t-1})$ given that the distribution of Y_t depends of $\boldsymbol{\theta}_t$ only through the quantity μ_t. Hence, the one-step ahead forecasting distribution is written as

$$p(Y_t|\boldsymbol{D}_{t-1}) \;=\; \int p(Y_t|\mu_t)p(\mu_t|\boldsymbol{D}_{t-1})d\mu_t. \tag{3.5}$$

So, it may be verify that $(Y_t|\boldsymbol{D}_{t-1})$ is normally distributed as

$$(Y_t|D_{t-1}) \;\sim\; \mathrm{N}[f_t, Q_t], \quad \text{where} \;\; Q_t = q_t + V_t. \tag{3.6}$$

Step 3: Updating for μ_t

Observed Y_t, the posterior distribution for μ_t is given by

$$(\mu_t|\boldsymbol{D}_t) \;\sim\; \mathrm{N}[f_t^*, q_t^*], \tag{3.7}$$

where,

$$f_t^* = f_t + (q_t/Q_t)(Y_t - f_t) \quad \text{ e } \quad q_t^* = q_t - q_t^2/Q_t,$$

since,

$$p(\mu_t|\boldsymbol{D}_t) \propto p(\mu_t|\boldsymbol{D}_{t-1})p(Y_t|\mu_t). \tag{3.8}$$

Step 4: Conditional structure for $(\boldsymbol{\theta}_t|\mu_t, D_{t-1})$

The main purpose is always calculate the posterior distribution of $\boldsymbol{\theta}_t$. That may be done from the joint posterior distribution of μ_t and $\boldsymbol{\theta}_t$. That is

$$p(\mu_t, \boldsymbol{\theta}_t|\boldsymbol{D}_t) \;\propto\; p(\boldsymbol{\theta}_t|\mu_t, \boldsymbol{D}_{t-1})\; p(\mu_t|\boldsymbol{D}_t). \tag{3.9}$$

Wherein,

$$(\boldsymbol{\theta}_t|\mu_t, \boldsymbol{D}_{t-1}) \;\sim\; \mathrm{N}[\boldsymbol{a}_t + \boldsymbol{R}_t\boldsymbol{F}_t(\mu_t - f_t)/q_t,\; \boldsymbol{R}_t - \boldsymbol{R}_t\boldsymbol{F}_t\boldsymbol{F}'_t\boldsymbol{R}_t/q_t].$$

Hence, our posterior distribution of interest may be obtained as

$$p(\boldsymbol{\theta}_t|\boldsymbol{D}_t) \;=\; \int p(\boldsymbol{\theta}_t|\mu_t, \boldsymbol{D}_{t-1})\; p(\mu_t|\boldsymbol{D}_t)d\mu_t. \tag{3.10}$$

Step 5: Updating for $\boldsymbol{\theta}_t$

Since all the components of equation (3.10) are normally distributed, so is the posterior distribution $p(\boldsymbol{\theta}_t|D_t)$, and it can be completely characterized by its mean and the variance matrix. The mean is given by

$$\boldsymbol{m}_t \;=\; E[\boldsymbol{\theta}_t|\boldsymbol{D}_t] \;=\; E[E\{\boldsymbol{\theta}_t|\mu_t, \boldsymbol{D}_{t-1}\}|\boldsymbol{D}_t],$$

so,

$$\boldsymbol{m}_t \;=\; \boldsymbol{a}_t + \boldsymbol{R}_t\boldsymbol{F}_t(f_t^* - f_t)/q_t. \tag{3.11}$$

The variance matrix can be expressed through

$$\boldsymbol{C}_t \;=\; V[\boldsymbol{\theta}_t|\boldsymbol{D}_t] \quad = V[E\{\boldsymbol{\theta}_t|\mu_t, \boldsymbol{D}_{t-1}\}|\boldsymbol{D}_t] + E[V\{\boldsymbol{\theta}_t|\mu_t, \boldsymbol{D}_{t-1}\}|\boldsymbol{D}_t],$$

hence,

$$\boldsymbol{C}_t \;=\; \boldsymbol{R}_t - \boldsymbol{R}_t\boldsymbol{F}_t\boldsymbol{F}_t'\boldsymbol{R}_t(1 - q_t^*/q_t)/q_t. \tag{3.12}$$

3.2.1 DGLM updating

Dropping the normality assumption, West et al. (1985) proposed an approximate mechanism to the DGLM analysis based on the steps developed in the DLM reformulated analysis. One crucial difference here is that the required distributions are specified only in terms of their moments. Then, initially, the model specification is completed by the posterior moments at $t-1$, that is

$$(\boldsymbol{\theta}_{t-1}|\boldsymbol{D}_{t-1}) \;\sim\; [\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1}].$$

Through the evolution equation in (3.3), the prior distribution at t is

$$(\boldsymbol{\theta}_t|\boldsymbol{D}_{t-1}) \;\sim\; [\boldsymbol{a}_t, \boldsymbol{R}_t],$$

where

$$\boldsymbol{a}_t = \boldsymbol{G}_t\boldsymbol{m}_{t-1} \quad \text{and} \quad \boldsymbol{R}_t = \boldsymbol{G}_t\boldsymbol{C}_{t-1}\boldsymbol{G}_t' + \boldsymbol{W}_t.$$

With these, the previously steps may be rewritten as described below.

Step 1: Joint prior distribution for λ_t and $\boldsymbol{\theta}_t$

Notice that λ_t is a function of the vector $\boldsymbol{\theta}_t$. So, under the prior distribution for $\boldsymbol{\theta}_t$ defined previously, the joint prior distribution for λ_t and $\boldsymbol{\theta}_t$ is partially specified in terms of moments

$$\left(\begin{array}{c} \lambda_t \\ \boldsymbol{\theta}_t \end{array} \Big| \ \boldsymbol{D}_{t-1} \right) \sim \left[\left(\begin{array}{c} f_t \\ a_t \end{array} \right) , \left(\begin{array}{cc} q_t & \boldsymbol{F}'_t \boldsymbol{R}_t \\ \boldsymbol{R}_t \boldsymbol{F}_t & \boldsymbol{R}_t \end{array} \right) \right], \qquad (3.13)$$

where

$$f_t = \boldsymbol{F}'_t \boldsymbol{a}_t \quad \text{and} \quad q_t = \boldsymbol{F}'_t \boldsymbol{R}_t \boldsymbol{F}_t.$$

Step 2: One-step ahead forecast

The only relevant information to predict Y_t is totally summarized by the marginal prior $(\eta_t|\boldsymbol{D}_{t-1})$ given that the distribution of Y_t depends of $\boldsymbol{\theta}_t$ only through the quantity η_t. Hence, the one-step ahead forecasting distribution is written as

$$p(Y_t|\boldsymbol{D}_{t-1}) \ = \ \int \ p(Y_t|\eta_t) p(\eta_t|\boldsymbol{D}_{t-1}) d\eta_t. \qquad (3.14)$$

Assuming the conjugate form for the prior distribution of η_t we have

$$p(\eta_t|\boldsymbol{D}_{t-1}) \ = \ c(r_t, s_t) \exp[r_t \eta_t \ - \ s_t a(\eta_t)]. \qquad (3.15)$$

A consistent choice have to be done for the defining parameters r_t and s_t, since $\lambda_t = g(\eta_t)$, the following equations must be satisfied

$$E[g(\eta_t)|\boldsymbol{D}_{t-1}] = f_t \quad \text{and} \quad Var[g(\eta_t)|\boldsymbol{D}_{t-1}] = q_t. \qquad (3.16)$$

Hence, the one-step ahead forecast distribution is given by

$$p(Y_t|\boldsymbol{D}_{t-1}) \ = \ \frac{c(r_t, s_t) b(Y_t, V_t)}{c(r_t + \phi_t Y_t, s_t + \phi_t)}. \qquad (3.17)$$

Step 3: Updating for η_t

The posterior distribution for η_t is given by

$$p(\eta_t|\boldsymbol{D}_t) \;=\; c(r_t+\phi_t Y_t, s_t+\phi_t)\exp[(r_t+\phi_t Y_t)\eta_t \;-\; (s_t+\phi_t)a(\eta_t)]. \tag{3.18}$$

By analogy we have

$$E[g(\eta_t)|\boldsymbol{D}_t] = f_t^* \quad \text{and} \quad Var[g(\eta_t)|\boldsymbol{D}_t] = q_t^*. \tag{3.19}$$

Step 4a: Conditional structure for $(\boldsymbol{\theta}_t|\lambda_t, \boldsymbol{D}_{t-1})$

As in the DLM the main purpose is to obtain posterior information from $\boldsymbol{\theta}_t$. That may be done trough the following joint distribution

$$p(\lambda_t, \boldsymbol{\theta}_t|\boldsymbol{D}_t) \;\propto\; p(\boldsymbol{\theta}_t|\lambda_t, \boldsymbol{D}_{t-1})\; p(\lambda_t|\boldsymbol{D}_t), \tag{3.20}$$

given that the posterior distribution will be obtained by

$$p(\boldsymbol{\theta}_t|\boldsymbol{D}_t) \;=\; \int p(\boldsymbol{\theta}_t|\lambda_t, \boldsymbol{D}_{t-1})\; p(\lambda_t|\boldsymbol{D}_t) d\lambda_t. \tag{3.21}$$

Step 4b: Linear Bayes' estimation of moments of $(\boldsymbol{\theta}_t|\lambda_t, \boldsymbol{D}_{t-1})$

Within the class of linear functions of λ_t, and subject only to the prior information given in Equation (3.13), the linear Bayes estimate for the moments of $(\boldsymbol{\theta}_t|\lambda_t, \boldsymbol{D}_{t-1})$ is given by

$$\hat{E}[\boldsymbol{\theta}_t|\lambda_t, \boldsymbol{D}_{t-1}] \;=\; \boldsymbol{a}_t + \boldsymbol{R}_t\boldsymbol{F}_t(\lambda_t - f_t)/q_t, \tag{3.22}$$

and

$$\hat{\text{Var}}[\boldsymbol{\theta}_t|\lambda_t, \boldsymbol{D}_{t-1}] \;=\; \boldsymbol{R}_t - \boldsymbol{R}_t\boldsymbol{F}_t\boldsymbol{F}_t'\boldsymbol{R}_t/q_t. \tag{3.23}$$

Step 5: Updating for $\boldsymbol{\theta}_t$

Analogously to the DLM we express the posterior moments. The posterior mean can be written as

$$\boldsymbol{m}_t \;=\; E[\boldsymbol{\theta}_t|\boldsymbol{D}_t] \;=\; E[E\{\boldsymbol{\theta}_t|\lambda_t, \boldsymbol{D}_{t-1}\}|\boldsymbol{D}_t],$$

so,

$$\boldsymbol{m}_t \;=\; \boldsymbol{a}_t + \boldsymbol{R}_t\boldsymbol{F}_t(f_t^* - f_t)/q_t. \tag{3.24}$$

The posterior variance matrix can be expressed as

$$\boldsymbol{C}_t \;=\; V[\boldsymbol{\theta}_t|\boldsymbol{D}_t] \quad = V[E\{\boldsymbol{\theta}_t|\lambda_t, \boldsymbol{D}_{t-1}\}|\boldsymbol{D}_t] + E[V\{\boldsymbol{\theta}_t|\lambda_t, \boldsymbol{D}_{t-1}\}|\boldsymbol{D}_t],$$

so,

$$\boldsymbol{C}_t \;=\; \boldsymbol{R}_t - \boldsymbol{R}_t\boldsymbol{F}_t\boldsymbol{F}_t'\boldsymbol{R}_t(1 - q_t^*/q_t)/q_t. \tag{3.25}$$

That way,

$$(\boldsymbol{\theta}_t|\boldsymbol{D}_t) \;\sim\; [\boldsymbol{m}_t, \boldsymbol{C}_t] \tag{3.26}$$

As in the DLM, we can use Theorem 3.2, replacing the Gaussian distribution by just the required moments needed, to obtain the k-step filtered moments for the state vector at time t. The specification of the evolution variance W_t may be done, analogously to the DLM, by using the discount factor.

Example 3.1 *The Poisson model*
Suppose $Y_t \sim$ Poisson(μ_t) a time series associated to counts. Given $\mu_t > 0$, $Y_t|\mu_t$ has a probability function member of the exponential family. Assuming the conjugate prior for μ_t, we have

$$(\mu_t|\boldsymbol{D}_{t-1}) \;\sim\; Gama(r_t, s_t),$$

that is $\eta_t \sim$ Log-gamma(r_t, s_t), where $\eta_t = \ln \mu_t$.

From the DGLM definition it is given that

$$g(\eta_t) \;=\; \lambda_t \;=\; \boldsymbol{F}_t'\boldsymbol{\theta}_t,$$

hence, assuming $g(\eta_t) = \eta_t$ implies that

$$E(g(\eta_t)|\boldsymbol{D}_{t-1}) \;=\; E(\ln \mu_t|\boldsymbol{D}_{t-1}) \;=\; f_t,$$

$$\mathrm{Var}(g(\eta_t)|\boldsymbol{D}_{t-1}) \;=\; \mathrm{Var}(\ln \mu_t|\boldsymbol{D}_{t-1}) \;=\; q_t.$$

It can be shown that

$$E(\ln \mu_t) \;=\; \psi(r_t) - \ln(s_t),$$
$$\text{Var}(\ln \mu_t) \;=\; \psi_1(r_t),$$

where ψ is the digamma function and ψ_1 is the trigamma function, that can be approximated by the following functions

$$\psi(x) \;\approx\; ln(x) + \frac{1}{2x} \quad \text{and} \quad \psi_1(x) \;\approx\; \frac{1}{x}\left(1 - \frac{1}{2x}\right).$$

For larger values of x we can use the following approximations:

$$\psi(x) \;\approx\; ln(x) \quad \text{and} \quad \psi_1(x) \;\approx\; x^{-1}.$$

That way,

$$f_t \;=\; \psi(r_t) - \ln(s_t) \;\approx\; \ln(r_t) - \ln(s_t),$$
$$q_t \;=\; \psi_1(r_t) \;\approx\; \frac{1}{r_t}.$$

Therefore,

$$s_t = \frac{\exp(-f_t)}{q_t},$$
$$r_t = \frac{1}{q_t}.$$

Analogously ,

$$E(\eta_t|D_t) \;=\; f_t^* \;\approx\; \ln\left(\frac{r_t + Y_t}{s_t + 1}\right),$$
$$\text{Var}(\eta_t|D_t) \;=\; q_t^* \;\approx\; \frac{1}{r_t + Y_t},$$

where $r_t + Y_t$ and $s_t + 1$ are also denoted as r_t^* and s_t^*, respectively, and are the parameters of the posterior distribution μ_t, that is, $\mu_t \sim Gamma(r_t^*, s_t^*)$.

Calculated the approximate values of r_t, s_t, f_t^* and q_t^*, the moments of $\theta_t|D_t$ can be obtained as Equations (3.11) and (3.12), as well as the one-step ahead forecast as Equation (3.17).

In Table 3.1 we present the approximate values of r_t, s_t, f_t^* e q_t^*, for some distributions in the uni-parametric exponential family, considering the identity link function. That way, for those distributions, we are able to proceed the DGLM updating through the five steps presented above, finding the moments of $\theta_t|D_t$ as Equations (3.11) and (3.12), and making forecasts as Equation (3.17).

Distribution	r_t	s_t	f_t^*	q_t^*
Poisson(μ_t)	$\frac{1}{q_t}$	$\frac{\exp(-f_t)}{q_t}$	$\ln\left(\frac{r_t+Y_t}{s_t+1}\right)$	$\frac{1}{r_t+Y_t}$
Bin.(m_t, μ_t)	$\frac{1+\exp(f_t)}{q_t}$	$\frac{1+\exp(-f_t)}{q_t}$	$\ln\left(\frac{r_t+Y_t}{s_t+m_t-Y_t}\right)$	$\frac{1}{r_t+Y_t} + \frac{1}{s_t+m_t-Y_t}$
Neg. Bin.(π_t, m_t)	$\frac{1-\exp(-f_t)}{q_t}$	$\frac{1-2\exp(f_t)+\exp(2f_t)}{\exp(f_t)q_t}$	$\ln\left(\frac{s_t+Y_t}{s_t+Y_t+r_t+m_t}\right)$	$\frac{1}{s_t+Y_t} - \frac{1}{s_t+Y_t+r_t+m_t}$
Normal(μ_t, V_t)	f_t	q_t	$\frac{s_tY_t+r_tV}{V+s_t}$	$\frac{s_tV}{V+s_t}$
Gamma(α_t, β_t)	$\frac{-f_t}{q_t}$	$\frac{f_t^2}{q_t}$	$\frac{-r_t}{s_t+Y_t}$	$\frac{r_t}{(s_t+Y_t)^2}$
Pareto(λ_t)	$\frac{\exp(-f_t)}{q_t}$	$\frac{1-q_t}{q_t}$	$\ln\left(\frac{s_t+\ln(Y_t)+1}{r_t+1}\right)$	$\frac{2s_t+2\ln(Y_t)+1}{2(s_t+\ln(Y_t)+1)}$

Table 3.1: Approximate values of r_t, s_t, f_t^* e q_t^* for some distributions in the uniparametric exponential family

The dynamic linear models can be represented graphically as shown in Figure 3.1, that is, Y_t, $t = 1, ..., n$, a random component is related to a parameter θ_t, and these parameters being stochastically related. Specifically, in the dynamic generalized linear model the random component is random variables which distribution is a member of the uniparametric exponential family, and it is related to the parameter θ_t through a link function.

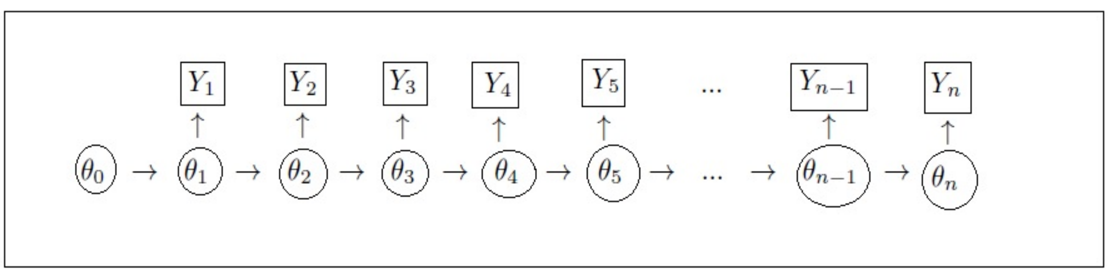


Figure 3.1: Graphical representation of the dynamic linear model.

Chapter 4

Product Partition Model

Once time series data is in study, it is reasonable to think about change point problem, result of some intervention that can change the structure of the series and provide a clustering structure. An attractive tool to work with this kind of problem was proposed by Hartigan (1990), known as the product partition model (PPM), which allowed the data to weight the partitions likely to hold, and assumed that observations in different components of random partitions of the data are independent given the partition. The work of Barry and Hartigan (1992) consider an approach were the PPM was defined in terms of a partition of a set of observations into contiguous blocks, wherein the partition has a prior product distribution, and given the partition, the parameters in different blocks have independent prior distributions, and is describe as follows.

Consider an observed time series $\boldsymbol{Y} = (Y_1, ..., Y_n)$, that is, a sequence of observations at consecutive points in time, wherein, conditionally on $\theta_1, ..., \theta_n$, has marginal densities $p_1(Y_1|\theta_1), ..., p_n(Y_n|\theta_n)$.

Let $T = \{1, ..., n\}$ be the index set of the observed time series $\boldsymbol{Y}$ and $\rho = \{t_0, t_1, ..., t_b\}$ a random partition of the set $T \cup \{0\}$, with $0 = t_0 < t_1 < ... < t_b = n$. Let B be a random variable which represents the number of blocks in ρ. Then, the partition ρ divides the time series $\boldsymbol{Y}$ into $B = b$ contiguous blocks denoted by $Y_\rho^{(j)} = (Y_{t_{j-1}+1}, ..., Y_{t_j})'$, $j = 1, ..., b$. Let $\boldsymbol{\theta}_k = \boldsymbol{\theta}_\rho^{(j)}$ for $t_{j-1}+1 \leq k \leq t_j$ be the parameters vector conditionally on the partition $\rho = \{t_0, t_1, ..., t_b\}$.

Consider $c_\rho^{(j)}$ the prior cohesion associated with block $Y_\rho^{(j)}$, which represents the degree of similarity among the observations $Y_\rho^{(j)}$. In the time series context, as pointed out by Loschi and Cruz (2002), the cohesion may be interpreted as the transition probabilities in the Markov chain defined by the endpoints of the blocks in ρ.

Following Loschi and Cruz (2002) we say that the random quantity $(Y_1, ..., Y_n; \rho)$ follows a PPM, that is, $(Y_1, ..., Y_n; \rho) \sim PPM$ if

1. The prior distribution of $\rho = \{t_0, t_1, ..., t_b\}$ is

$$p(\rho = \{t_0, t_1, ..., t_b\}) = \frac{\prod_{j=1}^{b} c_{\rho}^{(j)}}{\sum_{\mathcal{C}} \prod_{j=1}^{b} c_{\rho}^{(j)}}, \tag{4.1}$$

   where $\mathcal{C}$ is the set of all possible partitions of the set T into b contiguous blocks with endpoints $t_1, ..., t_b$ satisfying the condition $0 = t_0 < t_1 < ... < t_b = n$, for all $b \in T$;

2. Given the partition $\rho = \{t_0, t_1, ..., t_b\}$, the sequence $Y_1, ..., Y_n$ has joint density given by

$$p(\boldsymbol{Y}|\rho = \{t_0, t_1, ..., t_b\}) = \prod_{j=1}^{b} p(\boldsymbol{Y}_{\rho}^{(j)}), \tag{4.2}$$

   where

$$p(\boldsymbol{Y}_{\rho}^{(j)}) = \int p(\boldsymbol{Y}_{\rho}^{(j)}|\theta_{\rho}^{(j)})p(\theta_{\rho}^{(j)})d\theta_{\rho}^{(j)}, \tag{4.3}$$

   is the density of the random vector, with $p(\theta_{\rho}^{(j)})$ being the block prior density for $\theta_{\rho}^{(j)}$.

The graphical representation of the PPM is presented in Figure 4.1, that is, a random component $(Y_1, ..., Y_{t_n})$ that given $\theta_1, ..., \theta_{t_n}$ are conditionally independent, and given the partition the parameters $\theta_{\rho}^{(1)}, ..., \theta_{\rho}^{(b)}$ are independent. Loschi and Cruz (2002) emphasizes that the formulation of the product partition model allows the parameters to be time varying, and then it is a kind of dynamic model.

The prior distribution for B, that is, the number of blocks in the partition is given by

$$p(B = b) \propto \sum_{\mathcal{C}_1} \prod_{j=1}^{b} c_{[i_{j-1}i_j]}, \quad b \in T,$$

where $\mathcal{C}_1$ is the set of all partitions of T in b contiguous blocks with endpoint $t_1, ..., t_b$ satisfying the condition $0 = t_0 < t_1 < ... < t_b = n$.

Barry and Hartigan (1992) showed that the posterior distribution of ρ and B have the same form of the prior distribution, using the posterior cohesion for the j-th block presented below

$$c_{\rho}^{*(j)} = c_{\rho}^{(j)} p(\boldsymbol{Y}_{\rho}^{(j)}). \tag{4.4}$$

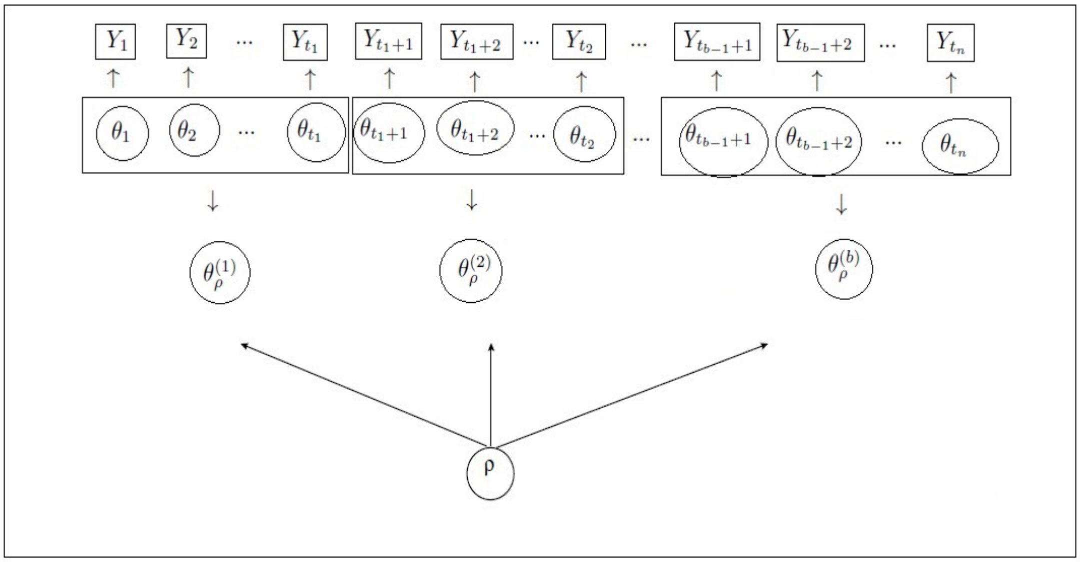


Figure 4.1: Graphical representation of the product partition model.

It was also showed that the posterior distributions of θ_k, $k = 1, ..., n$ is given by

$$p(\theta_k|\boldsymbol{Y}) = \sum_{t_{j-1} \leq k \leq t_j} r_\rho^{*(j)} p(\theta_\rho^{(j)}|\boldsymbol{Y}_\rho^{(j)}), \tag{4.5}$$

and the posterior expectation, or product estimate, of θ_k is given by

$$E(\theta_k|\boldsymbol{Y}) = \sum_{t_{j-1} \leq k \leq t_j} r_\rho^{*(j)} E(\theta_\rho^{(j)}|\boldsymbol{Y}_\rho^{(j)}), \tag{4.6}$$

where $r_\rho^{*(j)}$ is the posterior relevance for the block j, which given by

$$r_\rho^{*(j)} = P([t_{j-1}t_j] \in \rho|\boldsymbol{Y}).$$

According to Hartigan (1990), for any probability model over partitions, the relevance probability is defined to be the probability that the block $\boldsymbol{Y}_\rho^{(j)}$, $j = 1, ..., b$, is a component of the random partition ρ.

Although the product estimate in Equation (4.6) can be analytically calculated, it demands high computational efforts. Therefore, aiming to simplify the implementation of the method discussed previously, Barry and Hartigan (1993) suggested the use of an auxiliary random quantity U_i, in a Gibbs sampling scheme, given by

$$U_i = \begin{cases} 1 & if \;\; \theta_i = \theta_{i+1}, \\ 0 & if \;\; \theta_i \neq \theta_{i+1}, \end{cases}$$

which reflects whether or not a change point occurs at time i, for $i = 1, ..., n-1$. Once the vector $U = (U_1, ..., U_{n-1})$ is considered, the partition ρ is completely specified.

The Gibbs sampling is started with an initial value of U, that is, $(U_1^0, ..., U_{n-1}^0)$ and then generating each vector $(U_1^s, ..., U_{n-1}^s),\ s \geq 1$. The r-th element of step s is generated from the following conditional distribution

$$U_r | U_1^s, ..., U_{r-1}^s, U_{r+1}^{s-1}, ..., U_{n-1}^{s-1}; \boldsymbol{Y},$$

$r = 1, ..., n-1$.

Consider the ratio presented bellow

$$R_r = \frac{P(U_r = 1 | A_s; \boldsymbol{Y})}{P(U_r = 0 | A_s^r; \boldsymbol{Y})},$$

$r = 1, ..., n-1$, where $A_s^r = \{U_1^s = u_1, ..., U_{r-1}^s = u_{r-1}, U_{r+1}^{s-1} = u_{r+1}, ..., U_{n-1}^{s-1} = u_{n-1}\}$. Therefore, is sufficient to generate the samples of U considering the ratio above and the following criterion of choosing the values $U_i^s,\ i = 1, ..., n-1$

$$U_r^s = \begin{cases} 1 & if \quad R_r \geq \dfrac{1-u}{u}, \\ 0 & otherwise, \end{cases}$$

where $r = 1, ..., n-1$ and $u \sim \mathcal{U}[0, 1]$.

Following Barry and Hartigan (1993), a procedure to obtain the product estimates of θ_k can be described as follows. For each partition $(U_1^s, ..., U_{n-1}^s),\ s \geq 1$, consider $\hat{\theta_{ks}}$ the estimates per block. The product estimates of θ_k, are approximated by

$$\hat{\theta_k} = \frac{\sum_{s=1}^{M} \hat{\theta_{ks}}}{M},$$

where M is the net size of the generated sample.

Chapter 5

Dynamic Generalized Linear Model via Product Partition Model

The dynamic generalized linear model, defined in Section 3.2, can be extended to allow for random partitions of the observed time series by using the product partition model, in order to accommodate time series in the uni-parametric exponential family of distributions with experiences changes point through time. The new class, which we shall refer here as dynamic generalized linear model via product partition model (DGLM via PPM), can be represented graphically as Figure 5.1. The PPM conditional independence of $(Y_1, ..., Y_{t_n})$ given $\theta_1, ..., \theta_{t_n}$ is conserved, but given the partition ρ the parameters $\boldsymbol{\theta}_\rho^{(1)}, ..., \boldsymbol{\theta}_\rho^{(b)}$ will be dependent. The formulation of the proposed model is given as follows.

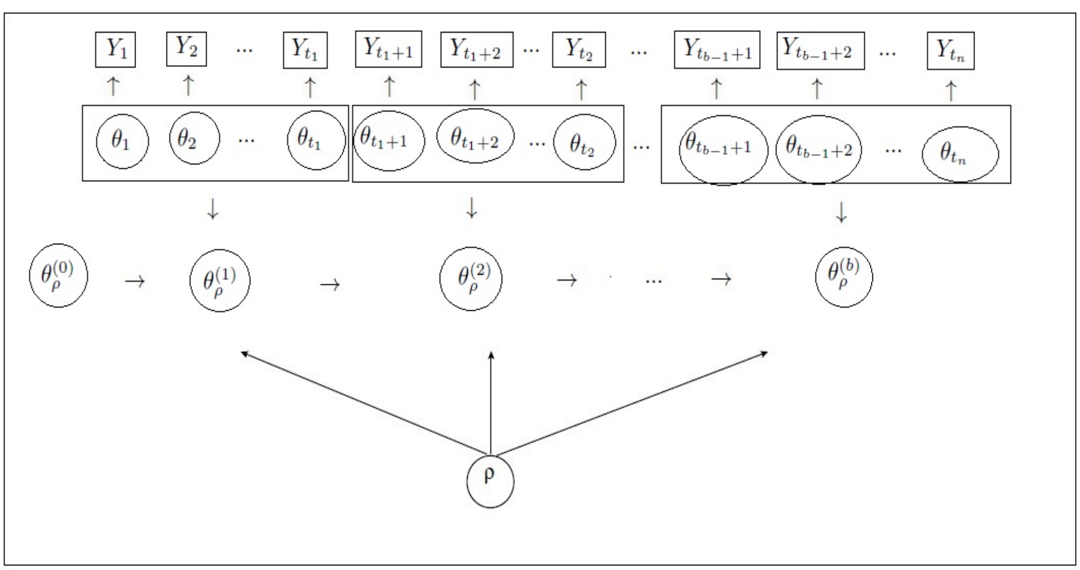


Figure 5.1: Graphical representation of the DGLM via PPM.

Consider $\boldsymbol{Y} = (Y_1, ..., Y_n)$ an observed time series that has marginal distributions $p_1(Y_1|\eta_1, V_1), ..., p_n(Y_n|\eta_n, V_n)$, member of the uni-parametric exponential family, and assume that $V_1, ..., V_n$ is dropped from the list of conditioning arguments as it is assumed to be known.

Let $T = (1, ..., n)$ be the index set. Consider $\rho = \{t_0, t_1, ..., t_b\}$, $0 = t_0 < t_1 < ... < t_b = n$ a random partition of the set $T \cup \{0\}$, which divides $Y_1, ..., Y_n$ into $B = b$ contiguous blocks, with B being a random variable representing the number of blocks. Each block is denoted by $\boldsymbol{Y}_\rho^{(j)} = (Y_{t_{j-1}+1}, ..., Y_{t_j})' \equiv (Y_{j1}, ..., Y_{jn_j})'$, $j = 1, ..., b$, where n_j is the number of elements in the block $\boldsymbol{Y}_\rho^{(j)}$. Notice that Y_{jk} denotes the k-th element of the block $\boldsymbol{Y}_\rho^{(j)}$.

Assume the following distribution of $\rho = \{t_0, t_1, ..., t_b\}$

$$p(\rho = \{t_0, t_1, ..., t_b\}) \;=\; \frac{\prod_{j=1}^{b} c_\rho^{(j)}}{\sum_{\mathcal{C}} \prod_{j=1}^{b} c_\rho^{(j)}}, \tag{5.1}$$

where $c_\rho^{(j)}$ denotes the prior cohesion associated with the block $\boldsymbol{Y}_\rho^{(j)}$ and $\mathcal{C}$ is the set of all possible partitions of the set T into b contiguous blocks with endpoints $t_1, ..., t_b$ satisfying the condition $0 = t_0 < t_1 < ... < t_b = n$, for all $b \in T$.

The prior distribution for B, that is, the number of blocks in the partition is given by

$$p(B = b) \;\propto\; \sum_{\mathcal{C}_1} \prod_{j=1}^{b} c_{[i_{j-1} i_j]}, \quad b \in T,$$

where $\mathcal{C}_1$ is the set of all partitions of T in b contiguous blocks with endpoint $t_1, ..., t_b$ satisfying the condition $0 = t_0 < t_1 < ... < t_b = n$.

Suppose that a set of independent explanatory variables, denoted by $\boldsymbol{F}_{jk}$, $k = 1, ..., n_j$, $j = 1, ..., b$, where $\boldsymbol{F}_{jk}$ is p-dimensional vector, known as design vector, is observed for Y_{jk}. Following West et al. (1985) it assumed that Y_{jk} is related to $\boldsymbol{F}_{jk}$ through an observational model given by

$$g(\eta_{jk}) \;=\; \lambda_{jk} \;=\; \boldsymbol{F}'_{jk}\boldsymbol{\theta}_\rho^{(j)}, \tag{5.2}$$

where $g(\eta_{jk})$ is a known, continuous and monotonic function mapping η_{jk} to the real line.

Given ρ, $\boldsymbol{\theta}_\rho = (\boldsymbol{\theta}_\rho^{(1)}, \boldsymbol{\theta}_\rho^{(2)}, ..., \boldsymbol{\theta}_\rho^{(b)})$, where $\boldsymbol{\theta}_\rho^{(j)}$ is the state vector associated with the block $Y_\rho^{(j)}$. In the dynamic modeling, the model specification is completed with the prior moments for $\boldsymbol{\theta}_\rho = (\boldsymbol{\theta}_\rho^{(1)}, \boldsymbol{\theta}_\rho^{(2)}, ..., \boldsymbol{\theta}_\rho^{(b)})$. For the canonical parameter, η_{jk}, it is assumed the conjugate prior distribution, denoted by $\eta_{jk} \sim \text{CF}(r_{jk}, s_{jk})$, given by

$$p(\eta_{jk}) = c(r_{jk}, s_{jk}) \exp\{r_{jk}\eta_{jk} - s_{jk}a(\eta_{jk})\}, \tag{5.3}$$

where CF denotes the conjugate family and the hyperparameters r_{jk} and s_{jk} are defining quantities obtained from the moments of $g(\eta_{jk}) = \boldsymbol{F}'_{jk}\boldsymbol{\theta}^{(j)}_{\rho}$, and $c(r_{jk}, s_{jk})$ is a normalizing constant.

Therefore, conditionally on $\rho = \{t_0, t_1, ..., t_b\}$ the observations have the following joint density

$$p(\boldsymbol{Y}|\rho = \{t_0, t_1, ..., t_b\}) = \prod_{j=1}^{b} p(\boldsymbol{Y}^{(j)}_{\rho}), \tag{5.4}$$

where

$$\begin{aligned}
p(\boldsymbol{Y}^{(j)}_{\rho}) &= \prod_{k=1}^{n_j} \int p(Y_{jk}|\eta_{jk}) p(\eta_{jk}) d\eta_{jk} \\
&= \prod_{k=1}^{n_j} c(r_{jk}, s_{jk}) b(Y_{jk}, V_{jk}) \int \exp\{V_{jk}^{-1}[Y_{jk}\eta_{jk} - a(\eta_{jk})]\} \exp\{r_{jk}\eta_{jk} - s_{jk}a(\eta_{jk})\}\, d\eta_{jk} \\
&= \prod_{k=1}^{n_j} \frac{c(r_{jk}, s_{jk}) b(Y_{jk}, V_{jk})}{c(r_{jk} + \phi_{jk}Y_{jk},\ s_{jk} + \phi_{jk})},
\end{aligned}$$

is the predictive distribution of the observations in the $Y^{(j)}_{\rho}$ block, known as data factor. For all distributions member of the exponential family $c(.)$ and $c(.)$ are determined functions, thus the predictive function can also be determined in all cases.

The posterior distribution of $\rho = \{t_0, t_1, ..., t_b\}$ is given by

$$p(\rho = \{t_0, t_1, ..., t_b\}|\boldsymbol{Y}) = \frac{\prod_{j=1}^{b} c^{*(j)}_{\rho}}{\sum_{\mathcal{C}} \prod_{j=1}^{b} c^{*(j)}_{\rho}},$$

where $c^{*(j)}_{\rho} = c^{(j)}_{\rho} p(\boldsymbol{Y}^{(j)}_{\rho})$ is the posterior cohesion.

The posterior distribution of B have the same form of the prior distribution, using the posterior cohesion for the j-th block.

Hence, the condition to apply the product partition model (see Loschi and Cruz (2002), for details) is satisfied since according to Barry and Hartigan (1992) any joint distribution on observations and partitions that satisfies the product condition for partitions and the independence condition for observations given the partition will be called product partition model.

Assume that the state parameters from different blocks are related through the evolution equation:

$$\boldsymbol{\theta}_{\rho}^{(j)} \;=\; \boldsymbol{G}_j \boldsymbol{\theta}_{\rho}^{(j-1)} + \boldsymbol{w}_j, \tag{5.5}$$

where $\boldsymbol{G}_j$ is a known evolution matrix of $\boldsymbol{\theta}_{\rho}^{(j)}$ and $\boldsymbol{w}_j$ is an evolution error associated to the block $\boldsymbol{Y}_{\rho}^{(j)}$ having mean vector equal to zero and covariance matrix equal to $\boldsymbol{W}_j$.

The prior for the state vector, that is, $\boldsymbol{\theta}_{\rho}^{(j)}$, will be specified as usual in dynamic modeling. Let $\boldsymbol{D}_{\rho}^{(j)}$ be the data information available up to the block j. The model specification is completed by the following posterior distribution

$$(\boldsymbol{\theta}_{\rho}^{(j-1)} | \boldsymbol{D}_{\rho}^{(j-1)}) \;\sim\; (\boldsymbol{m}_{j-1}, \boldsymbol{C}_{j-1}).$$

Then, using the evolution equation in (5.5), the prior distribution for $\boldsymbol{\theta}_{\rho}^{(j)}$ obtained is

$$(\boldsymbol{\theta}_{\rho}^{(j)} | \boldsymbol{D}_{\rho}^{(j-1)}) \;\sim\; (\boldsymbol{a}_j, \boldsymbol{R}_j),$$

where

$$\boldsymbol{a}_j = \boldsymbol{G}_j \boldsymbol{m}_{j-1} \quad \text{and} \quad \boldsymbol{R}_j = \boldsymbol{G}_j \boldsymbol{C}_{j-1} \boldsymbol{G}_j' + \boldsymbol{W}_j.$$

Denote by $\boldsymbol{D}_{\rho}^{(j-1;k-1)}$ the set of all information available until the block $j-1$ and the proceed information up to the element $k-1$. Assume the following joint prior distribution for λ_{jk} and $\boldsymbol{\theta}_{\rho}^{(j)}$

$$\left(\begin{array}{c} \lambda_{jk} \\ \boldsymbol{\theta}_{\rho}^{(j)} \end{array} \Big| \; \boldsymbol{D}_{\rho}^{(j-1;k-1)} \right) \sim \left[\left(\begin{array}{c} f_{jk} \\ \boldsymbol{a}_{jk} \end{array} \right) , \left(\begin{array}{cc} q_{jk} & \boldsymbol{F}_k' \boldsymbol{R}_{jk} \\ \boldsymbol{R}_{jk} \boldsymbol{F}_k & \boldsymbol{R}_{jk} \end{array} \right) \right] \tag{5.6}$$

where $f_{jk} = \boldsymbol{F}_k' \boldsymbol{a}_{jk}$ and $q_{jk} = \boldsymbol{F}_k' \boldsymbol{R}_{jk} \boldsymbol{F}_k$.

The one-step ahead forecast distribution is given by

$$p(Y_{jk} | \boldsymbol{D}_{\rho}^{(j-1;k-1)}) \;=\; \frac{c(r_{jk}, s_{jk}) b(Y_{jk}, V_{jk})}{c(r_{jk} + \phi_{jk} Y_{jk}, s_{jk} + \phi_{jk})} \tag{5.7}$$

The posterior distribution for η_{jk} is given by

$$p(\eta_{jk} | \boldsymbol{D}_{j-1;k}) \;=\; c(r_{jk} + \phi_{jk} Y_{jk}, s_{jk} + \phi_{jk}) \exp[(r_{jk} + \phi_{jk} Y_{jk}) \eta_{jk} - (s_{jk} + \phi_{jk}) a(\eta_{jk})]. \tag{5.8}$$

Matching the moments of the linear predictor and the canonical parameter we obtain that

$$E[g(\eta_{jk}) | \boldsymbol{D}_{j-1;k}] = f_{jk}^* \quad \text{and} \quad Var[g(\eta_{jk}) | \boldsymbol{D}_{j-1;k}] = q_{jk}^*. \tag{5.9}$$

Hence, we are able to update the state parameter as it is usually done in DGLM framework. That is, $(\boldsymbol{\theta}_\rho^{(j)}|\boldsymbol{D}_\rho^{(j-1;k-1)}) \sim (\boldsymbol{a}_{jk}, \boldsymbol{R}_{jk})$ is updated into $(\boldsymbol{\theta}_\rho^{(j)}|\boldsymbol{D}_\rho^{(j-1;k)}) \sim (\boldsymbol{m}_{jk}, \boldsymbol{C}_{jk})$, where

$$\boldsymbol{m}_{jk} \;=\; \boldsymbol{a}_{jk} + \boldsymbol{R}_{jk}\boldsymbol{F}_{jk}(f^*_{jk} - f_{jk})/q_{jk}, \qquad (5.10)$$

and

$$\boldsymbol{C}_{jk} \;=\; \boldsymbol{R}_{jk} - \boldsymbol{R}_{jk}\boldsymbol{F}_{jk}\boldsymbol{F}'_{jk}\boldsymbol{R}_{jk}(1 - q^*_{jk}/q_{jk})/q_{jk}. \qquad (5.11)$$

At the start of each block Y_ρ^j the updating is performed by taking $a_{j1} = a_j$, $R_{j1} = R_j$ and $\boldsymbol{D}_\rho^{j-1;0} = \boldsymbol{D}^{j-1}$. Within each block it is assumed that there is no parametric evolution, and we take $a_{j;k+1} = m_{jk}$ and $R_{j;k+1} = C_{jk}$ until the data information of all elements of the block is processed. Then, $m_{j;n_j} = m_j$, $c_{j,n_j} = C_j$ and $\boldsymbol{D}_\rho^{j;n_j} = \boldsymbol{D}_\rho^{(j)}$ and perform the parametric evolution. This cycle is repeated until $\boldsymbol{D}_\rho^{(b)} = \boldsymbol{D}$ is processed.

Example 5.1 *The Poisson model*
Assume a time series Y_{jk}, $k = 1, ..., n_j$, $j = 1, ..., b$, such that $Y_{jk}|\mu_{jk} \sim Poisson(\mu_{jk})$. From Table 2.1 it is known the canonical parameter and the function $b(.)$, that is,

$$\eta_{jk} \;=\; \ln \mu_{jk} \quad \textit{and} \quad b(Y_{jk}, V) \;=\; \frac{1}{Y_{jk}!}.$$

The conjugate prior for $\eta_{jk} = \ln \mu_{jk}$ is described in Example 2.3 as follows.

$$p(\eta_{jk}|r_{jk}, s_{jk}) \;=\; c(r_{jk}, s_{jk}) \exp\{r_{jk}\eta_{jk} - s_{jk}e^{\eta_{jk}}\},$$

with normalizing constant

$$c(r_{jk}, s_{jk}) = \left(\int \exp\{r_{jk}\eta_{jk} - s_{jk}e^{\eta_{jk}}\}\; d\eta_{jk}\right)^{-1} \;=\; \frac{s_t^{r_t}}{\Gamma(r_t)},$$

where, from Table 3.1, it is know that,

$$s_{jk} = \frac{\exp(-f_{jk})}{q_{jk}} \quad \textit{and} \quad r_{jk} = \frac{1}{q_{jk}}.$$

That way, the predictive distribution of the observations in the $Y_\rho^{(j)}$ block may be derived, that is

$$
\begin{aligned}
p(\boldsymbol{Y}_\rho^{(j)}) &= \prod_{k=1}^{n_j} \frac{c(r_{jk}, s_{jk}) b(Y_{jk}, V_{jk})}{c(r_{jk} + \phi_{jk} Y_{jk},\ s_{jk} + \phi_{jk})} \\
&= \frac{s_{jk}^{r_{jk}} \Gamma(r_{jk} + Y_{jk})}{(s_{jk}+1)^{r_{jk}+Y_{jk}} \Gamma(r_{jk}) Y_{jk}!} \\
&= \begin{pmatrix} r_{jk} + Y_{jk} - 1 \\ Y_{jk} \end{pmatrix} \left(\frac{s_{jk}}{s_{jk}+1} \right)^{r_{jk}} \left(\frac{1}{s_{jk}+1} \right)^{Y_{jk}}.
\end{aligned}
$$

From Table 3.1 we also have the following approximate values of

$$
f_{jk}^* = \ln\left(\frac{r_{jk} + Y_{jk}}{s_{jk}+1} \right) \qquad \textit{and} \qquad q_{jk}^* = \frac{1}{r_{jk} + Y_{jk}}
$$

Assuming a discrete uniform prior distribution for ρ, that is $c_\rho^{(j)} = 1\ \forall j$, the Gibbs sampling scheme described in Chapter 4 may be executed in order to obtain some inference about the partition. Then, given the partition, and computed the approximate values of r_{jk}, s_{jk}, f_{jk}^* and q_{jk}^*, the DGLM inference about the state vector as Equations (5.10) and (5.11), and future observations as Equation (5.7) may be proceed.

In Tables 2.1 and 3.1 we presented the values of the functions $c(.)$ and $b(.)$, and the approximate values of r_{jk}, s_{jk}, f_{jk}^* and q_{jk}^*, for some distributions in the uni-parametric exponential family. Then, for these distributions we are able to provide the DGLM via PPM inference as in Example 5.1.

5.1 Model comparison

In Bayesian context, the Bayes Factor (BF) and the Posterior Model Probability (PMP) (Kass and Raftery, 1995, see) are popular measures for model comparison and selection. We briefly describe how the BF and the PMP can be computed in the DGLM via PPM framework.

Comparing M_i, $i = 1, ..., M$ and $2 < M < \infty$, models procedure consists, usually, of computing the PMP for each model, and the criterion of selection is based on the model with highest PMP. The posterior probability of a model M_i given the data $\boldsymbol{Y}$ is given by Bayes theorem

$$
p(M_i|\boldsymbol{Y}) = \frac{p(\boldsymbol{Y}|M) p(M_i)}{\sum\limits_{i=1}^{M} p(\boldsymbol{Y}|M_i) p(M_i)},
$$

where $p(M_i)$ is the prior probability associated to model M_i, satisfying $\sum_{i=1}^{M} p(M_i) = 1$ and $p(\boldsymbol{Y}|M)$ is the corresponding marginal distribution of $\boldsymbol{Y}$.

Let M_ρ and M_0 be the DGLM via PPM and the DGLM, respectively. When the model selection consists in the choice between two models we may assess the Bayes factor, given by

$$BF(M_\rho, M_0) = \frac{p(\boldsymbol{Y}|M_\rho)}{p(\boldsymbol{Y}|M_0)} = \frac{\sum_\rho \int p(\eta_\rho|\boldsymbol{Y}, \rho, M_\rho) p(\boldsymbol{Y}, \eta_\rho, \rho, M_\rho) p(\rho|\boldsymbol{Y}) d\eta_\rho}{\int p(\eta_0|\boldsymbol{Y}, M_0) p(\boldsymbol{Y}, \eta_0, M_0) d\eta_0}.$$

Notice that the marginal distribution $p(\boldsymbol{Y}|M_\rho)$ is obtained by firstly computing the marginal distribution of $\boldsymbol{Y}$ conditioning on the partitions, and then averaging over all the partitions. The marginal distribution $p(\boldsymbol{Y}|M_0)$ is obtained straightforwardly.

Kass and Raftery (1995) emphasizes that the Bayes Factor is a summary of the evidence provided by the data in favor of one model, as opposed to another, and presented a suggested scale to interpret it, showed in Table 5.1.

$BF(M_\rho, M_0)$	Evidence against M_0
1 to 3.2	Not worth more than a bare mention
3.2 to 10	Substantial
10 to 100	Strong
> 100	Decisive

Table 5.1: Bayes factor interpretation scale.

Chapter 6

Application

In this chapter we present an analysis of two real time series data sets, already discussed in the literature, aiming to illustrate the usefulness of the proposed model, DGLM via PPM, and compare its inference to the DGLM inference. For both cases we provide a sensitivity analysis for different choices of the discount factor δ.

The Gibbs sampling scheme suggested by Barry and Hartigan (1993), and described in Chapter 4, was used to obtain some inference about the partition. Single chains of size 45,000 were considered. Posterior samples of size 4,000 were obtained after considering a burn-in period of 5,000 iterations and a lag of 10 to eliminate correlations. We assumed that the prior cohesion $c_{\rho}^{(j)} = 1$, $j = 1, ..., b$, implying in a discrete uniform prior distribution for $\rho = \{t_0, t_1, ..., t_b\}$.

Prior elicitation for the state parameter $\boldsymbol{\theta}$ is given by $(\boldsymbol{\theta}_{\rho}^{(1)}|\boldsymbol{D}_{\rho}^{(0)}) \sim [0, 100\mathbf{I}_p]$. We set the evolution matrix $\boldsymbol{G}_j = \mathbf{I}_p$, $j = 1, ..., b$, assuming then a random walk for the parametric evolution. The specification of the evolution covariance matrices W_j was made according to the discount factor strategy. In this case $p = 1$, due to the absence of explanatory variables in both data sets. The computational procedures were implemented in R Core Team (2013).

6.1 APCI Data

The Ample Price to Consumer Index, APCI, is measured monthly by the IBGE, Brazilian Institute of Geography and Statistics, and it is considered the official inflation index of Brazil. The APCI reflects the cost of living of families with incomes 1-40 minimum wages, residents of the metropolitan area of Sao Paulo, Rio de Janeiro, Belo Horizonte, Porto Alegre, Curitiba, Salvador, Recife, Fortaleza and Belem, besides of Distrito Federal and the city of Goiania. The data collection covers prices of sectors of trade, service providers, households and utilities. When the APCI rises, means that these items will suffer a price adjustment up, that is inflation. When the APCI decreases, it does not mean that these items will have a price decrease, but that the prices rose less then the previous period. Only when the APCI is negative, we have a price decrease, that is deflation. The government utilizes it to check if the established inflation goal is being fulfilled.

In this work we used the APCI series measured monthly in Belo Horizonte area, from the period of July, 1997 to June, 2008, resulting in 132 observations, which was previously analyzed in the work of Santos et al. (2010), presented in Figure 6.1. It can be observed an intervention around October, 2002, which occurred due to the concerns in the economy after the election of President Lula. We assumed that the series is Gaussian distributed.

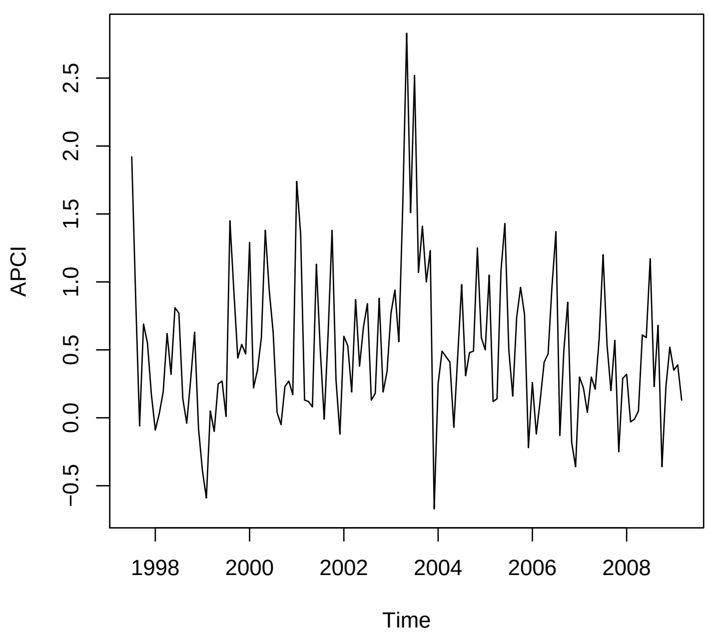


Figure 6.1: APCI series.

In Figure 6.2 we present the posterior model probability conditionally on the models considered, for $0.05 \leq \delta \leq 0.95$, that is, 19 values of discount factors. The analysis of the PMPs, conditionally on the models, provides a comparison of the effect of the discount factor on the model fitting. As we can observe in the DGLM case, the discount factor equals to 0.95 has the highest posterior probability, noting that for the other values of discount factor the posterior probabilities are really close to zero. A similar behavior is observed for the DGLM via PPM, with the discount factor equals to 0.95 having the highest posterior probability, except that in this case $\delta = 0.9$ have a higher probability than the other values at left.

Table 6.1 provides the Bayes factor for the DGLM via PPM over the DGLM for the discount factors considered. We can observe the superiority of the DGLM via PPM over the DGLM, which is classified, for the discount factors values smaller or equals to 0.90 as decisive, and for the discount factor equals to 0.95 as substantial, following the interpretation scale presented in Table 5.1. This superiority is confirmed in Figure 6.3 when analyzing the posterior model probability for the DGLM and the DGLM via PPM. We observe that for $\delta \leq 0.8$, the PMP of both models are close to zero. For $\delta > 0.8$ the models are more probables with the DGLM via PPM having bigger probabilities compared to the DGLM. The model with highest PMP is the DGLM via PPM with discount factor equal to 0.95.

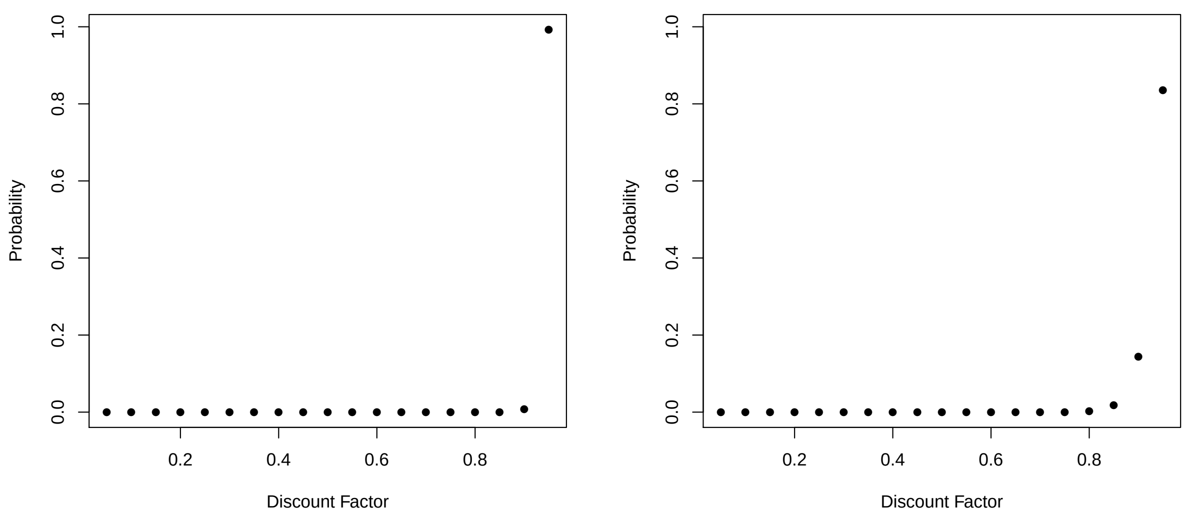


Figure 6.2: Posterior model probability for DGLM (left) and for DGLM via PPM(right) with the APCI data.

δ	Bayes Factor	δ	Bayes Factor	δ	Bayes Factor
0.05	3.48e+166	**0.40**	8.45e+36	**0.75**	5.86e+07
0.10	6.57e+122	**0.45**	3.74e+30	**0.80**	3.96e+05
0.15	6.71e+96	**0.50**	1.85e+25	**0.85**	5.45e+03
0.20	2.82e+78	**0.55**	4.25e+20	**0.90**	1.32e+02
0.25	3.76e+65	**0.60**	4.37e+16	**0.95**	6.11e+00
0.30	8.10e+53	**0.65**	1.86e+13		
0.35	7.54e+44	**0.70**	2.17e+10		

Table 6.1: Bayes Factor for different choices of δ with APCI data.

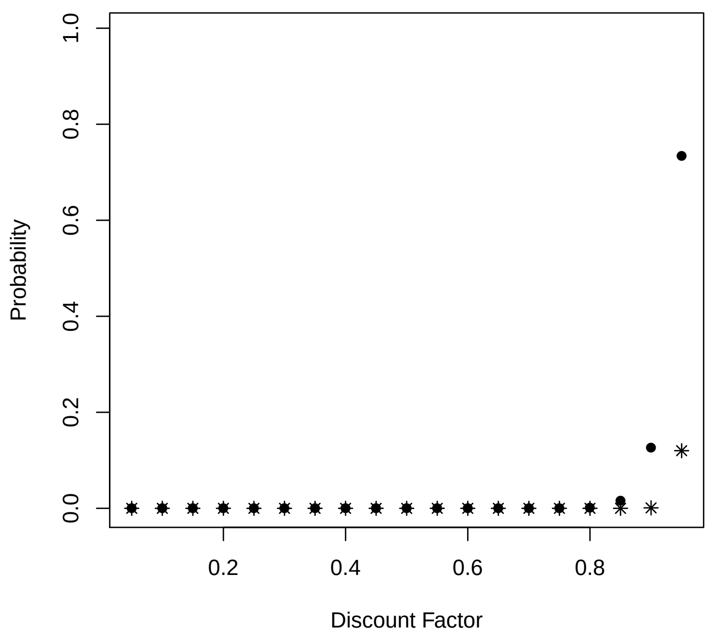


Figure 6.3: Posterior model probabilities for DGLM via PPM (solid circle) and for DGLM (asterisk symbol) with the APCI data.

We concentrate our analysis, from now on, on making inferences for the DGLM and the DGLM via PPM based on the results for $\delta = 0.95$, since it was the discount factor value with highest posterior model probability for both models.

The estimated posterior mean of the state parameter is presented in Figure 6.4, for both models considered. For this discount factor considered we detected that the most probable number of blocks is $B = 31$, with probability 0.114, and the estimated 95% HPD intervals for B is $[24, 38]$. From the formulation of the proposed model we expect the behavior of the estimated values as a consequence of the posterior partition, since we have a common state parameter for observations in the same block. Hence, in this case, we expect a smoother behavior of the estimated posterior mean as, indeed, observed. Furthermore, in Figure 6.4 we observe that the estimated values of the DGLM via PPM follows well the estimated values of the DGLM.

The forecast of the APCI series considering $\delta = 0.95$, and its relative forecast error is showed in Figure 6.5. The forecasting was obtained by the mean of the estimated forecasts obtained for each partition generated. The over smoothed behavior of the forecast was expected due to the choice of discount close to one. This choice means that a high amount of information is being allowed to pass from time $t-1$ to time t, and implies that the evolution error is close to zero. This behavior of the parametric evolution implies that we have smoother forecasts. Nevertheless, this does not indicate a bad adjustment. Indeed, we have evidences of a good adjustment when looking the relative forecast errors.

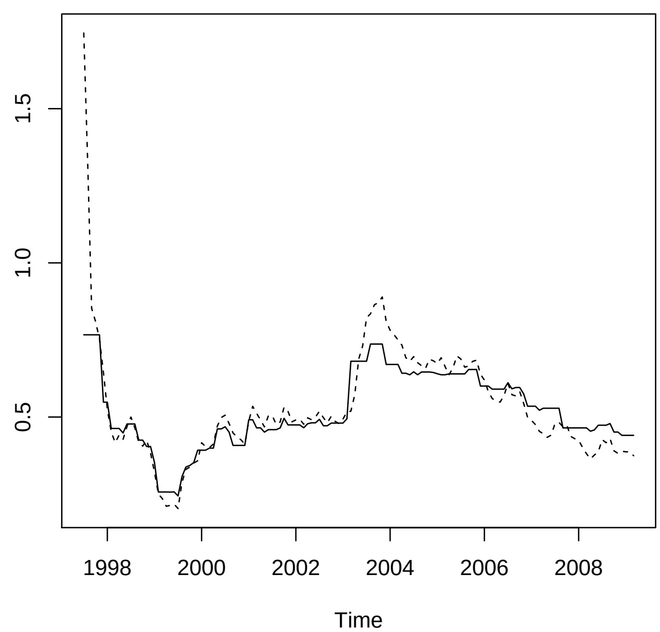


Figure 6.4: Posterior mean of the state parameter of the DGLM via PPM (solid line) and DGLM (dashed line), for $\delta = 0.95$.

Some summary measures related to the posterior distribution of the number of blocks are given in Figure 6.6 aiming to evaluate the convergence of the DGLM via PPM. We also performed the Geweke diagnostic, and conclude no problems related to the model convergence.

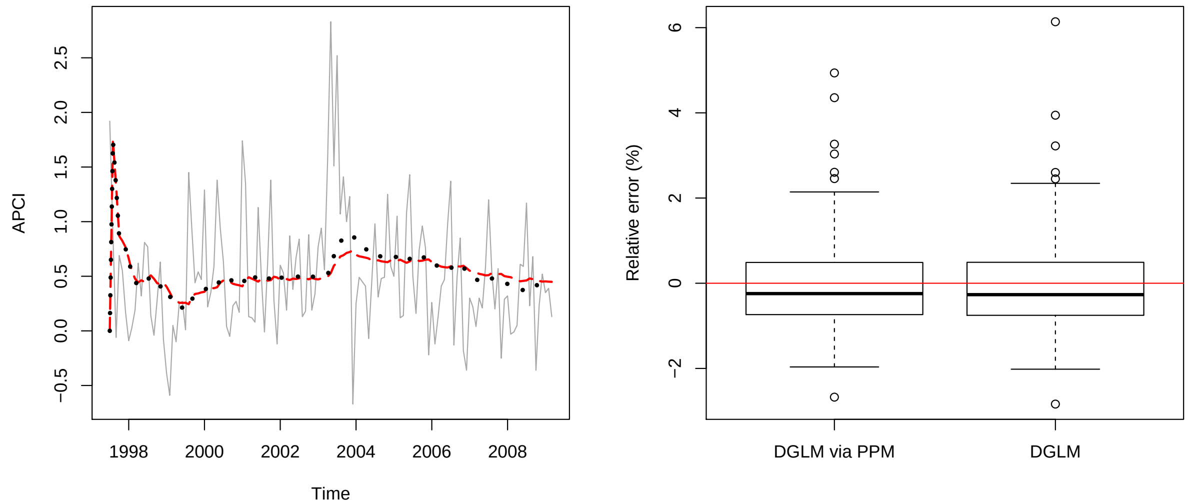


Figure 6.5: Forecast of the APCI series for the DGLM (pointed line) and DGLM via PPM (dashed line), and the relative forecast error for $\delta = 0.95$ (right).

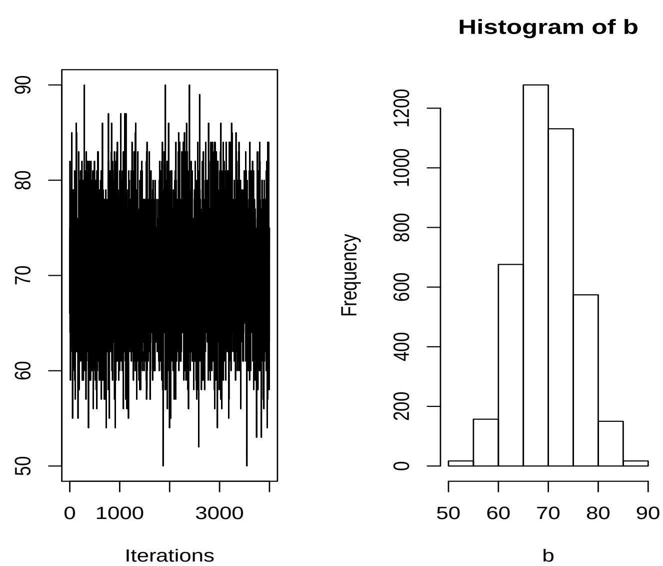


Figure 6.6: Posterior distribution of the number of intervals associated with the DGLM via PPM for $\delta = 0.95$.

One advantage of the proposed model is that we are able to make inferences about the partition as well the number of blocks. In Figure 6.7 we plotted the most probable number of blocks for both discount factors considered in order to investigate the behavior of the number of blocks in the partitions according to the choice discount factor. We observe that the number of blocks has a crescent behavior as the discount factor increase.

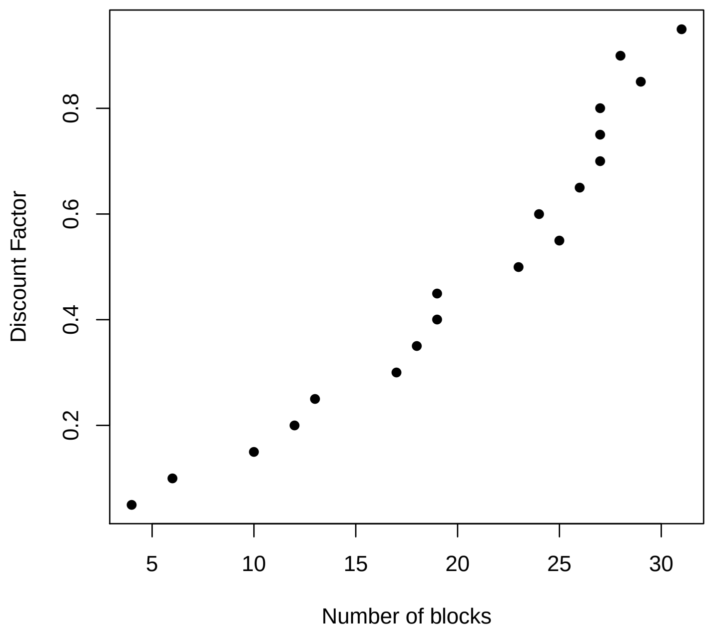


Figure 6.7: Most probable number of blocks for the discount factors considered with the APCI series

Aiming to investigate more about the partitions provided, we displayed in Figure 6.8 the change point probability in the DGLM via PPM, with $\delta = 0.95$. As it is well known, from the series, was expected that the model gives more probability to the observations around October, 2002 be a change point, due to the some economic changes related to the election of the President Lula. But, we observed that the model with this discount factor was given probability around 0.5 for all observations to be a change point. This behavior was not expected and it is not satisfactory from the inferential point of view.

In order to investigate the impact of the discount factor on the change point probability we plotted in Figure 6.21 those probabilities for some others discount factors values. For $\delta \leq 0.50$ we observed a higher probability around October, 2002, indicating that the expected change point was clearly detected. We also noted that these models had a higher, but not so clearly as before, change point probability around the beginning of the

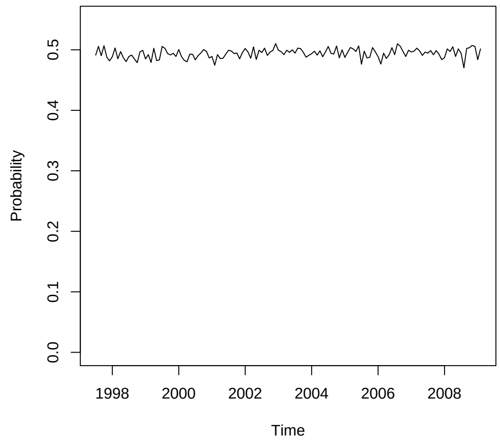
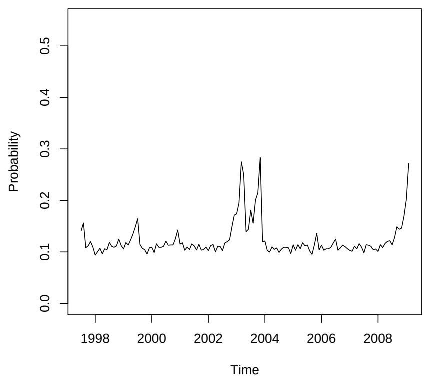
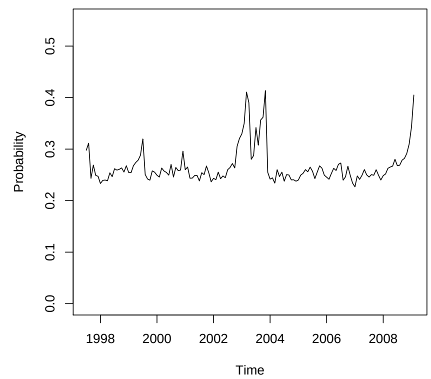


Figure 6.8: Change point probability for $\delta = 0.95$.

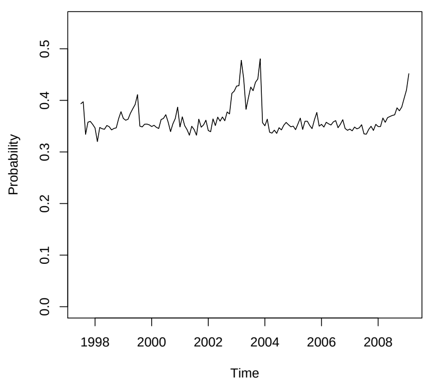
year of 1999, when the Brazilian Central Bank changed the exchange rate regime. When $\delta = 0.70$ this behavior, that indicates well the expected change point, begins dissipating , and the change points probabilities start to oscillate around 0.5. We still can observe, however, a peak around October, 2002. For $\delta = 0.80$ and $\delta = 0.9$ we can not detect anymore which observations could be a change point with higher probability, as well as in the case of $\delta = 0.95$. Hence, the results suggest that the discount factor has a great impact on the inference about the partition. A detailed studied with simulated data would be necessary to corroborate this conclusion.



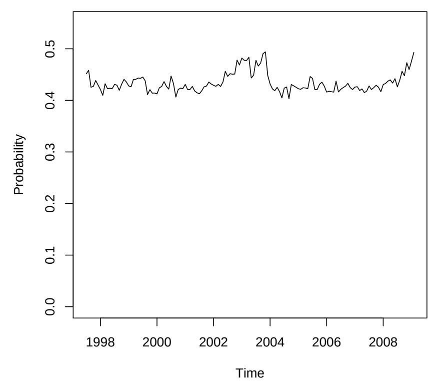
(a) Change point probability for $\delta = 0.10$.



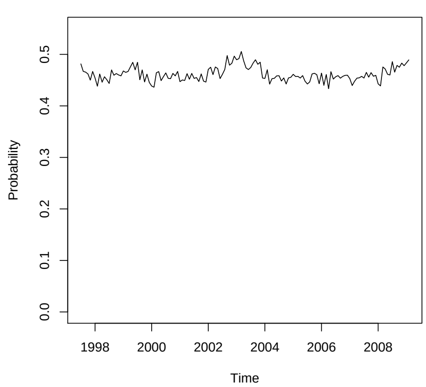
(b) Change point probability for $\delta = 0.30$.



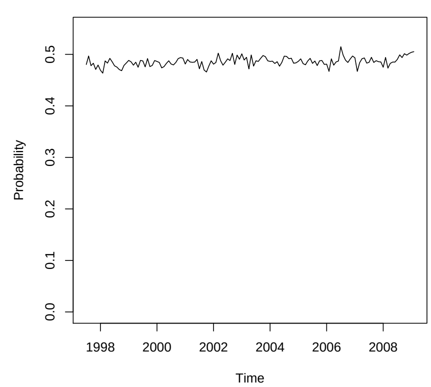
(c) Change point probability for $\delta = 0.50$.



(d) Change point probability for $\delta = 0.70$.



(e) Change point probability for $\delta = 0.80$.



(f) Change point probability for $\delta = 0.90$.

Figure 6.9: Change point probability for different values of discount factors with the APCI data.

6.2 Coal Mining Data

The coal mining data is composed by the number of fatal accidents in the period from 1851 to 1962, wherein the yearly count provides 112 observations, in coal mines of England and Wales, presented in Figure 6.10. This series is well known in the literature and was studied in the works of Worsley (1986); Gamerman (1992); Santos et al. (2010). It can be observed an intervention around the year of 1886, which changed the level of the series. We assumed that the series follows a Poisson distribution.

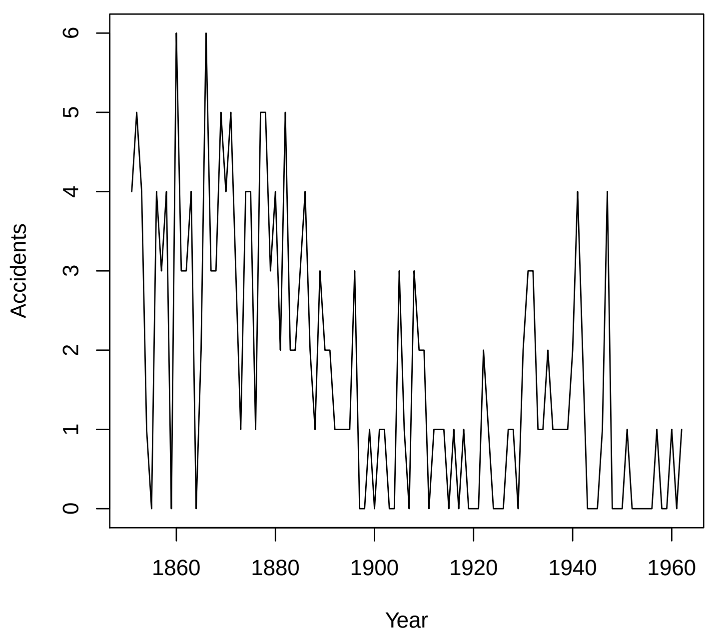


Figure 6.10: Coal mining series.

We considered 19 values of discount factors, $0.05 \leq \delta \leq 0.95$, for the DGLM and for the DGLM via PPM, resulting in 38 models evaluated. In Figure 6.11 is displayed the posterior model probabilities, conditionally on the model, which provides a sensitivity analysis allowing to compare the effect of the discount factor on the model fitting. We observe, in the DGLM case, that just the discount factors equals to 0.80 and 0.85 had posterior model probability significantly higher than zero. In the opposite, the DGLM via PPM case had a higher PMP when the discount factor was equals to 0.15, while for the other values the PMPs were really close to zero.

In Table 6.2 is displayed the Bayes factor for the DGLM via PPM over the DGLM, for these discount factors considered. We noted that for $\delta \leq 0.75$ the evidences against the DGLM are decisive, following Table 5.1. For $\delta = 0.8$ does not worth more than a bare mention the evidences against the DGLM, while for $\delta \geq 0.85$ there is no evidence against the DGLM. Analyzing Figure 6.12, that is, the posterior model probabilities associated with all 38 fitted models, we notice that the model with highest posterior probability is the DGLM via PPM with discount factor equals to 0.05. For $\delta = 0.1$ the DGLM via PPM stands out the DGLM. For the other discount factor considered the probabilities are not expressive, being considerably close to zero.

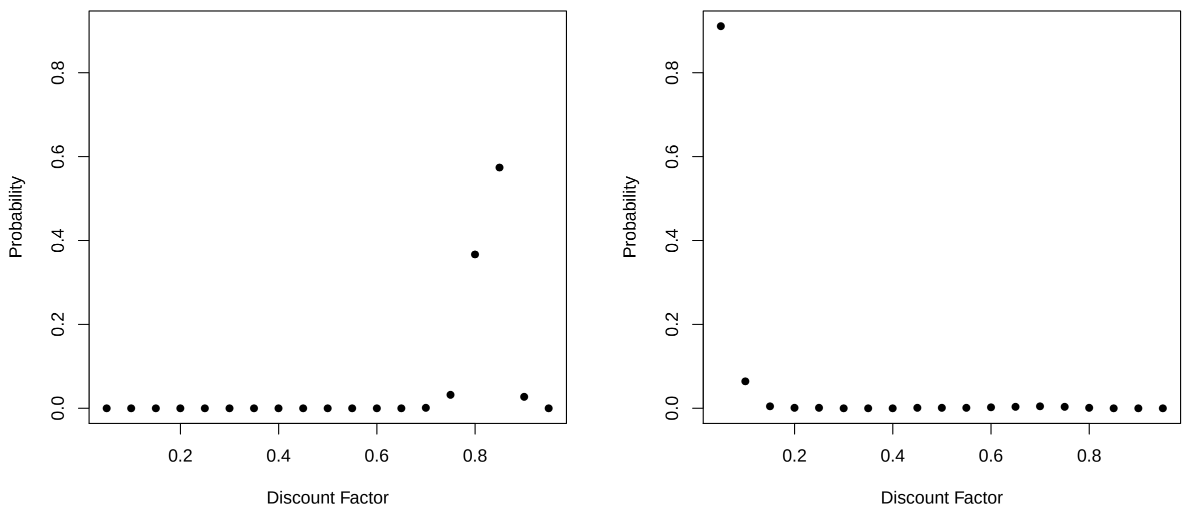


Figure 6.11: Posterior model probability for DGLM (left) and for DGLM via PPM(right) with the Coal mining data.

δ	Bayes Factor	δ	Bayes Factor	δ	Bayes Factor
0.05	6.82e+163	**0.40**	6.40e+23	**0.75**	1.25e+02
0.10	3.59e+109	**0.45**	6.05e+18	**0.80**	3.01e+00
0.15	2.21e+80	**0.50**	5.73e+14	**0.85**	1.13e-01
0.20	2.93e+61	**0.55**	2.17e+11	**0.90**	7.34e-03
0.25	1.29e+48	**0.60**	2.96e+08	**0.95**	3.69e-04
0.30	7.82e+37	**0.65**	1.06e+06		
0.35	8.61e+29	**0.70**	8.28e+03		

Table 6.2: Bayes Factor for different choices of δ with coal mining data.

In the remaining chapter we focus the analysis for the DGLM and DGLM via PPM adjusted considering $\delta = 0.05$, since this value of discount factors corresponds to the DGLM via PPM with highest posterior model probability, and for $\delta = 0.85$, once for the DGLM, this is the value with highest posterior model probability. Note that for this

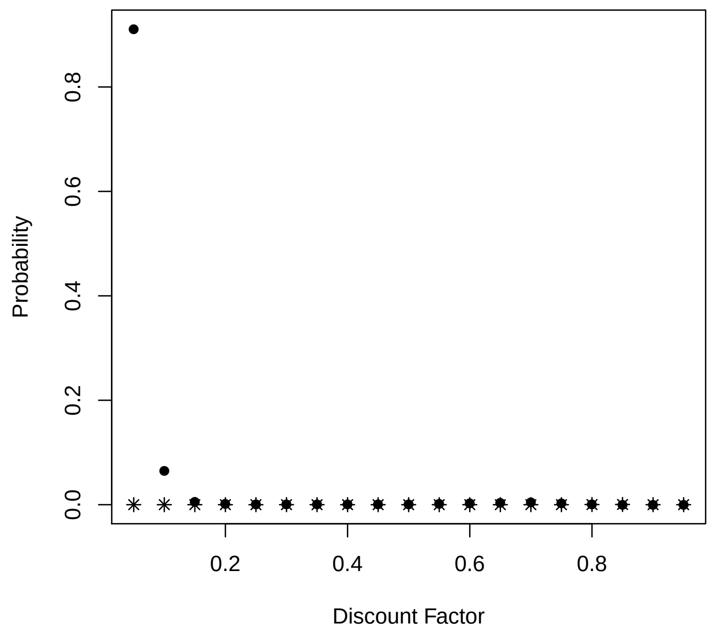


Figure 6.12: Posterior model probabilities for DGLM via PPM (solid circle) and for DGLM (asterisk symbol) with the Coal mining data.

application, we consider two different values of discount factor because we did not have an agreement of one best value for both models. We emphasize, however, that the model with highest PMP was the DGLM via PPM with $\delta = 0.05$.

The estimated posterior mean of the state parameter is presented in Figure 6.13 for both models and discount factors considered. Since in the DGLM via PPM formulation we have a common state parameter for observations in the same block, we expect the behavior of the estimated values of the state parameter as a consequence of the posterior partition. For $\delta = 0.05$ we have that the most probable number of blocks is $B = 7$ with probability 0.232 and the estimated 95% HPD interval for B is $[3, 10]$. For $\delta = 0.85$ the most probable number of blocks is $B = 34$ with probability 0.14 and the estimated 95% HPD interval for B is $[27, 40]$. That way, for $\delta = 0.05$, we have a clustering structure with a smaller number of blocks than for $\delta = 0.85$. This reflects on the behavior of the estimated values of the state parameter, as we observe in Figure 6.13, that is, for $\delta = 0.05$ we have bigger blocks with a common parameter. This also impacts in the fact that, for $\delta = 0.85$, the DGLM via PPM estimation is closer to DGLM estimation.

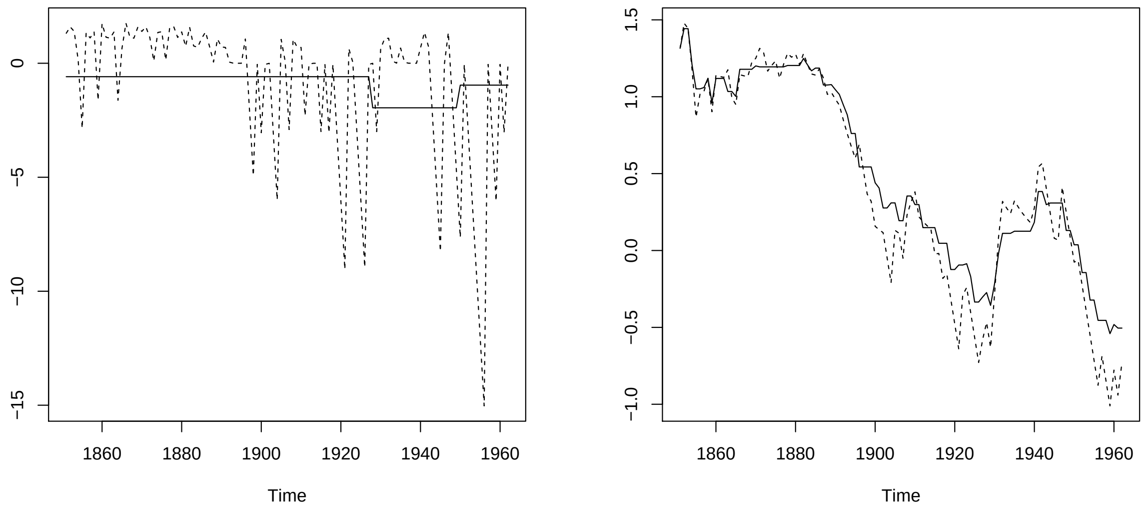


Figure 6.13: Posterior mean of the state parameter of the DGLM via PPM (solid line) and DGLM (dashed line), for $\delta = 0.05$ (left), and for $\delta = 0.85$ (right).

The forecast of the coal mining series considering $\delta = 0.05$, and its relative forecast errors is showed in Figure 6.14. The forecasting was obtained by the mean of the estimated forecasts obtained for each partition generated. Both models, DGLM and DGLM via PPM, seems to forecast well the series, noting that, as a consequence of the formulation of the product partition model, the DGLM via PPM estimated values are smoother than the DGLM ones. When a smaller discount factor is select a small amount of information is allowed to pass from time $t-1$ to time t, so we have higher evolution errors. This justifies the fact of the DGLM values do not be over smoothed. When looking the relative forecast error for both models we note that they oscillate around zero, indicating a good adjustment. For $\delta = 0.85$, the forecast of the coal mining series is displayed in Figure 6.15. In this case, the estimated values of the DGLM and the DGLM via PPM have a smoother behavior, as a consequence of the discount factor choice, that is, a higher amount of information is allowed to pass from time $t-1$ to time t, then we have evolution errors close to zero. For this discount factor we also have that the DGLM via PPM estimated values follows better the DGLM ones. The relative forecast error for both models also oscillate around zero, but now we have fewer discrepant values, indicating a good adjustment.

We displayed in Figure 6.16 the forecast of the coal mining series considering the DGLM and varying the discount factor, aiming to observe the impact of the discount factor on the forecasting. The same variation is displayed in Figure 6.17, now considering the DGLM via PPM. We note for the DGLM that the increase on the discount factor turns the forecasts smoother. The impact of the discount factor in the DGLM has already been discussed in the literature (West and Harrison, 1997, see). In the DGLM via PPM, would be necessary studies with simulated data, so we can say with more certainty what would be the impact of the choice of the discount factor on the forecasting.

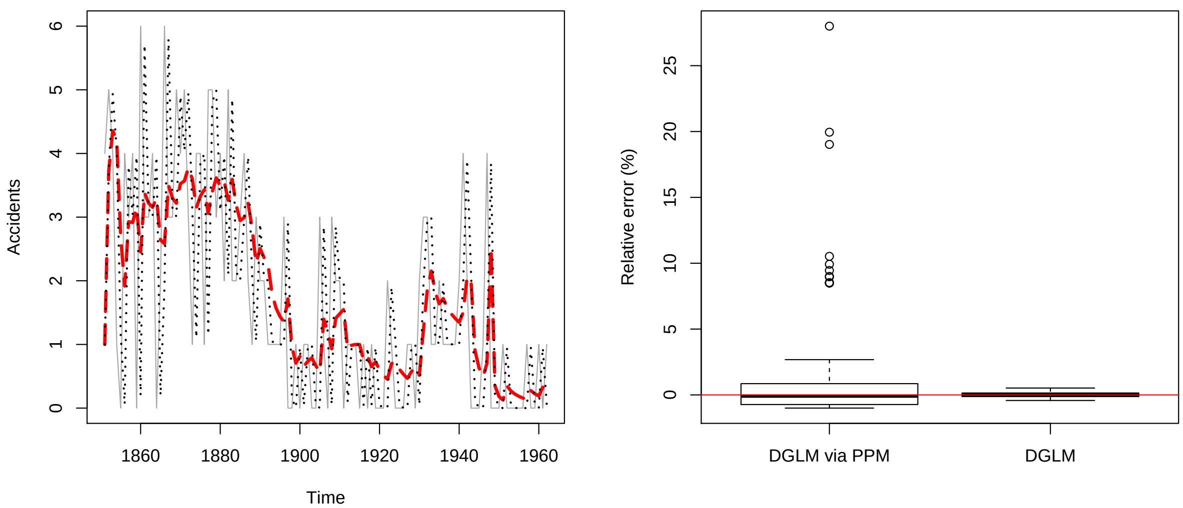


Figure 6.14: Forecast of the coal mining series for the DGLM (pointed line) and DGLM via PPM (dashed line), and the relative forecast error for $\delta = 0.05$ (right).

In Figure 6.19, we present some summary measures related to the posterior distribution of the number of blocks in order to check the convergence in the DGLM via PPM, and we did not detect any problems related to the convergence. The Geweke diagnostic was also evaluated aiming to confirm this convergence.

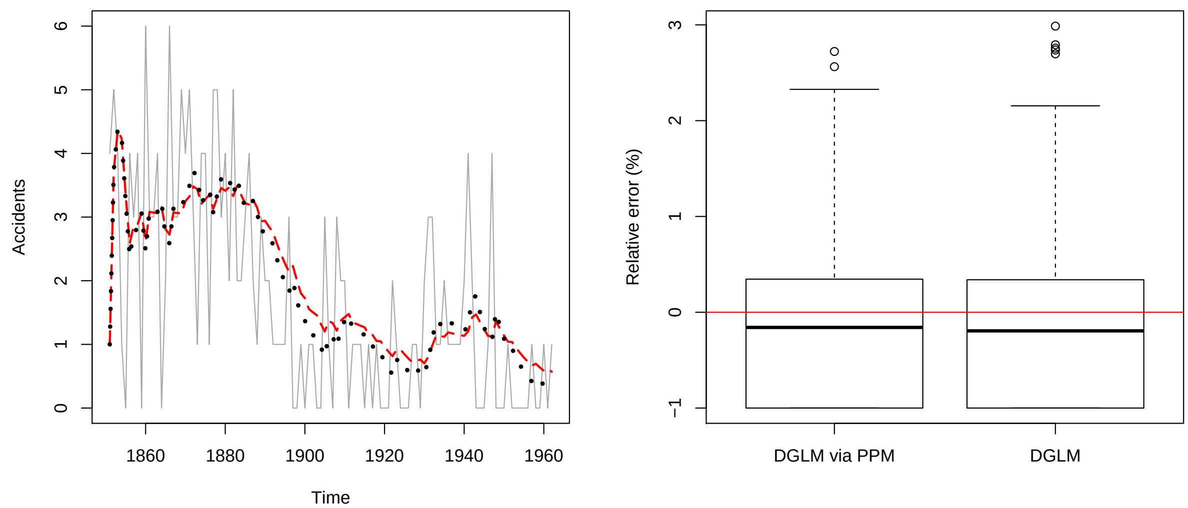


Figure 6.15: Forecast of the coal mining series for the DGLM (pointed line) and DGLM via PPM (dashed line), and the relative forecast error for $\delta = 0.85$ (right).

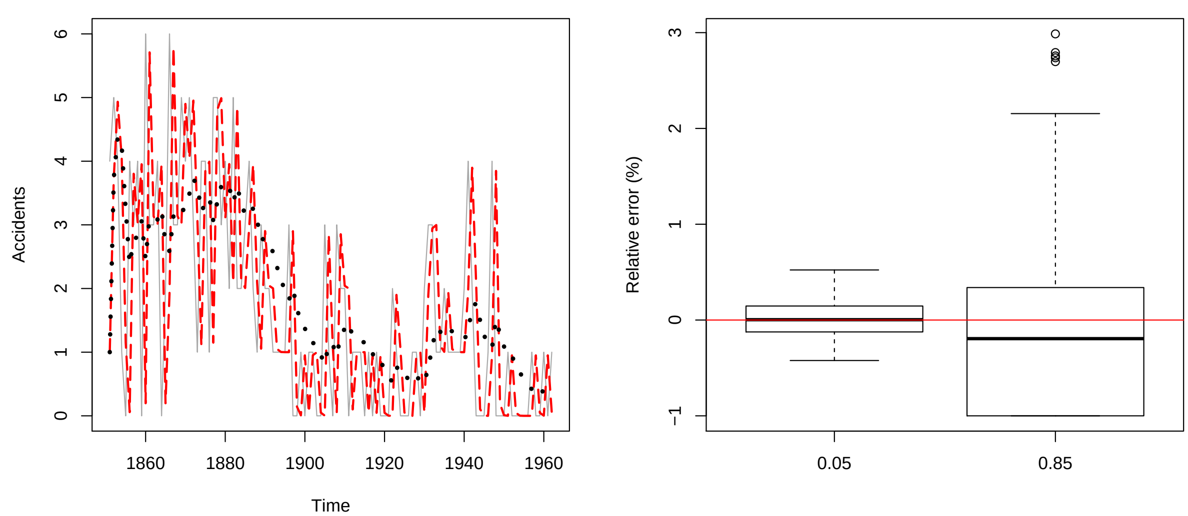


Figure 6.16: Forecast of the coal mining series with the DGLM for $\delta = 0.05$ (dashed line) and for $\delta = 0.85$ (pointed line), and its relative forecast error (right).

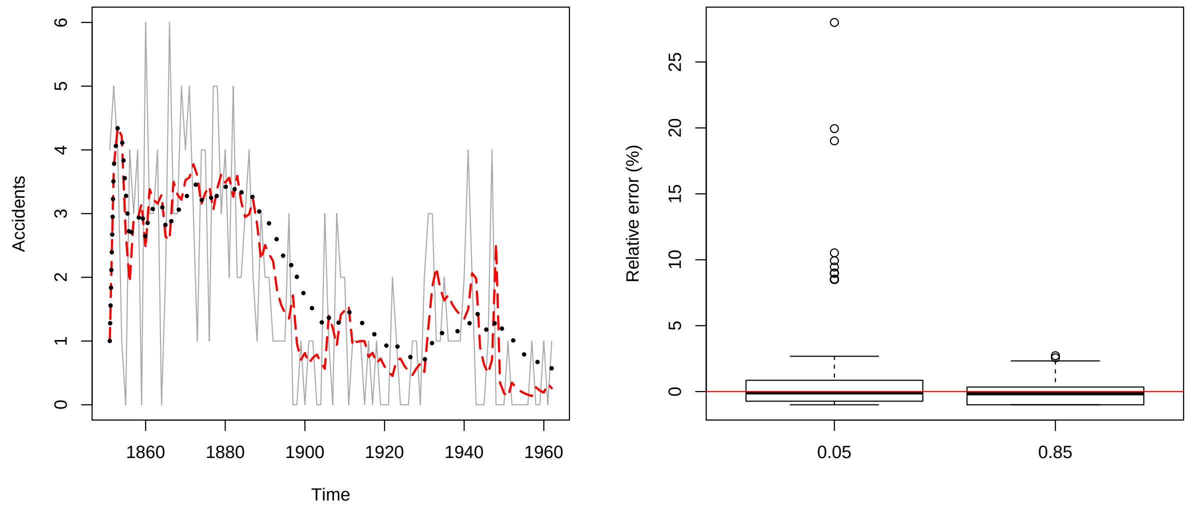


Figure 6.17: Forecast of the coal mining series with the DGLM via PPM for $\delta = 0.05$ (dashed line) and for $\delta = 0.85$ (pointed line), and its relative forecast error (right).

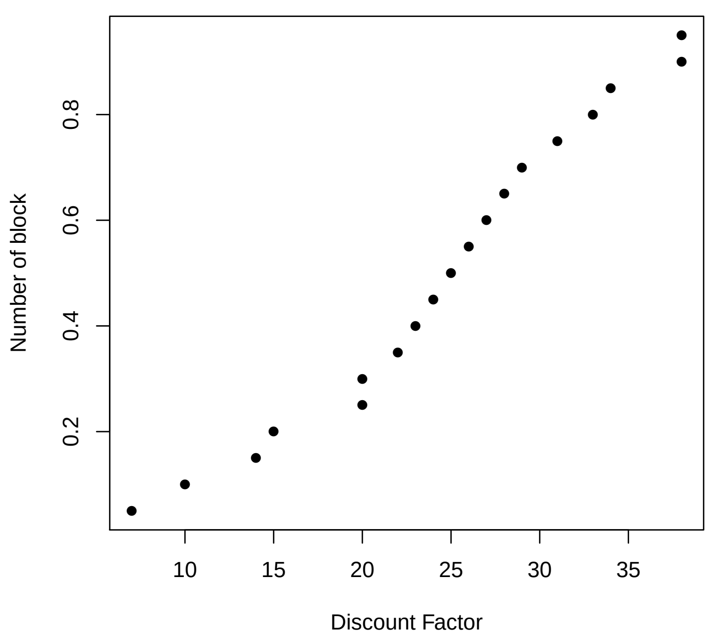


Figure 6.18: Most probable number of blocks for the discount factor considered with the Coal Mining series

We explore the behavior of the most probable number of blocks according to the choice of the discount factor in Figure 6.18 by plotting the this most probable number of block for both discount factor consider. We note that as higher is the value of the discount factor, higher is the most probable number of blocks in the partition.

The change point probability for both discount factors considered in the DGLM via PPM is displayed in Figure 6.20, in order to investigate more about the partitions provided. We observe that the model, for $\delta = 0.05$ gives more change point probability for some observations, but having a confusing behavior and not given more probability where it was expected, around the year of 1886. The high variability of the data could be a confounding factor in this case. In the case for $\delta = 0.85$ we observe a peak of highest probabilities around the year of 1886, as expected. Another peak around the year of 1950 is also detected, that in fact, may be noted when looking the coal mining series.

From Figure 6.20 we already verified that the discount factor seems to have some impact in the change point probability, as a consequence of its impact in the posterior partition. In order to investigated the behavior of this impact we plotted in Figure 6.21 the change point probability for some different discount factors. We note that the same confused behavior of the change point probability, already mentioned when $\delta = 0.05$, is observed for $\delta \leq 0.50$. In these cases, we can not affirm, for sure, which observation could be a change point. From $\delta \geq 0.70$ we begin to observe a more stable behavior, with a major peak of probabilities around the year of 1886, and a smaller peak around the year of 1950.

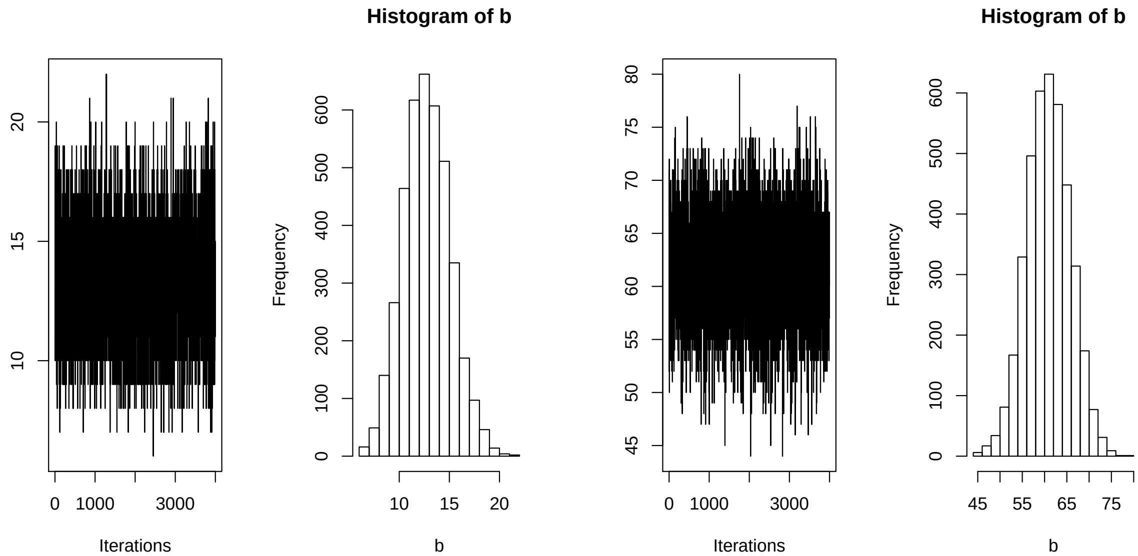


Figure 6.19: Posterior distribution of the number of intervals associated with the DGLM via PPM for $\delta = 0.05$ (left) and $\delta = 0.85$ (right).

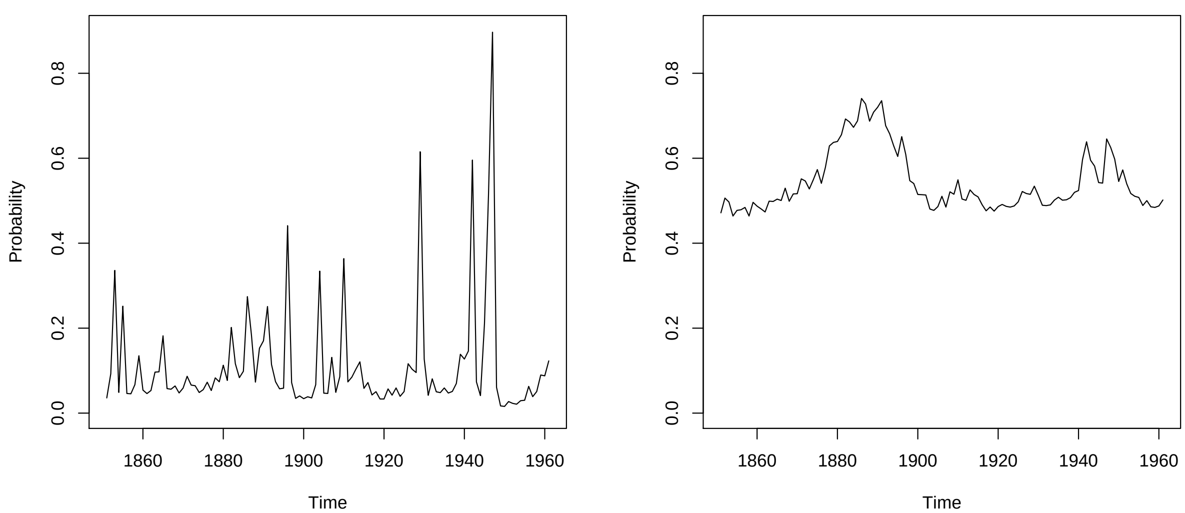
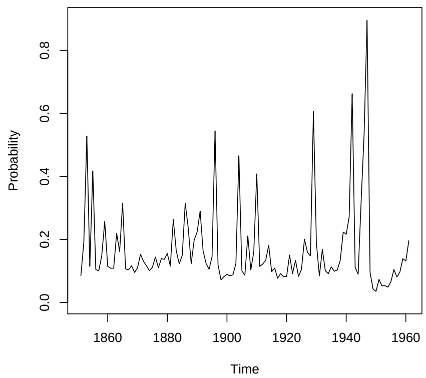
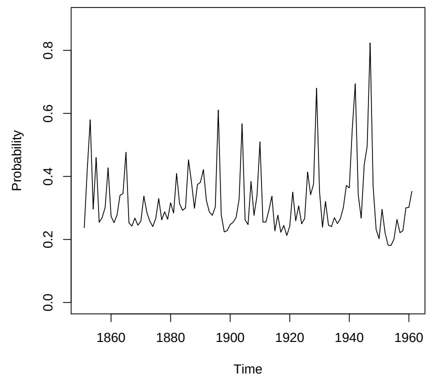


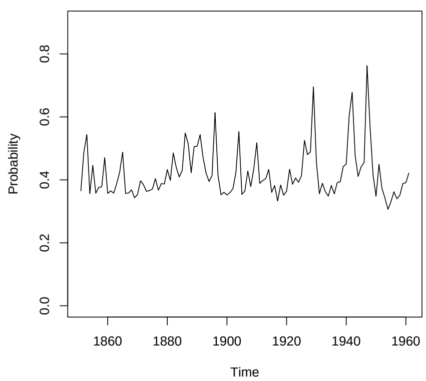
Figure 6.20: Change point probability for $\delta = 0.05$ (left) and $\delta = 0.85$ (right).



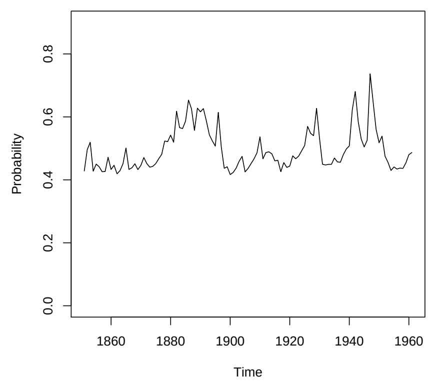
(a) Change point probability for $\delta = 0.10$.



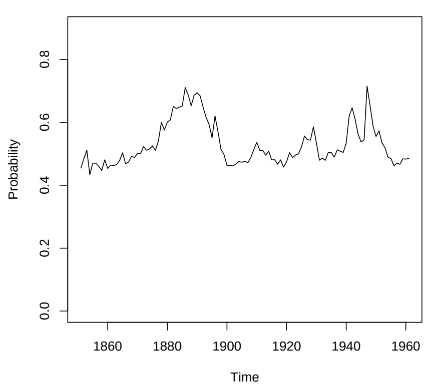
(b) Change point probability for $\delta = 0.30$.



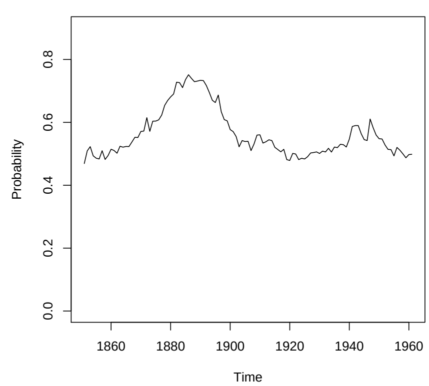
(c) Change point probability for $\delta = 0.50$.



(d) Change point probability for $\delta = 0.70$.



(e) Change point probability for $\delta = 0.80$.



(f) Change point probability for $\delta = 0.90$.

Figure 6.21: Change point probability for different values of discount factors with the coal mining data.

Chapter 7

Final Remarks

The class of dynamic generalized linear model is an attractive extension of the dynamic linear model, since it accommodates time series whose distribution is a member of the uni-parametric exponential family. In this work we extended this class by using the product partition model, aiming to accommodate time series in the exponential family with change point problems. The greatest attractive expected in our model is the inference about the partition, allowing us to construct blocks of observations, identify change points and evaluate the probability of the observation be a change point.

In order to investigate the usefulness of the proposed model we analyzed two real data sets, that have already been discussed in the literature: the APCI time series and the coal mining time series. The APCI series, which we assumed to be Gaussian distributed, has an expected intervention, in the year of 2002, due to the election of the President Lula, and the coal mining series is distributed as the Poisson, has an expect intervention around the year of 1856. We performed a sensitivity analysis aiming to evaluate the impact of the discount factor on the model fitting in our applications. Was detected in the Gaussian and the Poisson cases a superiority of the DGLM via PPM over the PPM, and that the discount factor had a great impact on the inference about the partition.

In this work we assumed that the prior cohesion of the partition is a discrete uniform. We intend to provide a sensitivity analysis for the choice of the prior cohesion in order to evaluate its impact in the partitions provided in the DGLM via PPM class.

The computational coast of the proposed class was expensive. Then, we wish to provide a R package with the already implemented methodology, using the C++ language, in order to improve the computational time. The development of this package will allow us to provide a study with simulated data in order to corroborate the results obtained in the analysis with real time series data.

We expect in future works to propose variations and extensions of the already defined

DGLM via PPM. These new versions would keep the inferential gain already reached and provide more flexibility. We briefly describe these proposals as follows.

We may be interested in add a correlation structure within the parameters in the same block. So, we intend to define an alternative approach by passing the evolution equation after the proceed information of each observation and providing a parametric evolution within the block. We intend to provide a non-linear version of the DGLM via PPM using the first order Taylor series approximations for inference, and providing the clustering structure for the observations via the product partition model aiming to obtain the same inferential gain of the DGLM via PPM added to some flexibility. We also intend to provide a similar work of the DGLM via PPM, considering the non-Gaussian family of state-space models (NGSSM) aiming to obtain a competitive class of models. Besides, the NGSSM is class of models of scale for volatility data, making possible to treat change point problems in the volatility. We may be interested in this when the time series presents change points or jumps also in the volatility, not only in the mean. This would implies in an arduous work in the DGLM framework, but not with the NGSSM.

Bibliography

Barry, D. and Hartigan, J. A. (1992). Product partition models for change point problems. *The Annals of Statistics*, 20:260–279.

Barry, D. and Hartigan, J. A. (1993). A bayesian analysis for change point problems. *Journal of the American Statistical Association*, 88:309–319.

Box, G. E. P. and Jenkins, G. M. (1970). Time series analysis: Forecasting and control. *San Francisco, CA: Holden-Day*, page 575.

Box, G. E. P. and Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*, 70:70–79.

Carlin, P. B., Polson, N. G., and Stoffer, D. S. (1992). A monte carlo approach to nonnormal and nonlinear state-space modeling. *Journal of the American Statistical Association*, 87:493–500.

Demarqui, F. N., Loschi, R. H., Dey, D. K., and Colosimo, E. A. (2012). A class of dynamic piecewise exponential models with random time grid. *Journal of Statistical Planning and Inference*, 142, Issue 3:728–742.

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

Fearnhead, P. (2006). Exact and efficient bayesian inference for multiple changepoint problems. *Statistics and Computing*, 16:203–213.

Ferreira, J. A., Loschi, R. H., and Costa, M. A. (2014). Detecting changes in time series: A product partition model with across-cluster correlation. *Signal Processing*, 96, Part B:212–227.

Gamerman, D. (1992). A dynamic approach to the statistical analysis of point processes. *Biometrika*, 79:39–50.

Gamerman, D., dos Santos, T. R., and Franco, G. C. (2013). A non-gausian family of state-space models with exact marginal likelihood. *Journal of Time Series Analysis*, 34:625–645.

Harrison, P. J. and Stevens, C. F. (1976). Bayesian forecasting. *J. R. Statist. Soc. A*, page 205.

Hartigan, J. A. (1990). Partition models. *Communications in Statistics - Theory and Methods*, 19:2745–2756.

Kass, R. E. and Raftery, A. E. (1995). Bayes factor. *Journal of the American Statistical Association*, 90:773–795.

Loschi, R. H. and Cruz, F. R. B. (2002). An analysis of the influence of some prior specification of change points via product partition model. *Computational Statistics & Data Analysis*, 39:477–501.

Loschi, R. H., Cruz, F. R. B., Takahashi, R. H. C., Iglesias, P. L., Arellano-Valle, R. B., and Smith, J. M. (2005). A note on bayesian identification of change points on data sequence. *Computers & Operations Research*, 35, Issue 1:156–170.

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*. Chapman & Hall/CRC.

Meinhold, R. J. and Singpurwalla, N. (1989). Robustification of kalman filter models. *Journal of the American Statistical Association*, 84:479–486.

Migon, H. S., Gamerman, D., Lopes, H. F., and Ferreira, M. A. R. (2005). Dynamic models. *Handbook of Statistics*, 25:557 – 592.

Monteiro, J. V. D., Assunção, R. M., and Loschi, R. H. (2011). Product partition models with correlated parameters. *Bayesian Analysis*, 6, Number 4:691–726.

Muller, P., Quintana, F., and Rosner, G. L. (2011). A product partition model with regression on covariates. *Journal of Computational and Graphical Statistics*, 20:260–278.

Nelder, A. and Wedderburn, R. W. M. (1972). Generalized linear models. *J. R. Statist. Soc. A*, 135, Part 3:370.

Perron, P. (2005). Dealing with structural breaks.

Pole, A., West, M., and Harrison, J. (1994). *Applied Bayesian Forecasting and Time Series Analysis*. Chapman & Hall.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Santos, T. R., Franco, G. C., and Gamerman, D. (2010). Comparasion of classical and bayesian approaches for intervantion analysis in structural models. *International Statistical Review*, 00:1–22.

Shumway, R. H. and Stoffer, D. S. (2006). *Time Series Analysis and Its Applications*. Springer.

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

West, M., Harisson, J. P., and Migon, H. S. (1985). Dynamic generalized linear models and bayesian forecasting. *Journal of the American Statistical Association*, pages 83 – 93.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*. Springer, 2nd edition.

Worsley, K. J. (1986). Confidence regions and tests for a change-point in a sequence of exponential family random variables. *Biometrika*, 73:91–104.