

A Bayesian Joinpoint Regression model with an unknown number of break-points

Miguel A Martínez-Beneito¹, Gonzalo García-Donato^{2,1} and Diego Salmerón^{3,4}

¹ *Centro Superior de Investigación en Salud Pública, Valencia, Spain.*

² *Universidad de Castilla la Mancha, Albacete, Spain.*

³ *Servicio de Epidemiología, Consejería de Sanidad, Murcia, Spain.*

⁴ *CIBER de Epidemiología y Salud Pública (CIBERESP), Spain.*

Abstract: Joinpoint regression is used to determine the number of segments needed to adequately explain the relationship between two variables. This methodology can be widely applied to real problems but we focus on epidemiological data, the main goal being to uncover changes in the mortality time trend of a specific disease under study. Traditionally, joinpoint regression problems have paid little or no attention to the quantification of uncertainty in the estimation of the number of change-points. In this context, we found a satisfactory way to handle the problem in the Bayesian methodology. Nevertheless, this novel approach involves significant difficulties (both theoretical and practical) since it implicitly entails a model selection (or testing) problem. In this study, we face these challenges through i) a novel reparameterization of the model; ii) a conscientious definition of the prior distributions used and iii) an encompassing approach which allows the use of MCMC simulation-based techniques to derive the results. The resulting methodology is flexible enough to make it possible to consider mortality counts (for epidemiological applications) as Poisson variables. The methodology is applied to the study of annual breast cancer mortality during the period 1980-2007 in Castellón, a province in Spain.

Keywords and phrases: Bayesian statistics, Model selection, Bayes factors, Joinpoint Regression, Epidemiological time series.

1. Introduction

Joinpoint Regression is a statistical modeling technique that explains the relationship between two variables by means of a segmented linear regression constrained to be continuous everywhere, in particular in those places where the slope of the regression function changes. This technique is widely applied to the modeling of time trends in mortality or incidence series in epidemiological studies. In these applications, the number (if any) and the location of the

*Contract/grant sponsor: Ministerio de Educación y Ciencia; Contract grant numbers: MTM2007-61554, MTM2010-19528.

†Contract/grant sponsor: Fondo de Investigaciones Sanitarias. Instituto de Salud Carlos III; Contract grant number: ISCIII06-PI1742.

changes in trends (known as change-points or joinpoints) is usually unknown, the main goal being to assess their existence and determine their location. Joinpoint Regression is applied in a wide variety of contexts. Nevertheless, for clarity of presentation of the main ideas in this paper, the above-mentioned epidemiological setting is assumed throughout the paper.

The underlying problem in this context is a model selection (or testing) problem, with uncertainty about which is the model (number of change-points) that (most likely) produced the data observed. In this paper, we approach the problem from a Bayesian perspective which is fully detailed below. Nevertheless, the predominant techniques in the context of Joinpoint Regression are frequentist. Within them, the main goal is to find “the smallest number of joinpoints such that, if one more joinpoint is added, the improvement is not statistically significant” (Statistical Research and Applications Branch, National Cancer Institute (2009)). This goal is usually met by means of non-parametric permutation tests (Kim et al. (2000)), the final result of that analysis being the determination of the model that meets the former condition and the estimation of its parameters and their variability. The National Cancer Institute (NCI) of the United States has developed a tool to carry out these kinds of analyses (Statistical Research and Applications Branch, National Cancer Institute (2009)). This software has become a standard tool in epidemiological literature, see for example: Cayuela et al. (2004); Bosetti et al. (2005); Stracci et al. (2007); Karim-Kos et al. (2008); Bosetti et al. (2008); Qiu et al. (2009).

The proposed permutation test approach has, in our opinion, two main limitations:

- (i) The underlying model selection criterion selects the simplest model such that if a new joinpoint is added, it does not yield a statistically significant improvement. Therefore, the model selected is not the most likely one but it is chosen according to both its capacity to describe the time trend within the data and the informativeness of data to highlight temporal changes (in the case of mortality or disease incidence time series this last feature will depend heavily on the average number of annual observed cases in the dataset). Hence, in a situation when the data is not very informative, the criteria is clearly biased (by definition) towards more simplistic models. This conservative behavior of the permutation procedure could seem reasonable to some authors (Kim, Yu and Feuer (2009)), but we find this systematic tendency unsatisfactory. Instead, in this situation we would expect to be “informed” that a number of alternative models are equally plausible.
- (ii) It is hard to quantify to what extent one selected model is more likely than others. In practice, a single model is chosen, with a fixed number of joinpoints, regardless of whether the choice is much more likely or not than the other alternatives. As a consequence, the main goal of the inference in these kinds of models (how many joinpoints can adequately explain the time trend that we are observing?) lacks an estimate of its variability in contrast to the remaining parameters of the selected model. This is because

the asymptotics of the number of joinpoints is an involved issue. Relevant advances in this regard include the work of Yao (1988); Liu, Wu and Zidek (1997); Kim, Yu and Feuer (2009), focusing mainly on conditions where estimators of this parameter are consistent (converge to the parameter as the sample size increases).

There is some previous work devoted to the application of Bayesian ideas to Joinpoint problems. Carlin, Gelfand and Smith (1992) is one of the pioneering contributions in this area, proposing the application of MCMC methods to fit these kinds of models. Moreno, Torres and Casella (2005) derived the intrinsic priors for a possibly heteroscedastic normal model under the assumption of a fixed change-point. In a similar problem but under homoscedasticity, more recently Bayarri and García-Donato (2007), explicitly derived the Zellner-Siow priors (Zellner and Siow (1980); Zellner (1984)). Bearing in mind its epidemiological application, Tiwari et al. (2005) describe the calculus of Bayes factors for model selection in Gaussian Joinpoint models. A common limitation of the application of all these studies to epidemiological time series modeling is the assumption of normal errors. This restriction is relaxed in the work of Ghosh, Basu and Tiwari (2009), who propose semi-parametric regression models by means of Dirichlet processes. Nevertheless, the original data in mortality studies are usually the annual observed death counts, for which in this paper we assume a Poisson regression model to take into account the discrete nature of these data. As far as we know, the only Bayesian model selection approach that considers Poisson counts in the context of joinpoint regression is Ghosh et al. (2009). These authors also acknowledge the advantage of the Poisson assumption specially when the observed counts are smaller, in which case the Gaussian modeling of incidence or mortality rates is clearly not convenient. That is probably the work most closely related to ours. Nevertheless, our approach differs from it in at least two main ways. First, Ghosh et al. (2009) model the hazard rate in a context of relative survival, while we propose a model for the, more usual in this context, incidence or death rates (see Section 3.1). Second, the model selection performed in Ghosh et al. (2009) is based on popular model selection criteria like CPO and DIC, but not on posterior probabilities as we do. The advantages of posterior probabilities over other criteria are a straightforward interpretation and the richness of the results produced (eg. to produce predictions). These are put in practice in Section 4.

The Bayesian approach is straightforward at least conceptually. Bayes factors allow the selection of models from among several alternatives, strictly according to their posterior probabilities. Furthermore, through Bayes factors it is possible not only to select one of the models entertained, according to the evidence (posterior probability) provided by the data, but to quantify the difference between the one selected and the remaining competing models. In a broad sense, Bayes factors make it possible to evaluate the uncertainty involved with the selection made. Furthermore, their use is the basis for what has been called Model Averaging (Clyde (1999)), under which it is possible to average the fit of all the models weighted by their posterior probabilities. Therefore, uncertainty in the

selection of the “correct” model is propagated to the inferential exercise.

The goal of this study is to propose a Joinpoint Regression modeling that evaluates and incorporates the uncertainty in both model selection and model parameters into the analysis. We found the Bayesian approach, for the aforementioned reasons, a very appealing way of doing so. Of course, the Bayesian approach to model selection problems is not free of difficulties, nicely explained in Berger and Pericchi (2001). These can basically be summarized in two main problems: a strong influence of the prior on the results and a very challenging numerical problem to compute these results. Much of the material presented in this paper focuses on how we manage to overcome these problems.

The rest of the paper is organized as follows: Section 2 introduces a reparameterization of the usual Joinpoint Regression model which will be really convenient as a first step to assign prior distributions. In Section 3 we introduce a Joinpoint modeling proposal with an unknown number of change-points. That proposal will be carried out as a variable selection process (George and McCulloch (1993); Dellaportas, Forster and Ntzoufras (2000, 2002)), and prior distributions will be discussed in detail in order to get reasonable results from the previous model selection problem. In Section 4 our new model will be applied to the study of breast cancer mortality in the Spanish province of Castellón to illustrate the possibilities of the Bayesian approach for exploiting the results from the inference. Finally, in Section 5 we will summarize the main advances of our model and some future lines of work will be outlined.

2. A convenient parameterization of the Joinpoints

Suppose that we want to describe the behavior of the variable $\{Y_i; i = 1, \dots, n\}$ as a function of the explanatory covariate $\{t_i, i = 1, \dots, n\}$. We find it convenient to assume that t_i represents time, although it could represent any other magnitude. It is usual to model the presence of J change-points (locations in which the functional describing the relationship between Y and t changes) through the expected value of the dependent variable as

$$g(E(Y_i|t_i)) = \alpha + \beta_0 \cdot t_i + \sum_{j=1}^J \beta_j \cdot (t_i - \tau_j)^+, \quad (1)$$

for certain g , a linking function. In this equation, τ_j represents the location of the j -th change-point and $(\theta)^+$ is θ if $\theta > 0$, and 0 otherwise. We label the model in (1) as M_J , to make it explicit that it contains exactly J change-points. Similarly, we call the model with no change-points M_0 .

In what follows, we assume that the maximum number of change-points is J^* , a number which is fixed (more on this aspect later). Then the problem that we face is to find the posterior probability of each of the models in $\{M_0, M_1, \dots, M_{J^*}\}$ and in the case that a single model needs to be selected, the one with the highest posterior probability is preferred.

In the Bayesian framework, the distinction between common and new parameters for the assignment of prior distributions (first used by Jeffreys (1961)) is

crucial. The terminology is very intuitive: common parameters appear in all the competing models while the new parameters are model-specific. In the problem above, α and β_0 are common and the remaining β 's are new parameters.

As we fully describe in the following section, the scheme that we adopt for the assignment of the priors proposes the same marginal (non informative) prior for the common parameters. Clearly, using the same prior (either subjectively elicited or not) for common parameters does not make sense unless the meaning of these parameters does not change throughout the different models entertained. Unfortunately, common parameters do not (in general) represent the same magnitude, this being a major difficulty in Bayesian model selection (see Berger and Pericchi (2001)). Although it is hard to establish precise conditions under which common parameters have the same meaning (a notion which can be viewed as subjective), it is clear that, as it stands, this is not the case for the problem presented above. For instance, in M_0 , α and β_0 represent the parameters explaining the global trend of the series. Nonetheless, in M_1 (a model with one joinpoint) α and β_0 are the intercept and the slope respectively of a line adjusted *up to* the change-point (τ_1), $\beta_0 + \beta_1$ being the slope from the joinpoint on. We propose an alternative parameterization, for which we argue that the hypothesis of common parameters with a same meaning across models holds reasonably. More concisely, let

$$E(Y_i|t_i) = \alpha + \beta_0 \cdot (t_i - \bar{t}) + \sum_{j=1}^J \beta_j \mathcal{B}_{\tau_j}(t_i) \quad (2)$$

where $\mathcal{B}_{\tau_j}(t)$, which we call the *break-point* or the *break-point centered at τ_j* , is defined as the following piecewise linear function

$$\mathcal{B}_{\tau_j}(t) = \begin{cases} a_{0j} + b_{0j} \cdot t & \forall t \leq \tau_j \\ a_{1j} + b_{1j} \cdot t & \forall t > \tau_j \end{cases}$$

restricted to a number of conditions which are fully specified below. What we finally obtain is, conditioned on the joinpoints, a known function of the original regressors at times $\{t_1, t_2, \dots, t_n\}$ and describing a peak at the moment τ_j . Alternatively, the set of break-points can be interpreted as a base of continuous piecewise linear functions generating the joinpoints needed to describe the time trend in the data. The linear component in (2) has been introduced as $(t_i - \bar{t})$ (instead of t_i) to avoid the dependency between α and β_0 .

The conditions imposed on $\mathcal{B}_{\tau_j}(t)$ are:

- $\lim_{t \rightarrow \tau_j^-} \mathcal{B}_{\tau_j}(t) = \lim_{t \rightarrow \tau_j^+} \mathcal{B}_{\tau_j}(t)$; that is, $\mathcal{B}_{\tau_j}(t)$ is a continuous function. Hence, it is guaranteed that the regression function (2) is continuous all around, independently of the number of break-points.
- $\sum_{i=1}^n \mathcal{B}_{\tau_j}(t_i) = 0$; that is, the sum of the elements of the break-point evaluated in all the points observed must be zero. This way the addition of any break-point in the model would not alter the mean value of the regression function and it would not change the meaning and the estimation of parameter α across models. In other words, this condition can be understood

as if the new break-points are imposed to be geometrically orthogonal to the intercept term.

- $\sum_{i=1}^n \mathcal{B}_{\tau_j}(t_i) \cdot t_i = 0$; that is, the slope of the break-points along the whole period of study is zero. This way the addition of any break-points in the model would neither alter the slope of the regression function nor would it change the meaning and the estimate of parameters β_0 across models. In other words, this condition can be understood as if the new break-points are imposed to be geometrically orthogonal to the slope term.
- $\mathcal{B}_{\tau_j}(\tau_j) = 1$; so that the parameter β_j has the role of measuring the magnitude of the break-point in the location where the change in the tendency takes place. Hence, β_j has to be interpreted as the value of the deviation produced at τ_j as a consequence of including this break-point in the model. Without a restriction of this kind the value of β_j would not be identifiable.

Subject to these conditions and given τ_j , the function $\mathcal{B}_{\tau_j}(t)$ is unambiguously determined (details are provided as supplementary material).

The main motivation for introducing the above basis of functions is to improve robustness over prior specification by reparameterization as is further explained below. In this sense, our approach is different from other basis of functions used (like MARS in Friedman (1991) and Nott, Kuk and Duc (2005)) where the aim was to approximate surfaces which could potentially be highly non linear on their multiple arguments. With this parameterization, it is now reasonable to assume that the common parameters have the same meaning: in all models, α and β_0 are parameters of a line representing the global trend. The remaining parameters are used to modify this common line to incorporate changes in the trend without changing the original meaning of α and β_0 in M_0 .

As an illustration, a graphical representation of several break-points, corresponding to different locations of the change-points (for 20 observed values), is shown in Figure 1 (left). As can clearly be seen, all break-points are equal to 1 at their corresponding change-points, are zero on average and their slope is also null. Following this same context, Figure 1 (right) shows how the inclusion of break-points modifies a straight line common to all models ($\alpha + \beta_0 \cdot t$). In this figure we have plotted the line $y(t) = 10 + 0.1 \cdot t$, jointly with the functions $y(t) = 10 + 0.1 \cdot t + 0.5 \cdot \mathcal{B}_5(t)$ and $y(t) = 10 + 0.1 \cdot t + 0.5 \cdot \mathcal{B}_5(t) - 0.3 \cdot \mathcal{B}_{13}(t)$. The effect of including break-points over a common regression function is the torsion of this function, instead of a divergence from a certain place as was done in the parameterization in (1). In other words, the effect of the introduced break-points is a global modification of the regression curve, as opposed to the previous parameterization which produced a local modification, changing the meaning of the parameters α and β_0 depending on whether the break-points are (or are not) included in the model and how many of them are included.

Additionally, a number of nice side effects are derived from this parameterization:

- With the definition of the break-points, the columns of the design matrix corresponding to the common parameters are orthogonal to the columns of the new parameters. Hence, we can reasonably expect the Fisher in-

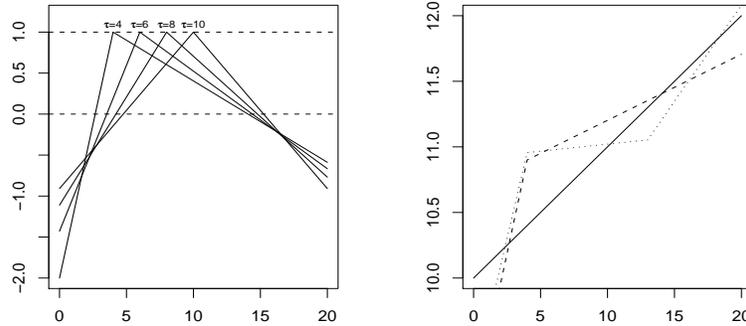


Figure 1: Left hand side: Shape of different break-points as a function of their change-point locations. Right hand side: Trends with 0 (solid line), 1 (dashed line) and 2 (dotted line) change-points. All these functions have the same mean value and slope for the whole period.

formation to be approximately block diagonal (of course, it depends on the particular distribution assumed for Y , and will be true for the normal case). In this scenario, it is known that the Bayes factors (and consequently any other result derived from them, like the posterior probabilities) are quite robust to the prior distribution used for common parameters (see Jeffreys (1961); Kass and Vaidyanathan (1992)). This has been used as a justification for the strategy of using non informative priors, possibly improper, for common parameters (see Liang et al. (2008); Bayarri and García-Donato (2008)). We use a similar approach, as we explain in the next section.

- With this reparameterization, the common parameters α and β_0 conserve their meaning regardless of the model (i.e. regardless the number of joinpoints). This allows us to make inferences about them in a model-averaged way, that is, taking into account the uncertainty regarding which is the true model. We put this into practice in Section 4 where we draw inferences on these parameters.

3. A Joinpoint Regression model with an unknown number of change-points

3.1. An encompassing model

To address the question of how many joinpoints are needed, or equivalently to select from among $\{M_0, M_1, \dots, M_{J^*}\}$, we introduce an encompassing model

in which all these models are nested. We also assume the scenario of modeling epidemiological series of mortality or incidence counts of a certain disease, since this is the most extended application of joinpoint models. Nevertheless, the main ideas that we present hold for other types of data, not necessarily counting data.

Let Y_i be the number of cases of a specific disease observed during a period of time, represented by t_i . To account for the discrete nature of these values, we suppose that

$$Y_i \sim \text{Poisson}(\mu_i) \quad i = 1, \dots, n$$

whose mean is defined as:

$$\log(\mu_i) = \log(P_i) + \alpha + \beta_0 \cdot (t_i - \bar{t}) + \sum_{j=1}^{J^*} \delta_j \cdot (\beta_j \cdot \mathcal{B}_{\tau_j}(t_i)) \quad (3)$$

where P_i is the population under study during year i , $\{\delta_j; j = 1, \dots, J^*\}$ are binary variables which include (for $\delta_j = 1$) or exclude (for $\delta_j = 0$) each change-point in the model and τ_j is the unknown location of the j -th break-point. This model can be interpreted as an encompassing model, which contains all possible joinpoint models unambiguously identified through $\boldsymbol{\delta} = (\delta_1, \dots, \delta_{J^*})$, to explain the data. In particular, $M_0 \equiv \{\boldsymbol{\delta} = \mathbf{0}\}$, $M_1 \equiv \{\boldsymbol{\delta} : \sum_{j=1}^{J^*} \delta_j = 1\}$, ..., and $M_{J^*} \equiv \{\boldsymbol{\delta} = \mathbf{1}\}$. Hence, the posterior probability over the model space $\{M_0, M_1, \dots, M_{J^*}\}$ is completely specified through the posterior distribution of $\boldsymbol{\delta}$. A detailed description of the computation of $\pi(\boldsymbol{\delta} | \mathbf{y})$ is provided as supplementary material. The proposal in (3) can in a way be seen as an order-2 regression spline (Hastie, Tibshirani and Friedman (2009)) as this is a piecewise linear continuous regression function.

To avoid identifiability problems we impose a number of restrictions on the locations of the change-points on which there is broad consensus in the related literature. These are imposed so as to ensure a minimum distance between change-points and a restriction of order. In particular, we assume that the parametric space for $\boldsymbol{\tau} = (\tau_1, \tau_2, \dots, \tau_{J^*})$, which we call Ω is:

$$\Omega = \{(\tau_1, \dots, \tau_{J^*}) : t_1 + d < \tau_1, \tau_1 + d < \tau_2, \tau_2 + d < \tau_3, \dots, \tau_{J^*} + d < t_n\}.$$

With this assumption, there is a minimum number of periods d between any two change-points. By default, we use $d = 2$ in our applications. Notice that this restriction avoids the existence of two or more change-points between consecutive observations. Other ways of implementing this restriction have been proposed, as for example in Ghosh, Basu and Tiwari (2009). Notice that Ω is a bounded set in \mathcal{R}^{J^*} .

3.2. The prior distribution

Let $\pi(\alpha, \beta_0, \boldsymbol{\delta}, \boldsymbol{\tau}, \boldsymbol{\beta})$ be the prior distribution of the parameters in the model (3). We express this prior as

$$\pi(\alpha, \beta_0, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\delta}) = \pi(\boldsymbol{\beta} | \alpha, \beta_0, \boldsymbol{\delta}, \boldsymbol{\tau})\pi(\alpha, \beta_0)\pi(\boldsymbol{\tau})\pi(\boldsymbol{\delta}),$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_{J^*})$.

The prior $\pi(\alpha, \beta_0)$ corresponds to common (with the same meaning, as argued in the previous section) parameters in all models. Under this condition, it is common to use a non informative prior (see eg. Berger and Pericchi (2001); Bayarri and García-Donato (2007, 2008)), in this case the constant prior: $\pi(\alpha, \beta_0) \propto 1$.

All the other parameters are not common and it is well-known (Berger and Pericchi (2001)) that the posterior distribution is very sensitive to their prior distribution. In particular, non informative (improper) or vague priors would produce arbitrary Bayes factors.

Our approach to assigning $\pi(\boldsymbol{\beta} \mid \alpha, \beta_0, \boldsymbol{\delta}, \boldsymbol{\tau})$, is based on an approximation of the divergence based (DB) prior, introduced by Bayarri and García-Donato (2008) as a broad generalization of the pioneering ideas of Jeffreys (1961); Zellner and Siow (1980) and Zellner (1984). When comparing two nested models, an approximation of the DB prior for the new parameters in the complex model (possibly reparameterized to range within the real line), is a heavy tailed density, centered at zero and scaled by the inverse of the corresponding block of the unitary Fisher information matrix evaluated in the simpler model. This would lead us to the proposal

$$\pi(\boldsymbol{\beta} \mid \alpha, \beta_0, \boldsymbol{\delta}, \boldsymbol{\tau}) = \text{Cauchy}_{J^*}(\boldsymbol{\beta} \mid \mathbf{0}, \boldsymbol{\Sigma}), \quad (4)$$

the matrix $\boldsymbol{\Sigma}$ to be specified. Using a hierarchical scheme this is equivalent to the proposal

$$\pi(\boldsymbol{\beta} \mid \alpha, \beta_0, \boldsymbol{\delta}, \boldsymbol{\tau}, \gamma) = \text{Normal}_{J^*}(\boldsymbol{\beta} \mid \mathbf{0}, \gamma \boldsymbol{\Sigma}), \quad \gamma \sim \text{Gamma}^{-1}(0.5, 0.5),$$

where γ acts as a ‘mixing’ parameter.

In our problem, it is easy to see that the block (corresponding to $\boldsymbol{\beta}$) of the Fisher information matrix of the encompassing model (3) evaluated at $\boldsymbol{\beta} = \mathbf{0}$ is $\mathcal{I} = \boldsymbol{\Delta} \mathbf{B}^t \mathbf{W} \mathbf{B} \boldsymbol{\Delta}$ where $\mathbf{B} = \{B_{\tau_j}(t_i)\}$ (the matrix of covariates), $\mathbf{W} = \text{diag}\{w_i\}$ with $w_i = P_i \exp(\alpha + \beta_i(t_i - \bar{t}))$ and $\boldsymbol{\Delta} = \text{diag}(\boldsymbol{\delta})$. Clearly, \mathcal{I} is not (for every $\boldsymbol{\delta}$) a positive-definite matrix so it cannot be used directly to define $\boldsymbol{\Sigma}$ above. Instead, we propose

$$\boldsymbol{\Sigma} = n(\boldsymbol{\Delta} \mathbf{B}^t \mathbf{W} \mathbf{B} \boldsymbol{\Delta} + \text{diag}(\mathbf{B}^t \mathbf{W} \mathbf{B} - \boldsymbol{\Delta} \mathbf{B}^t \mathbf{W} \mathbf{B} \boldsymbol{\Delta}))^{-1}, \quad (5)$$

which is (for every $\boldsymbol{\delta}$) a positive-definite matrix.

As we argue below, the resulting proposed prior (4) with a scale matrix as in (5) has a very interesting interpretation. Given a particular $\boldsymbol{\delta}^*$, with say $\sum \delta_i^* = J$, let $\boldsymbol{\beta}_\delta$ be the J -dimensional subvector of $\boldsymbol{\beta}$ which corresponds to the non null δ_i 's. Similarly, denote $\boldsymbol{\beta}_{\delta^c}$ as the remaining parameters in $\boldsymbol{\beta}$. Finally, \mathbf{B}_δ denotes the matrix with the columns in \mathbf{B} which corresponds to the non null δ_i 's. Hence it can easily be seen that $\boldsymbol{\beta}_\delta$ and $\boldsymbol{\beta}_{\delta^c}$ are independent, conditional on the mixing parameter γ . In fact

$$\pi(\boldsymbol{\beta} \mid \alpha, \beta_0, \boldsymbol{\delta} = \boldsymbol{\delta}^*, \boldsymbol{\tau}, \gamma) = \text{Normal}_J(\boldsymbol{\beta}_\delta \mid \mathbf{0}, \gamma n(\mathbf{B}_\delta^t \mathbf{W} \mathbf{B}_\delta)^{-1}) f(\boldsymbol{\beta}_{\delta^c} \mid \alpha, \beta_0, \boldsymbol{\tau}, \gamma),$$

where

$$f(\boldsymbol{\beta}_{\delta^c} | \alpha, \beta_0, \boldsymbol{\tau}, \gamma) = \prod_{\{i:\delta_i=0\}} \text{Normal}(\beta_i | 0, \gamma n[\mathbf{B}^t \mathbf{W} \mathbf{B}]_{ii}^{-1}).$$

For every $\boldsymbol{\delta}$, the joint prior $\pi(\boldsymbol{\beta} | \alpha, \beta_0, \boldsymbol{\delta}, \boldsymbol{\tau})$ can be seen as the product of the (approximated) DB prior for the *active* parameters in the model, times a proper density for those inactive parameters. This proper density has no effect on the corresponding Bayes factors, thereby acting as a pseudoprior (see for example: Carlin and Chib (1995); Han and Carlin (2001); Dellaportas, Forster and Ntzoufras (2002)). Nevertheless, as noticed previously in the literature, these may have a great impact on the numerical results. In our experience, partially described in the supplementary material, this particular form of the pseudoprior leads to quite satisfactory results.

With this approach, the priors are in some sense defined encompassingly. That is, instead of using one (DB) prior for each possible submodel nested in (3), implicitly defined by each $\boldsymbol{\delta}$, we use a single prior which contains all these priors. The main advantage of the resulting procedure is that it makes it possible (as we describe in the next section) to use standard estimation procedures (like MCMC) to easily solve the model selection problem, whose direct resolution usually requires the help of sophisticated numerical techniques.

The proposal for $\pi(\boldsymbol{\tau})$ is not as delicate an issue as the prior for $\boldsymbol{\beta}$. This is because the corresponding parametric space Ω is a bounded set which ensures that $\pi(\boldsymbol{\delta})$ is proper under very mild conditions. Hence, the normalizing constant is unambiguously defined. In this situation, the constant prior is clearly a reasonable default choice: $\pi(\boldsymbol{\tau}) \propto 1$ for $\boldsymbol{\tau} \in \Omega$.

Finally, we introduce our proposal for $\pi(\boldsymbol{\delta})$, which is deeply related to our prior beliefs over the model space $\{M_0, M_1, \dots, M_{J^*}\}$. One possibility would be to use independent Bernoulli distributions for δ_i with a probability of success of 0.5. Nevertheless, this would lead us to the binomial($J^*, 0.5$) distribution over the model space, clearly favoring those models with around $J^*/2$ joinpoints and also giving an undesirable important role to the fixed value J^* (which is usually posed as arbitrarily large).

Instead, we experimented with two different proposals. In the first one (which we refer to as *BayesI*), all models M_i have the same probability (i.e. $1/(J^* + 1)$), this probability being equally distributed over the same number of joinpoints. This prior distribution can be formulated in a hierarchical way as:

$$\pi(\boldsymbol{\delta} | \mathbf{p}) \propto \frac{\prod_{i=1}^{J^*} p_i^{\delta_i} (1 - p_i)^{(1 - \delta_i)}}{\binom{J^*}{\sum_{j=1}^{J^*} \delta_j}}, \quad \pi(p_i) = \text{Beta}(1/2, 1/2), \quad i = 1, \dots, J^*$$

The term in the denominator of the prior for $\boldsymbol{\delta}$ yields the same prior probability for any number of change-points in the model if $p_1 = \dots = p_{J^*} = 0.5$ (the prior expected values of these terms). This hierarchical formulation, once the

parameters \mathbf{p} are integrated out, can also be expressed as:

$$\pi_1(\boldsymbol{\delta}) = (J^* + 1)^{-1} \left(\sum_{j=1}^{J^*} \delta_j \right)^{-1}.$$

The main drawback of this proposal is that, *a priori*, the mean number of joinpoints included in the model ($J^*/2$) depends on the arbitrary quantity J^* . While this is not a problem when J^* is carefully assigned, it could be an issue when this parameter is arbitrarily assigned (as is the case in many studies). To avoid this dependency on J^* as much as possible, while keeping the essence of the first proposal, we alternatively explore a slight modification on $\pi_1(\boldsymbol{\delta})$. The prior (to be called *Bayes2*) can be defined hierarchically as:

$$\begin{aligned} \pi(\delta_i|p_i) &= \text{Bernoulli}(p_i), \quad i = 1, \dots, J^* \\ \pi(p_i) &= \text{Beta}(1/2, (J^* - 1)/2), \quad i = 1, \dots, J^*. \end{aligned}$$

With this second proposal, the prior expected number of break-points (1 break-point) does not depend on the value of J^* .

In the same way as for *Bayes1*, this prior distribution can also be expressed in a non-hierarchical way, once the p_i parameters are marginalized, as:

$$\pi_2(\boldsymbol{\delta}) = (J^*)^{-J^*} (J^* - 1)^{J^* - \sum_{j=1}^{J^*} \delta_j}.$$

Interestingly, as pointed out by an Associated Editor, for large J^* , the prior probability of zero and one joinpoint tends to $e^{-1} \approx 0.37$, emphasizing the robustness of π_2 to the choice of J^* .

4. Breast cancer mortality trend in Castellón province

We analyze the breast cancer mortality time trend in Castellón province, for the period 1980-2007. Castellón is one of the 50 provinces that make up Spain, where around 285,000 women resided in 2007 and with 62.4 women dying annually of breast cancer, on average, during that period. Two facts have occurred in those years that have presumably changed breast cancer mortality trends, which are the progressive introduction of the Breast Cancer Screening Program in that province in 1992 (Vizcaíno et al. (1998)), and the introduction of new therapies for the treatment of this disease around the world, at approximately the same time. In fact, there is some controversy over which of these two factors could have a higher impact on mortality variation (Peto (1996); Blanks et al. (2000); Berry et al. (2005)). In any case, we would expect to find at least one joinpoint on the breast cancer mortality trend studied. Moreover, changes in breast cancer mortality trends have already been described for the first half of the 1990s for other Spanish regions (Ascunce et al. (2007); Salmerón et al. (2009); Pérez Lacasta et al. (2010)).

Regarding Bayesian criteria, if we focus on the mode of the posterior distribution of the number of joinpoints, in Figure 2 it can be seen that both methods

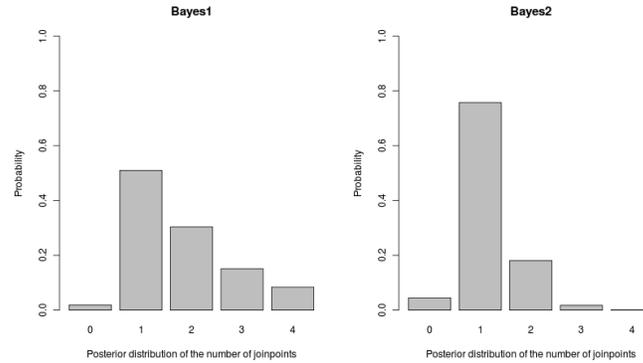


Figure 2: Posterior distribution of the number of joinpoints in breast cancer mortality trend in Castellón Province for *Bayes1* and *Bayes2* criteria

point to the existence of exactly one joinpoint, indeed *Bayes1* yields a 51.0% probability of one joinpoint while *Bayes2* estimates that probability at 73.6%. Moreover, the probability of no joinpoint in the time trend for these two models is 1.9% and 3.0% respectively. Hence the probability of a 'simply linear' trend for breast cancer mortality is low.

On the other hand we have used the *Permutation* and *BIC* criteria, by means of the NCI tool. The *Permutation* criterion (the one suggested by the NCI tool's authors) does not find any joinpoints in the period under study, possibly due to the limited information on the trend provided by the low number of deaths by year. The *BIC* criterion however, finds one joinpoint around 1995, with a 95% confidence interval: [1990,2005]. As a consequence, the conservative behavior observed of the *Permutation* criterion makes it yield results which neither agree with the remaining criteria nor with the results that we expected, based on further knowledge of this cause of mortality. It has to be acknowledged that the *Permutation* test yields a p-value of 0.0136 when testing 1 versus 0 joinpoints, whereas the significance level of the test is 0.0125 (the result of the Bonferroni correction of the original level 0.05), that is, the model without any joinpoint is really close to being rejected. Therefore, the arguments for selecting between the models with either none or just one joinpoint are not conclusive at all for this example but, on the contrary, the consequences derived from that selection are dramatic. This result warns against the danger of those criteria choosing a particular number of joinpoints and ignoring the uncertainty in that choice. In our example the permutation test would presumably have led to a wrong answer and from that moment on all the results from the analysis would have been completely misleading, as the premises of the chosen model are accepted as true and those from alternative models are completely ignored.

The left hand side of Figure 3 shows the least squares estimated trend for a different number of joinpoints (recall that the *Permutation* test chooses the

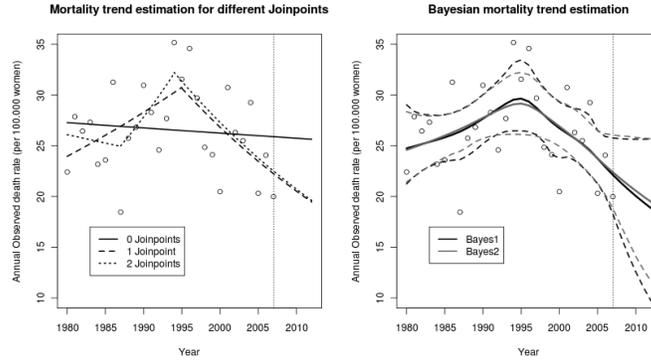


Figure 3: Left: Least Squares estimated trends for a different number of joinpoints for the annual observed death rate (per 100,000 women). Right: Estimated time trend for the annual observed death rate (per 100,000 women) for both *Bayes1* and *Bayes2*.

curve with 0 joinpoints from among all these). As can be noticed, fitted trends are quite different and the repercussion of the choice of one or another curve, and as a consequence of ignoring alternative models are dramatic. Those consequences could be even worse in the case of temporal forecasting. The former figure also shows the forecasted trend for every one of these curves for the following 5 years (2008-2012) and, as can be appreciated, predictions for the model without joinpoints diverge completely from those of the models with joinpoints. As the Permutation based criteria lacks an estimate of the probability of these scenarios, they cannot be averaged and one of them has to be chosen with the previously outlined risks. This is not the case of our proposed methods. The right hand side of Figure 3 shows the estimated trend by both *Bayes1* and *Bayes2* where predictions based on different numbers of joinpoints are averaged to provide a single answer weighting all the scenarios considered. Moreover, as pointed out in George (1999), the predictions derived with this Model Averaging procedure are optimal in several senses, the square error loss being one of them. Predicted trends for both Bayesian criteria are really close, although the *Bayes1* prediction depicts more detail than the *Bayes2* as a consequence of the higher number of joinpoints that this model occasionally considers. From now on we will focus on the results of *Bayes1* to improve the clarity of the presentation.

The estimation of β_0 from *Bayes1* yields a posterior mean of -0.0017 and a 95% probability interval [-0.0077,0.0042]. Consequently, the global trend during the whole period studied has a slope which is not very different from 0. The interest of this result is that it is valid regardless of the true number of joinpoints underlying the true trend, as we are not conditioning our results on a specific number of change-points. Conversely, the frequentist results condition all their conclusions on the selected model, and therefore they are valid if and only if

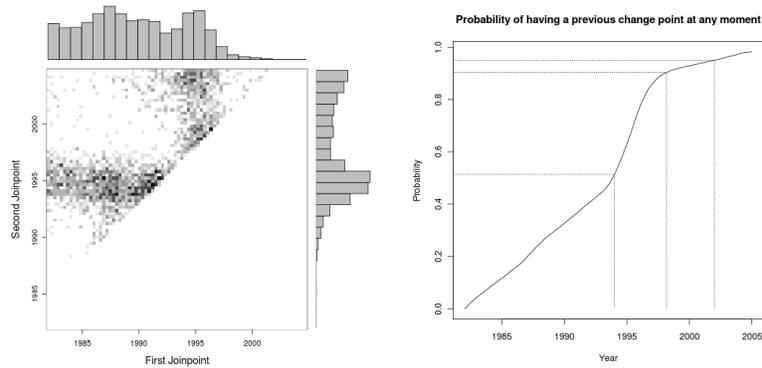


Figure 4: Left: joint posterior distribution of the location of joinpoints conditioned to the model with just 2 joinpoints. Histograms correspond to the marginal posterior distribution of the location of the first and second joinpoint. Right: probability of the trend having described a change-point before any specific moment.

that model agrees with the reality.

Figure 4 (left) shows the joint posterior distribution of the location of these two joinpoints. As can be noticed, this figure points out the existence of one joinpoint from about 1993 to 1997 and another one less precisely placed in the range from 1983 to 2005. For this second joinpoint there are several places that seem to be likely to host the second joinpoint. This information is much richer than that provided by the least squares estimate in Figure 3 that just points to 1987 as the most likely year for the second joinpoint. Focussing in the results in Figure 4 (left), our impression is that reducing the results of the 2-joinpoints-fit to a single curve (although it corresponds to the least squares fit) could be a really poor summary of all these results. Moreover, the marginal distribution of the locations of both joinpoints are multimodal and clearly asymmetric. Therefore confidence intervals obtained for these quantities under asymptotic approximations will have to be treated with caution.

Finally, Figure 4 (right) shows, at every moment of the period under study, the probability of the trend having described a change-point before that precise moment. As it is evident that the curve converges to 0.981, the complement of the probability of not describing a joinpoint along the whole period of study. Moreover, from the start of 1994 the probability of already having described a joinpoint is higher than 50%, at the start of 1998 that probability is higher than 90% and we have to wait until 2002 to be sure, with a 95% probability, of having described at least one joinpoint. This way we can also measure (without any asymptotic assumptions) the degree of certainty that we have about the trend having described a change at any moment. It is possible to derive many posterior summaries of this kind with this approach. These will be able to answer most of

the epidemiological questions that could be of particular interest in real studies.

5. Conclusions

This study has introduced a novel approach to Joinpoint analysis, taking advantage of considering the model selection problem as an inference problem on an encompassing model which contains all the candidate alternatives. Moreover, the possibility of carrying out its inference in WinBUGS is an important added value as it makes it available to a wide community of users for further applications. We also think that the reparameterization made of the original problem has also been important as it has made it possible to incorporate many previous model selection theory results. These results have given us a really valuable insight into the prior distributions of non common parameters, which is the main challenge from the Bayesian point of view, in order to give a reasonable answer to model selection problems.

Our proposal models the observed mortality as Poisson counts. This distributional assumption has several advantages: i) it can cope with zero counts without any problem; ii) no additional assumption has to be made about the variability of the observations, since it is implicitly established and iii) avoids any Gaussian assumption that may not be appropriate at all, especially when the observed counts are lower. On the other hand, the location of change-points is now considered continuous in time, as in Yu et al. (2007); Ghosh, Basu and Tiwari (2009). In our opinion this is a more realistic assumption, as if a rupture point really has existed it could have occurred at any moment in the period under study regardless of whether we have observed counts aggregated for subintervals of that same period.

A secondary question is to know how the Bayesian approach compares with the existing methods in the frequentist arena. We do so through an intensive simulation study (details provided as supplementary material). They are useful to know under what circumstances which method is expected to choose, on average and over replicated data sets, the “correct” model. What we found is that the Bayesian proposals are more sensitive (compared with the *Permutation* approach), although less specific in the detection of joinpoints. This seems to be an expected consequence of the known conservative behavior of the *Permutation* method and the probabilistic essence of Bayesian approaches. Therefore, in a context where data may not be very informative, the use of Bayesian methodology should really be encouraged. As a consequence, those methods shown in this paper could be of interest for those disease registries of a moderate size, not as big as those usually used by the National Cancer Institute of the United States.

As just outlined, statistical power is an issue of concern in joinpoint studies. This concern is even greater when covariates are considered from the frequentist approach as in those cases the original data set is usually split for independent analyses (one for every value of the covariate) and as a consequence the statistical power of every one of those subanalyses is decreased. The new proposals

introduced in this paper are model-based and hence the new covariates could also be included in our model as main effects or as an interaction with the terms already considered in the model. Indeed, from the frequentist point of view some progress has been made towards this kind of modeling in the case of having at most one joinpoint, see for example Pollán et al. (2009). In that case we would not be forced to make independent data analyses for every value of the covariate and that way we would retain the statistical power of the original analysis that did not consider any covariate. Also, this kind of analysis would be the straightforward way to analyse age standardized rates from the methodology that we have just outlined. This possibility is very attractive and will be one of the main lines of development following this study.

References

- ASCUNCE, E. N., MORENO-IRIBAS, C., BARCOS URTIAGA, A., ARDANAZ, E., EDERRA SANZ, M., J, C. and EGUÉS, N. (2007). Changes in breast cancer mortality in Navarre (Spain) after introduction of a screening programme. *Journal of Medical Screening* **14** 14-20.
- BAYARRI, M. J. and GARCÍA-DONATO, G. (2007). Extending Conventional priors for Testing General Hypotheses in Linear Models. *Biometrika* **94** 135-152.
- BAYARRI, M. J. and GARCÍA-DONATO, G. (2008). Generalization of Jeffreys' Divergence Based Priors for Bayesian Hypothesis testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70** 981-1003.
- BERGER, J. O. and PERICCHI, L. R. (2001). *Model Selection Objective Bayesian Methods for Model Selection: Introduction and Comparison* 135-207. Institute of Mathematical Statistics.
- BERRY, D. A., CRONIN, K. A., PLEVITIS, S. K., FRYBACK, D. G., CLARKE, L., ZELEN, M., MANDELBLATT, J. S., YAKOVLEV, A. Y., HABBEMA, J. D. F. and FEUER, E. J. (2005). Effect of Screening and Adjuvant Therapy on Mortality from Breast Cancer. *New England Journal of Medicine* **353** 1784-1792.
- BLANKS, R. G., MOSS, S. M., MCGAHAN, C. E., QUINN, M. J. and BABB, P. J. (2000). Effect of NHS breast screening programme on mortality from breast cancer in England and Wales, 1990-8: comparison of observed with predicted mortality. *BMJ* **321** 665-669.
- BOSETTI, C., LEVI, F., LUCCHINI, F., NEGRI, E. and VECCHIA, C. L. (2005). Lung cancer mortality in european women: recent trends and perspectives. *Annals of Oncology* **16** 1597-1604.
- BOSETTI, C., BERTUCCIO, P., LEVI, F., LUCCHINI, F., NEGRI, E. and LA VECCHIA, C. (2008). Cancer mortality in the European Union, 1970-2003, with a joinpoint analysis. *Annals of Oncology* **19** 631-640.
- CARLIN, B. P. and CHIB, S. (1995). Bayesian model choice via Markov chain Monte Carlo. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **57** 473-484.

- CARLIN, B. P., GELFAND, A. E. and SMITH, A. F. M. (1992). Hierarchical Bayesian Analysis of Change-point Problems. *Applied Statistics* **41** 389-405.
- CAYUELA, A., RODRÍGUEZ-DOMÍNGUEZ, S., LÓPEZ-CAMPOS, J. L., OTERO CANDELERIA, R. and RODRÍGUEZ MATUTES, C. (2004). Joinpoint regression analysis of lung cancer mortality, Andalusia 1975-2000. *Annals of Oncology* **15** 793-796.
- CLYDE, M. (1999). *Bayesian Statistics Bayesian Model Averaging and Model Search Strategies* 157-185. Oxford University Press.
- DELLAPORTAS, P., FORSTER, J. J. and NTZOUFRAS, I. (2000). Bayesian variable selection using the Gibbs sampler. In *Generalized Linear Models: A Bayesian Perspective* (D. K. Dey, S. K. Ghosh and B. K. Mallick, eds.) 273-286. Chemical Rubber Company Press.
- DELLAPORTAS, P., FORSTER, J. J. and NTZOUFRAS, I. (2002). On bayesian model and variable selection using MCMC. *Statistics and Computing* **12** 27-36.
- FRIEDMAN, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics* **19** 1-141.
- GEORGE, E. (1999). Discussion to Bayesian Model Averaging: a Tutorial. *Statistical Science* **14** 409-412.
- GEORGE, E. I. and MCCULLOCH, R. E. (1993). Variable Selection Via Gibbs Sampling. *Journal of the American Statistical Association* **88** 881-889.
- GHOSH, P., BASU, S. and TIWARI, R. C. (2009). Bayesian Analysis of Cancer Rates from SEER Program Using Parametric and Semiparametric JoinPoint Regression Models. *Journal of the American Statistical Association* **104** 439-452.
- GHOSH, P., HUANG, L., YU, B. and TIWARI, R. C. (2009). Semiparametric Bayesian approaches to joinpoint regression for population-based cancer survival data. *Computational Statistics and Data Analysis* **53** 4073-4082.
- HAN, C. and CARLIN, B. P. (2001). Markov Chain Monte Carlo Methods for computing Bayes factors: a comparative Review. *Journal of the American Statistical Association* **96** 1122-1132.
- HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning*. Springer-Verlag.
- JEFFREYS, H. (1961). *Theory of Probability, third edition*. Oxford University Press.
- KARIM-KOS, H. E., DE VRIES, E., SOERJOMATARAM, I., LEMMENS, V., SIESLING, S. and COEBERGH, J. W. W. (2008). Recent trends of cancer in Europe: A combined approach of incidence, survival and mortality for 17 cancer sites since the 1990s. *European Journal of Cancer* **44** 1345 - 1389.
- KASS, R. E. and VAIDYANATHAN, S. (1992). Approximate Bayes factors and orthogonal parameters, with application to testing equality of two Binomial proportions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **54** 129-144.
- KIM, H. J., YU, B. and FEUER, E. (2009). Selecting the Number of Change-Points in Segmented Line Regression. *Statistica Sinica* **19** 597-609.
- KIM, H.-J., FAY, M. P., FEUER, E. J. and MIDTHUNE, D. N. (2000). Permuta-

- tion tests for joinpoint regression with applications to cancer rates. *Statistics in Medicine* **19** 335-351.
- LIANG, F., PAULO, R., MOLINA, G., CLYDE, M. and BERGER, J. O. (2008). Mixtures of g-priors for bayesian variable selection. *Journal of the American Statistical Association* **103** 410-423.
- LIU, J., WU, S. and ZIDEK, J. (1997). On segmented multivariate regression. *Statistica Sinica* **7** 497-52.
- MORENO, E., TORRES, F. and CASELLA, G. (2005). Testing equality of regression coefficients in heteroscedastic normal regression models. *Journal of Statistical Planning and Inference* **131** 117-114.
- NOTT, D., KUK, A. and DUC, H. (2005). Efficient sampling schemes for Bayesian MARS models with many predictors. *Statistics and Computing* **15** 93-101.
- PÉREZ LACASTA, A. MARÍA JOSÉ AND GREGORI GOMIS, CARLES LAVILA, M., GISPERT MAGAROLAS, R., MARTÍNEZ ALONSO, M., VILAPRINYO TERRÉ, E., PLA FARNÓ, R. and RUÉ MONNÉ, M. (2010). Evolución de la mortalidad por cáncer de mama y diseminación de la mamografía de cribado en Cataluña. Un análisis por regiones sanitarias. *Revista Española de Salud Pública* **84** 691-703.
- PETO, R. (1996). Five Years of Tamoxifen- Or More? *Journal of the National Cancer Institute* **88** 1791-1793.
- POLLÁN, M., PASTOR-BARRIUSO, R., ARDANAZ, E., ARGÜELLES, M., MARTOS, C., GALCERÁN, J., SÁNCHEZ-PÉREZ, M. J., CHIRLAQUE, M. D., LARRAÑAGA, N., MARTÍNEZ-COBO, R., TOBALINA, M.-C., VIDAL, E., MARCOS-GRAGERA, R., MATEOS, A., GARAU, I., ROJAS-MARTÍN, M.-D., JIMENEZ, R., TORRELLA-RAMOS, A., PERUCHA, J., PÉREZ-DE RADA, M.-E., GONZÁLEZ, S., RABANAQUE, M.-J., BORRÁS, J., NAVARRO, C., HERNÁNDEZ, E., IZQUIERDO, A., LÓPEZ-ABENTE, G. and MARTÍNEZ, C. (2009). Recent Changes in Breast Cancer Incidence in Spain, 1980-2004. *Journal of the National Cancer Institute* **101** 1584-1591.
- QU, D., KATANODA, K., TOMOMI, M. and TOMOTAKA, S. (2009). A joinpoint regression analysis of long-term trends in cancer mortality in Japan (1958-2004). *International Journal of Cancer* **24** 443-448.
- SALMERÓN, D., CIRERA, L., SAEZ, M. and NAVARRO, C. (2009). Influence of the introduction of the ICD-10 on tendencies of mortality by causes (1980-2004). *Gaceta Sanitaria* **23** 144-146.
- STATISTICAL RESEARCH AND APPLICATIONS BRANCH, NATIONAL CANCER INSTITUTE, (2009). Joinpoint Regression Program, Version 3.4 Available at: <http://srab.cancer.gov/joinpoint>.
- STRACCI, F., CANOSA, A., MINELLI, L., PETRINELLI, A. M., CASSETTI, T., ROMAGNOLI, C. and LA ROSA, F. (2007). Cancer mortality trends in the Umbria Region of Italy, 1978-2004: A joinpoint regression analysis. *BMC Cancer* **7**.
- TIWARI, R. C., CRONIN, K., DAVIS, W., FEUER, E. J., YU, B. and CHIB, S. (2005). Bayesian model selection for join point regression with application to age-adjusted cancer rates. *Applied Statistics* **54** 919-939.

- VIZCAÍNO, I., SALAS, D., VILAR, J. S., RUIZ-PERALES, F., HERRANZ, C. and IBAÑEZ, J. (1998). Breast Cancer Screening Program: First Round in the Population-based Program in Valencia, Spain. *Radiology* **206** 253-260.
- YAO, Y.-C. (1988). Estimating the number of change-points via Schwarz criterion. *Statistics and Probability letters* **6** 181-189.
- YU, B., BARRET, M. J., KIM, H.-J. and FEUER, E. J. (2007). Estimating joinpoints in continuous time scale for multiple change-point models. *Computational Statistics & Data Analysis* **51** 2420-2427.
- ZELLNER, A. (1984). *Basic Issues in Econometrics* Posterior Odds Ratios for Regression Hypothesis: General Considerations and Some Specific Results 275-305. Chicago: University of Chicago Press.
- ZELLNER, A. and SIOW, A. (1980). Posterior Odds Ratio for Selected Regression Hypotheses. In *Bayesian Statistics 1* (J. M. BERNARDO, M. H. DEGRUOT, D. V. LINDLEY and A. F. M. SMITH, eds.) 585-603. Valencia: University Press.