Bayesian prediction of disability insurance frequencies using economic indicators

C. Donnelly*
Heriot-Watt University, Edinburgh, UK

Mario V. Wüthrich
ETH Zurich, RiskLab, Department of Mathematics, Switzerland

Abstract
We use economic indicators to improve the prediction of the number of incurred but not recorded disability insurance claims, assuming that there is a link between the number of claims and the chosen economic indicators. We propose a Bayesian model where we model the claims development in three directions: along incurred periods, recording lag periods and calendar periods. A stochastic model of the economic indicators is incorporated into the calendar period development direction. Thus we allow for the impact of the economic environment on the number of claims. Applying the proposed model to data, we illustrate how the inclusion of economic indicators affects the prediction of the number of incurred but not recorded disability claims.

Keywords
Bayesian model; Calendar period; Claims prediction; Credit spreads; Disability insurance; Economic indicators; Prediction uncertainty; Stochastic model; Unemployment

1 Introduction

Disability insurance is a form of insurance payable to policyholders who are unable to work either permanently or for an extended time period, due to some impairment. There can be a significant delay, called a *recording lag*, between the date an impairment occurred (the *incurred date*) and the date on which a valid claim is recorded on the insurer's administrative system. The lag can be broken down into two parts:

(i)   the time between the incurred date and the date on which a claim is received for consideration by the insurer. This delay can be due to a waiting period in the policy or uncertainty on the part of the insured on whether their current incapacity is temporary or not;

(ii) the time between the initial consideration date and the date on which the valid claim is recorded on the insurer's administrative system. This delay can be due to the time taken to assess both the validity of the claim and the extent of the work incapacity.

*Correspondence to: Catherine Donnelly, Department of Actuarial Mathematics and Statistics, and the Maxwell Institute for Mathematical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, United Kingdom. Email: C.Donnelly@hw.ac.uk

The insurer must take the recording lag into account when calculating reserves for claims which are incurred but not recorded. This prediction of incurred but not recorded claims is an important problem for the insurer. In disability insurance, it seems intuitive that claims experience should be linked to economic conditions; it is generally observed that many disability claims are a choice not to participate in the workforce, and the size and prospects of the workforce are affected by economic conditions. There are arguments as to the exact impact of the economic conditions, as highlighted in Schriek & Lewis (2010, Section 2). For example, there are arguments that disability rates should increase as the economy declines and other arguments that disability rates should increase as the economy booms. Studies have shown both these effects in different countries (see Schriek & Lewis 2010, Section 2 for references). Whatever the precise impact on individual policyholders in a particular industry sector or country, the broad message is that changes in economic conditions should be reflected by changes in the disability experience. For this reason, we model the development of incurred but not recorded claims using economic factors in addition to the information gained from the past evolution of claims. Furthermore, we might expect that there is a delay before changes in economic conditions affect policyholder behaviour, since it takes time for industry and the policyholders to recognize the effects of a different economic environment. For claims prediction, this means that we may be able to use economic indicators observed in the past, for example one year ago, in order to improve the prediction of the claims development.

The use of economic indicators for the prediction of disability claims has been examined through linear regression in two papers. In Schriek & Lewis (2010), a linear regression of South African disability rates against various economic indicators is performed, with the goal of finding if there is a link between disability claims incidence and the state of the economy. They find that unemployment and consumer confidence indicators are strongly correlated with the disability experience. In König *et al.* (2011), a Poisson model is applied in a Bayesian framework to Swiss data for the purposes of claim prediction. A strong correlation between the posterior mean of one of the fitted model parameters and the spread of corporate bonds over government bonds is found.

However, König *et al.* (2011) do not incorporate a stochastic model of the economic indicators into the claims prediction model. Consequently, while their model is simple, it is not helpful for the quantification of prediction uncertainty, which is a drawback from a risk management perspective. In particular, their approach was to develop a Bayesian claims prediction model independently of the chosen economic indicator. Then the posterior mean of a parameter of the fitted model was linearly regressed against the economic indicator. If we use the resultant linear relationship to predict future claims then, as we are forced to consider only posterior means, we lose the powerful Bayesian predictive distribution and thus are unable to quantify prediction uncertainty.

The primary aim in this paper is to propose a Bayesian model which overcomes this limitation and hence allow us to

(i)   justify the findings in König *et al.* (2011);
(ii)   stochastically project the economic indicators into the future;
(iii) consistently predict disability rates within the model; and
(iv)   quantify prediction uncertainty.

The model we propose makes the chosen economic indicator an integral part of the Bayesian model, and thus melds both insurance information and economic information into the model in a natural way. We allow one of the model parameters to be a function of the economic indicator and choose a

382

stochastic model for the future development of the economic indicator. Together, these assumptions enable us to obtain the full predictive distribution of the future claims, allowing for the impact of the economic indicator on the development of the claims. We can use the predictive distribution to calculate statistics of the future claims, calculate risk measures, such as value-at-risk or expected shortfall, and do more sophisticated analyses, for example applying extreme value theory to estimate the upper tail of the numbers of claims. We are not constrained to the use of the posterior means of the parameters in the prediction of claims, as in König *et al.* (2011). We use the claims data from König *et al.* (2011), which is income protection disability insurance for both temporary and permanent disability claims, with a minimum waiting period of three months. Our analysis found that the disability experience is strongly linked to the spread observed 1.25 years ago of corporate bond yields over government bond yields, which supports the findings of König *et al.* (2011). We also examined an unemployment indicator, but we did not find it to be helpful for claims prediction.

The idea of directly incorporating an economic indicator into a Bayesian model for claims prediction is new. Although we use a specific set of claims data to illustrate our model, our approach can be applied to other claims data where there are solid reasons for assuming that the claims are affected by economic indicators, e.g. inflation. We emphasise the importance of a well-founded argument for assuming a link between the data and the chosen indicator.

2 Notation

An annual summary of the disability claims data that we work with is shown in Table 1. The data corresponds to the calendar years 1997 to 2008 and concerns only claims for which a disability payment will eventually be made. The data $\mathcal{D}_I$ is shown as a claims development table, where the time period in which the claim was incurred is shown vertically and the lag before it was recorded on the insurer's administrative system is shown horizontally. The lower triangle is empty since this corresponds to claims which have been incurred but have not yet been recorded on the insurer's administrative system. The right-most column shows the number of insured lives in each incurred period. Our aim is to find a suitable model for the incurred but not recorded claims so that we can predict the lower triangle.

Table 1. Claims development table (annual figures).

Incurred period i	Recording lag period j												Number of policies s_i
	0	1	2	3	4	5	6	7	8	9	10	11	
0	2 016	3 560	1 049	316	130	52	16	8	4	0	0	0	480 199
1	1 774	3 660	1 049	361	97	74	70	12	4	0	0		502 661
2	2 292	3 493	1 019	405	125	62	20	12	0	8			515 803
3	1 968	4 081	1 291	426	121	31	8	0	8				536 556
4	2 511	5 070	1 598	378	70	55	12	4					582 452
5	2 850	5 933	1 504	262	90	47	16						601 253
6	3 304	5 476	1 090	285	94	47							609 116
7	2 738	5 031	1 008	320	90								591 749
8	2 617	4 297	1 242	293									600 378
9	2 086	4 457	930										622 947
10	2 144	3 746											627 236
11	2 379												669 942

We denote by $N_{i,j}$ the number of claims which were incurred in period i and were recorded by the insurer j time periods later. Thus in Table 1, $N_{5,1} = 5933$ is the number of claims which were incurred in calendar year 2002 and were recorded in calendar year 2003 on the insurer's administrative system.

We denote by I the last row of the claims development table and denote the set of observed claims by

$$\mathcal{D}_I := \{N_{i,j}; i + j \leq I, 0 \leq i, j \leq I\}. \tag{2.1}$$

For example, in Table 1, $I = 11$ and the upper triangle corresponds to $\mathcal{D}_{11}$. Correspondingly, we denote the unknown lower triangle by the complement

$$\mathcal{D}_I^c := \{N_{i,j}; i + j > I, 0 \leq i, j \leq I\}. \tag{2.2}$$

3 Bayesian models for disability prediction

In this paper, we consider Bayesian models for the disability claims data. There have been various papers written about Bayesian methods in a non-life insurance claims reserving context; for example, see de Alba (2002, 2006), de Alba & Nieto-Barajas (2002), England & Verrall (2006), Ntzoufras & Dellaportas (2002), Peters *et al.* (2009), Scollnik (2001) and Verrall (2004). Usually, models for the claims table consist of modeling the development of claims vertically (along incurred periods) and horizontally (along recording lag periods). We model additionally the development of claims diagonally (along calendar periods). The motivation is that, for disability claims, we expect the economic indicators in calendar period ℓ to impact all the claims recorded in calendar period k, for some $k \geq \ell$. This means that we must model the changes in the claims data which occur between calendar time periods.

We use the parameters

(i) $\{\pi_i; i = 0, 1, \ldots\}$ to model the incurred period direction;

(ii) $\{\gamma_j; j = 0, 1, \ldots\}$ to model the recording lag period direction; and

(iii) $\{\lambda_k; k = 0, 1, \ldots\}$ to model the calendar period direction.

3.1 A brief summary of Bayesian inference

Suppose we wish to find a model for the incurred but not recorded claims data. To combine prior information, expert judgment and the information contained in the observations in the upper triangle of the claims table in order to predict the lower triangle, we use Bayesian inference.

Based on our past experience in dealing with similar data and our subjective judgment, we decide on a model of the upper triangle $\mathcal{D}_I$ with joint density function $f(\mathcal{D}_I|\Theta)$, where Θ is a vector of unknown constants called the *parameter*. Thus, if we know Θ, the model allows us to determine the distribution of the claims data in the upper triangle.

In Bayesian inference, both the unknown parameter Θ and the data before we observed it have a joint probability distribution function. The distribution of the parameter Θ is called the *prior distribution* and here we denote its density function by g. We choose the prior distribution in accordance with our own prior subjective beliefs about the parameter Θ. Using Bayes' formula, the density of the parameter Θ conditional on seeing the data $\mathcal{D}_I$ is calculated as

$$g(\theta \mid \mathcal{D}_I) = \frac{g(\theta)f(\mathcal{D}_I \mid \Theta = \theta)}{\int g(x)f(\mathcal{D}_I \mid \Theta = x)\mathrm{d}x}. \tag{3.1}$$

We call $g(\theta|\mathcal{D}_I)$ the *posterior density*, since it captures what we know about the distribution of Θ at the point θ after seeing the data in the upper triangle $\mathcal{D}_I$. Having calculated the posterior density, we can use it to make statements about the parameter Θ, such as its mean or standard deviation. Furthermore, if we postulate a model of the lower triangle $\mathcal{D}_I^c$ with joint density function $f(\mathcal{D}_I^c|\Theta)$ as well as conditional independence between $\mathcal{D}_I$ and $\mathcal{D}_I^c$ given Θ, then we can also use the posterior density to compute the distribution of the incurred but not recorded claims $\mathcal{D}_I^c$. We call this latter distribution the *predictive distribution* and from it we can calculate statistics such as the mean and standard deviation of the incurred but not recorded claims (see Section 4 for an example of these calculations), as well as risk measures such as the value-at-risk and expected shortfall.

3.2 A Bayesian model with no economic effects

The first model we present is a Poisson-gamma-lognormal model. We use Gamma(α, β) to denote a gamma distribution with mean α/β.

Model 3.1   There exist fixed volumes $s_i > 0$, for $I = 0,1,\ldots,I$, and define the parameter vector

$$\Theta := (\pi_0,\ldots,\pi_I, \gamma_0,\ldots,\gamma_I, \lambda_0,\ldots,\lambda_{2I}).$$

Set

$$\theta_{i,j} := s_i \pi_i \gamma_j \lambda_{i+j}.$$

Then we assume

(a) the elements of Θ are mutually independent and positive almost surely; and

(b) the random variables $N_{i,j}|\Theta$ are mutually independent and Poisson distributed with mean $\theta_{i,j}$ for $i, j = 0,1,\ldots,I$, that is
$$N_{i,j}|\Theta \sim \text{Poisson}(\theta_{i,j}).$$

The prior distributions are

(c) $\pi_i \sim \text{Gamma}(\alpha_\pi, \beta_\pi)$ for $i = 0,1,\ldots, I$;

(d) $\gamma_j \sim \text{Gamma}(\alpha_\gamma, \beta_{\gamma j})$ for $j = 0,1,\ldots, I$; and

(e) $\lambda_k \sim \log\mathcal{N}(\mu_\lambda, \sigma_\lambda^2)$ for $k = 0,1,\ldots, 2I$,

for appropriate prior parameters $\alpha_\pi, \beta_\pi, \alpha_\gamma, \sigma_\lambda > 0$, $\beta_{\gamma_j} > 0$ for $j = 0,1,\ldots, I$, and $\mu_\lambda \in \mathbb{R}$.

Remark 3.2   Model 3.1 is a variation of König *et al.* (2011, Model 2.3). The difference lies in assumption (e); we use a lognormal prior distribution for the calendar period development factors λ_k whereas König *et al.* (2011) use a gamma prior distribution, selected for practical simulation reasons (we prefer to use a lognormal prior as it fits more naturally with the model of the calendar year development factors that we choose in Model 3.5 and Model 3.9 below). König *et al.* (2011) used expert judgment to specify the prior distributions. As we analyze the same data, we used the same parameters as König *et al.* (2011) for the prior distributions of π_i and γ_j, that is

$$\alpha_\pi := 225, \ \ \beta_\pi := 15000, \ \ \alpha_\gamma := 1/0.0009 \ \text{ and } \ \beta_{\gamma_j} := 1/\left(0.0009\hat{\gamma}_j^P\right),$$

where $\hat{\gamma}_j^P$ is the maximum likelihood estimate of γ_j for the Poisson Model (Remark 3.3 details how $\hat{\gamma}_j^P$ is calculated). We chose the parameters of the lognormal prior distribution so that λ_k has a mean

of 1 and a coefficient of variation of 0.2 (recall that the coefficient of variation of a random variable X is $\text{Vco}(X) = \sqrt{\text{Var}(X)}/\mathbb{E}(X)$). This results in the specifications

$$\mu_\lambda := -\frac{1}{2}\log(1.04) \quad \text{and} \quad \sigma_\lambda^2 := \log(1.04).$$

Remark 3.3 The maximum likelihood estimate $\hat{\gamma}_j^P$ used in the prior distribution of γ_j is the maximum likelihood estimate for the Poisson Model, a well-known model in non-life insurance claims reserving, for which details can be found in Denuit *et al.* (2007, Chapter 1). It is obtained iteratively by first initializing

$$\hat{p}_0 := \frac{1}{s_0}\sum_{j=0}^{I} N_{0,j} \quad \text{and} \quad \hat{\gamma}_I^P := \frac{N_{0,I}}{s_0\hat{p}_0},$$

and then iterating

$$\hat{p}_n := \frac{\sum_{j=0}^{I-n} N_{n,j}}{s_n\left(1 - \sum_{j=I-n+1}^{I} \hat{\gamma}_j^P\right)} \quad \text{and} \quad \hat{\gamma}_{I-n}^P := \frac{\sum_{i=0}^{n} N_{i,I-n}}{\sum_{i=0}^{n} s_i\hat{p}_i},$$

for each $n = 1, 2, \ldots, I$.

3.3 The disability frequency

As a measurement of the disability risk of a portfolio, the *disability frequency* is an important quantity. (Although the duration of the disability claim should also be considered for income disability insurance, as data on this was not available to us, we have ignored this second quantity in our analysis.) The disability frequency p_i is the average number of disability claims per life insured which occur in period i, that is

$$p_i := \frac{1}{s_i}\sum_{j=0}^{I} N_{i,j}.$$

To calculate the posterior mean predicted disability frequency $\hat{p}_i$, we require the posterior mean of the predicted claims. For Model 3.1 this is given by

$$\bar{N}_{i,j} := \mathbb{E}(N_{i,j}\,|\,\mathcal{D}_I) = \mathbb{E}(\theta_{i,j}\,|\,\mathcal{D}_I) = s_i\mathbb{E}\left(\pi_i\gamma_j\,|\,\mathcal{D}_I\right)\mathbb{E}(\lambda_{i+j}), \quad \text{for } i + j > I, \tag{3.2}$$

Note that, for $i + j > I$, λ_{i+j} is in the lower triangle and is thus, due to our assumptions, independent of the information $\mathcal{D}_I$. The posterior mean predicted disability frequency is then given by

$$\hat{p}_i := \frac{1}{s_i}\sum_{j=0}^{I-i} N_{i,j} + \sum_{j=I-i+1}^{I} \mathbb{E}\left(\pi_i\gamma_j\,|\,\mathcal{D}_I\right)\mathbb{E}(\lambda_{i+j}), \quad \text{for } i > 1. \tag{3.3}$$

For our calibration of Model 3.1, we have $\mathbb{E}(\lambda_k) = 1$, for $k = I+1, \ldots, 2I$. In the sequel, we propose an alternative distribution for λ_k which links it to an economic indicator. In that case, the posterior mean of the predicted claims does not decouple as in (3.2) and hence neither does the posterior mean predicted

disability frequency. This is due to the incorporation of the economic indicator data into the model, which induces a non-trivial dependence structure and changes the prediction.

3.4 Incorporating an economic indicator into the model

Here we detail an empirical Bayesian model which incorporates an economic indicator. We relate the calendar period development factors to an appropriately time-lagged economic indicator before performing a Bayesian analysis. Up to a number of future calendar periods equal to the chosen time lag, the incorporation of a time-lagged economic indicator should improve the prediction of the claim numbers compared to a model using a non-lagged economic indicator. Thus, ideally, we prefer an economic indicator which not only is a good model for the calendar period development factors, but also requires a large time lag since this allows prediction over several future periods.

Remark 3.4 In König *et al.* (2011) (see also Remark 3.2), the future calendar period development factors are assumed to satisfy

$$\lambda_k = \hat{\alpha} + \hat{\beta} S_{k-5}, \text{ for } k = I + 1, \ldots, 2I, \tag{3.4}$$

where S_k is the spread of corporate bond yields over government bond yields in period k (which corresponds to the calendar year $1997 + k/4$ in a quarter-year view). The coefficients $\hat{\alpha}$ and $\hat{\beta}$ are obtained by fitting a linear regression model to the posterior means $\{\mathbb{E}(\lambda_k | \mathcal{D}_I); k = 0, 1, \ldots, I\}$ and lagged credit spreads, with the lag of 5 quarter years determined as the lag which maximizes the empirical correlation of the posterior means and credit spreads.

3.4.1 A first model incorporating an economic indicator

Model 3.5 Assume Model 3.1 but with the additional assumptions that we are given a scalar factor $\rho \in [0,1]$, a fixed time lag $\Delta \in \{0, 1, 2, \ldots\}$ and replace assumption (e) with the following two assumptions:

(e′) we are given a series $(S_k)_{k \in \mathbb{Z}}$ of economic indicators which follow a random walk:

$$S_k = S_{k-1} + \varepsilon_k, \ \varepsilon_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(-\alpha \sigma_\varepsilon^2/2, \sigma_\varepsilon^2), \text{ for all } k \in \mathbb{Z}.$$

We set the variance σ_ε^2 equal to the sample variance of the observed economic indicator data $(S_k)_{k \le I}$. The calculation of the scaling factor α is detailed in the next assumption.

(f′) We set

$$\ln(\lambda_k) := \sqrt{1-\rho} \ln\left(\lambda_k^{(1)}\right) + \sqrt{\rho} \ln\left(\lambda_k^{(2)}\right), \text{ for } k = 0, 1, \ldots, 2I,$$

in which the random variables $\lambda_j^{(1)}$ and $\lambda_k^{(2)}$ are independent for all j, k,

$$\lambda_k^{(1)} \overset{\text{i.i.d.}}{\sim} log\mathcal{N}(\mu_\lambda, \sigma_\lambda^2), \ \lambda_k^{(2)} := \exp(\alpha S_{k-\Delta})$$

and the scaling factor α is chosen so that the coefficient of variation of $\lambda_k^{(1)}$ is equal to that of $\lambda_k^{(2)}$ for $k \ge I + \Delta + 1$.

Remark 3.6 If we choose $\rho = 0$ then Model 3.5 reduces to Model 3.1. Choosing $\rho = 1$ means that we believe the calendar period development factors $(\lambda_k)_k$ to be fully explained by the economic

indicators $(S_k)_{k \geq -\Delta}$. Thus, we can think of ρ as the credibility weight we give to Model 3.1 (as represented by $\lambda_k^{(1)}$) compared to a fully economic model (as represented by $\lambda_k^{(2)}$). For this reason, we refer to ρ as the *credibility weight*.

Remark 3.7 The mean of the error term ε_k in assumption (e') is chosen so that $\lambda_k^{(2)}$ is a martingale, for $k \geq I + \Delta + 1$.

Remark 3.8 If we did not take the square root of $1 - \rho$ and ρ in the equation of (f'), then the variance of $\ln(\lambda_k)$ is a strictly convex combination of the variance of $\ln(\lambda_k^{(1)})$ and $\ln(\lambda_k^{(2)})$. If, for example, the variance of $\lambda_k^{(1)}$ is much less than that of $\lambda_k^{(2)}$ then this could result in a favouring of models which have more weight given to $\lambda_k^{(1)}$. We avoid this possibility by taking the square root of $1 - \rho$ and ρ so that the variance of $\ln(\lambda_k)$ is a linear combination of the variance of $\ln(\lambda_k^{(1)})$ and $\ln(\lambda_k^{(2)})$.

If we use a model which includes economic indicator data, then the available information consists not only of the claims table data, but also of the observed economic indicators. We represent this information as

$$\mathcal{D}_I^\star := \{\mathcal{D}_I, (S_k)_{k \leq I}\}. \tag{3.5}$$

When we do Bayesian inference on Model 3.5, we use the information $\mathcal{D}_I^\star$. For example, the posterior density of Θ is $g(\Theta | \mathcal{D}_I^\star)$ instead of $g(\Theta | \mathcal{D}_I)$.

Applying Model 3.5 to the data

We analyzed the quarterly claims data corresponding to the data summarized by Table 1. This means that $I = 47$, corresponding to 48 quarter-year's worth of data from 1997 to 2008. We applied Model 3.5 with the parameters of the prior distributions as in Remark 3.2 and using two economic indicators: credit spreads and the unemployment rate. The economic indicator data is plotted in Figure 1. Although we also examined a consumer confidence index, the analysis showed that it was not useful for claims prediction and we do not show the results here.

Running Model 3.5 and assessing the output

To compute the posterior distributions, we used Markov chain Monte Carlo (MCMC) simulation methods. The MCMC methodology provides us with a simulated Markov chain $\Theta^{[1]}, \Theta^{[2]}, \Theta^{[3]}, \ldots$ with

$$\Theta^{[n]} = \left(\pi_0^{[n]}, \ldots, \pi_I^{[n]}, \gamma_0^{[n]}, \ldots, \gamma_I^{[n]}, \lambda_0^{[n]}, \ldots, \lambda_{2I}^{[n]} \right), \tag{3.6}$$

which is an empirical approximation of the posterior distribution $g(\Theta | \mathcal{D}_I^\star)$. The computation was implemented in *WinBUGS*, which is a software program specially designed for such a purpose, to produce 10 000 simulations from the posterior densities of each of the model's parameters. Scollnik (2001) gives an overview of MCMC techniques and how they can be implemented in *WinBUGS* in an actuarial context.

A selection of autocorrelation plots and traceplots for the parameters are shown in Figures 2 and 3, respectively. We used a thinning interval of 50 to reduce the autocorrelations which were observed without any thinning. Boxplots of the posterior parameter distributions are shown in Figure 4. The diagnostic plots in Figure 3 show that convergence has been obtained.

388
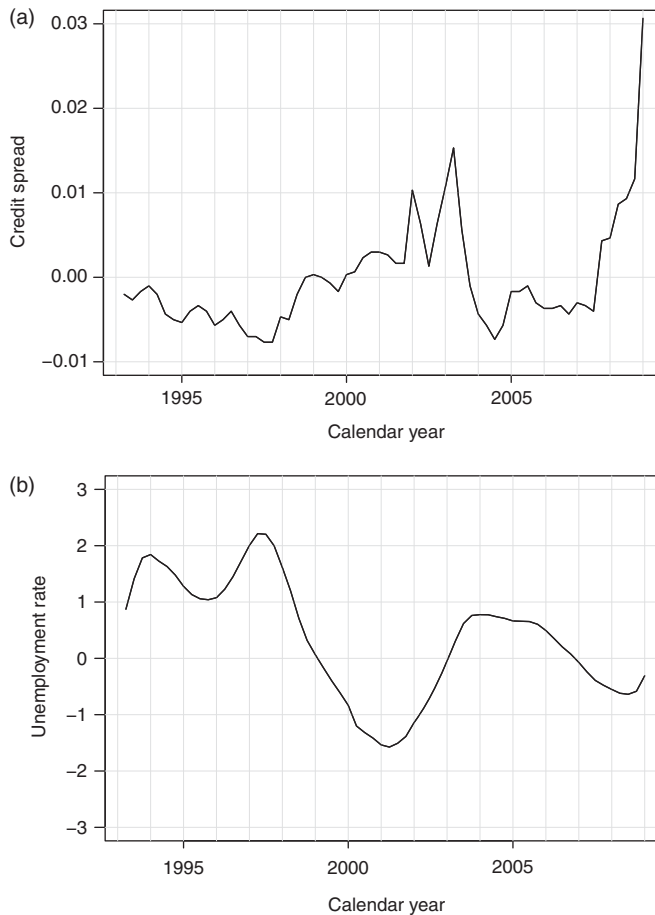
Figure 1. The economic indicator data. The first quarter of 1997 corresponds to calendar period 0. Each series has been normalized by subtracting the average value, so that the normalized series has empirical mean zero. (a) Credit spread data between government and corporate bonds (b) Unemployment rate data.
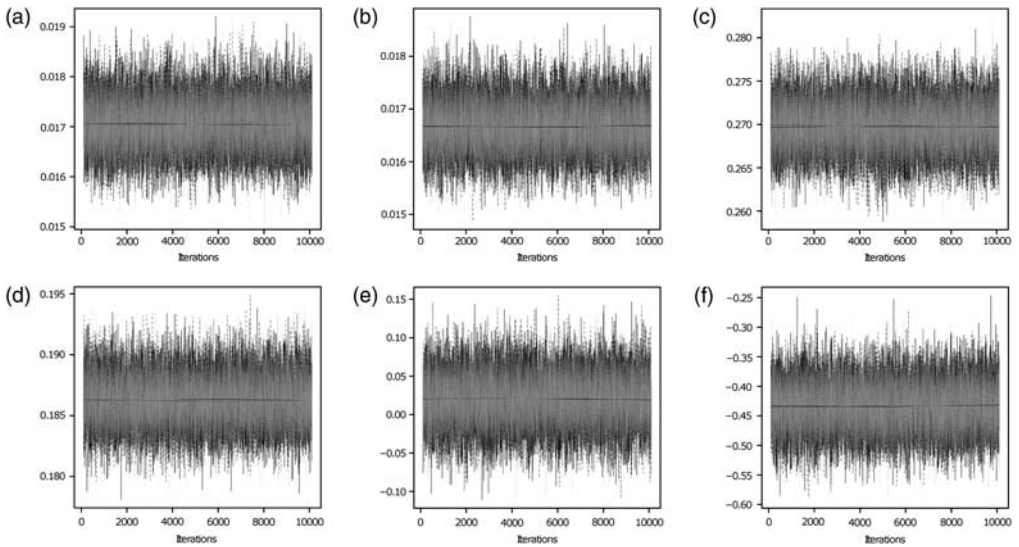
Selecting a time lag and economic indicator for Model 3.5

To compare Model 3.5 for different choices of the scalar ρ, time lag Δ and economic indicator series, we used a model selection criterion called the Deviance Information Criterion (DIC). Introduced in Spiegelhalter *et al.* (2002), the DIC is a way of comparing Bayesian models by measuring the trade off between the fit of the model to the data and the complexity of the model. Using the DIC as a model selection criterion suggests that we should choose the model with the smallest DIC. However, as it is a relatively *ad hoc* measure (for criticisms of DIC, see, for example, the discussion in Spiegelhalter *et al.* 2002), we do not apply this criterion rigorously. Instead, we use it as an approximate guide to the selection of a model. Note that *WinBUGS* can automatically calculate the DIC.

Choosing credit spreads as the economic indicator, we plot the DIC against the time lag Δ in Figure 5(a). Each line corresponds to a fixed choice of $\rho \in \{0, 0.1, 0.2, \ldots, 0.9\}$. In particular, the horizontal line corresponding to $\rho = 0$, which corresponds to having no economic indicators in
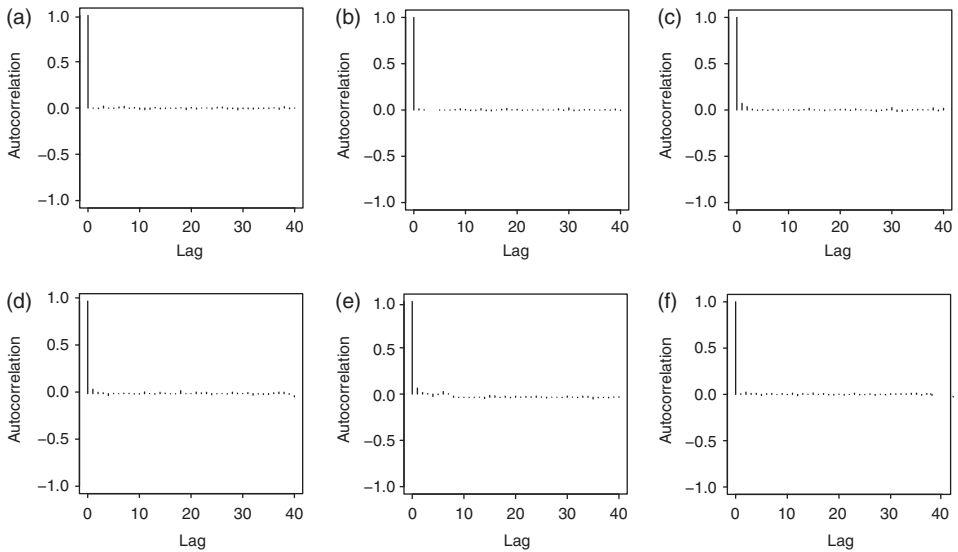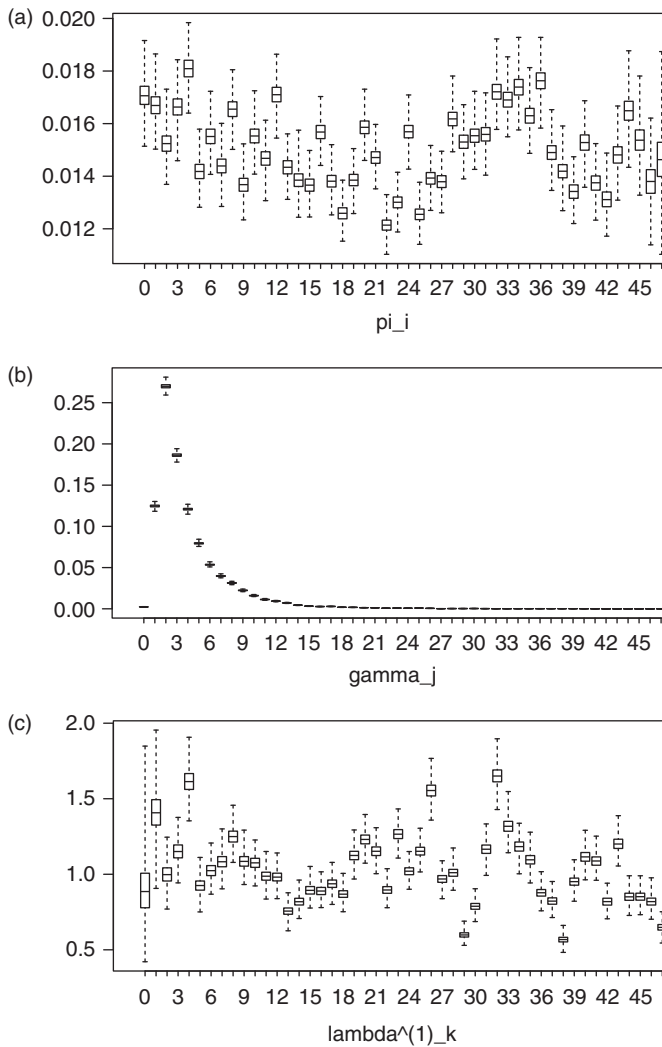
Figure 2. Autocorrelation plots for a selection of the parameters of Model 3.5, using credit spreads as the economic indicator, with a time lag $\Delta = 5$ quarter years and a credibility weight $\rho = 0.5$. (a) π_1 (b) π_2 (c) γ_3 (d) γ_4 (e) $\log(\lambda_{25}^{(1)})$ (f) $\log(\lambda_{48}^{(1)})$.


Figure 3. Traceplots for a selection of the parameters of Model 3.5, using credit spreads as the economic indicator, with a time lag $\Delta = 5$ quarter years and a credibility weight $\rho = 0.5$. (a) π_1 (b) π_2 (c) γ_3 (d) γ_4 (e) $\log(\lambda_{25}^{(1)})$ (f) $\log(\lambda_{48}^{(1)})$.

Model 3.5, gives the DIC value for Model 3.1. The DIC values for $\rho = 1$ are not plotted since they are much higher. For each fixed credibility weight $\rho \neq 0$, the lowest DIC is attained when the time lag is $\Delta = 5$, corresponding to a time lag of 5 quarter years. This suggests that the optimal time lag for the data analyzed is $\Delta = 5$, which is consistent with the optimal time lag obtained by König *et al.* (2011).

Figure 4. Boxplots showing the posterior distribution of the parameters of Model 3.5 using credit spreads as the economic indicator, with a time lag $\Delta = 5$ quarter years and a credibility weight $\rho = 0.5$. (a) Posterior distribution of the parameters $\pi_0, \pi_1, \ldots, \pi_{47}$ (b) Posterior distribution of the parameters $\gamma_0, \gamma_1, \ldots, \gamma_{47}$ (c) Posterior distribution of the parameters $\lambda_0^{(1)}, \lambda_1^{(1)}, \ldots, \lambda_{47}^{(1)}$.

Fixing the time lag $\Delta = 5$, we see from Figure 5(a) that the lowest DIC value is attained at a credibility weight of $\rho = 0.6$. However, for $\rho \in \{0.1, 0.2, \ldots, 0.8\}$, the differences in DIC are not very substantial, being less than 5 units in magnitude. This means that we cannot state with statistical conviction that the model with $\rho = 0.6$ is the "best", based on the lowest DIC criterion.

Figure 5(b) shows the results when we choose the unemployment rate as the economic indicator. For each fixed choice of the credibility weight $\rho \in \{0, 0.1, 0.2, \ldots, 0.7\}$, the lowest DIC is obtained when the time lag is $\Delta = 0$, corresponding to no time lag. As the credibility weight ρ is increased, the DIC increases. Again, we do not plot the DIC values for $\rho \in \{0.8, 0.9, 1\}$ since they are much higher.

391
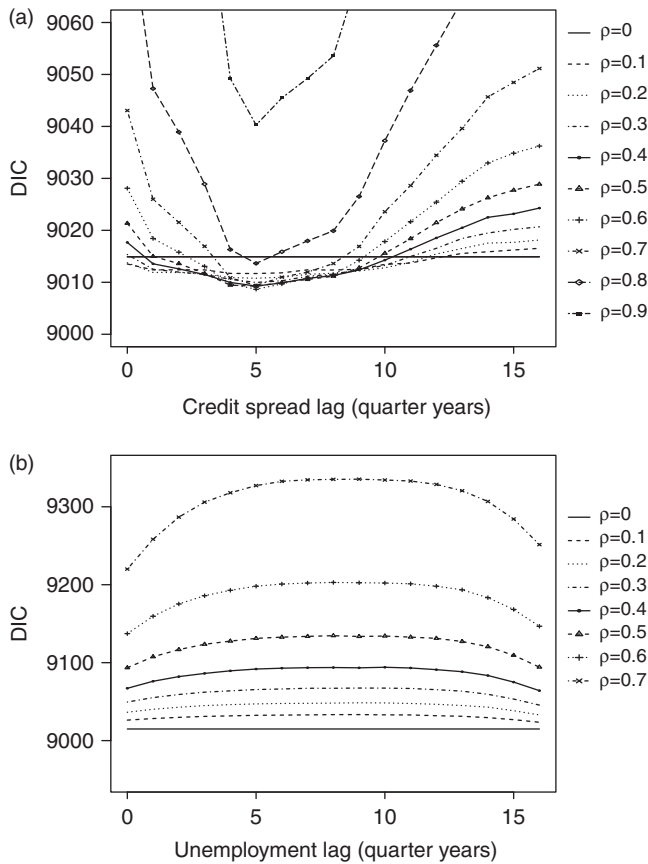
Figure 5. DIC against lags for various fixed values of the credibility weight ρ in Model 3.5. Note the difference in scales. (a) DIC against credit spread lag (b) DIC against unemployment lag.

In summary, based on the DIC, using credit spreads as the economic indicator suggests an optimal time lag $\Delta = 5$ and using the unemployment rate as the economic indicator suggests an optimal time lag $\Delta = 0$. In order to improve claims prediction, we prefer an economic indicator which maximizes the time lag and, on this criterion, we prefer to use credit spreads as an economic indicator. Indeed, for the claims data summarized by Table 1, the unemployment rate is not particularly useful as an economic indicator since it has an optimal time lag of zero.

3.4.2 A second model incorporating an economic indicator

Using credit spreads as an economic indicator in Model 3.5, the differences in the DIC at time lag $\Delta = 5$ were not large enough to enable us to choose a particular value of the credibility weight ρ. For this reason, we considered a model which is identical to Model 3.5 except that the credibility weight is modelled as a parameter with a prior distribution, rather than as a constant. The motivation is to find the time lag which allows us to give the most weight to the economic factor $\lambda_k^{(2)}$.

Model 3.9   Assume Model 3.5 but replace the parameter vector Θ by

$$\tilde{\Theta} := (\rho, \Theta)$$

392

and replace assumption (f′) with

(f″)

$$\ln(\lambda_k) := \sqrt{1-\rho}\,\ln\left(\lambda_k^{(1)}\right) + \sqrt{\rho}\,\ln\left(\lambda_k^{(2)}\right), \ \ \text{for } k = 0, 1, \ldots, 2I,$$

in which the credibility weight parameter ρ is independent of all the other parameters and has prior distribution

$$\rho \sim \text{Unif}[0, 1],$$

the random variables $\lambda_j^{(1)}$ and $\lambda_k^{(2)}$ are independent for all j, k,

$$\lambda_k^{(1)\,\text{i.i.d.}}\sim \log\mathcal{N}(\mu_\lambda, \sigma_\lambda^2), \ \ \lambda_k^{(2)} := \exp(\alpha S_{k-\Delta})$$

and the scaling factor α is chosen so that the coefficient of variation of $\lambda_k^{(1)}$ is equal to that of $\lambda_k^{(2)}$ for $k \geq I + \Delta + 1$.

Remark 3.10 Reflecting the inconclusive results about the optimal value of the credibility weight in Model 3.5, we assume that the prior distribution of the credibility weight parameter is uniformly distributed between 0 and 1. As in Model 3.5, the mean of the error term ε_k in assumption (e′) is chosen so that $\lambda_k^{(2)}$ is a martingale, for $k \geq I + \Delta + 1$.

Applying Model 3.9 to the data

We analyzed the same quarterly claims data, applying Model 3.9 with the values of the parameters of the prior distributions as detailed in Remark 3.2 and using credit spreads and the unemployment rate as economic indicators. As before, we use a MCMC method to obtain simulations

$$\widetilde{\Theta}^{[n]} = \left(\rho^{[n]}, \pi_0^{[n]}, \ldots, \pi_I^{[n]}, \gamma_0^{[n]}, \ldots, \gamma_I^{[n]}, \lambda_0^{[n]}, \ldots, \lambda_{2I}^{[n]}\right)$$

from a Markov chain which empirically approximates the posterior distribution $g(\widetilde{\Theta}|\mathcal{D}_I^\star)$.

Running Model 3.9 and assessing the output

A selection of autocorrelation plots and traceplots for the parameters are shown in Figures 6 and 7, respectively. The plots were obtained after using a thinning interval of 50 to reduce the autocorrelations. We observe more autocorrelation and thus a slower rate of convergence than for Model 3.5. Boxplots of the posterior parameter distributions of $(\lambda_k^{(1)})_{k \leq I}$ and $(\lambda_k)_{k \leq I}$ are shown in Figure 8. The effect of the credit spreads are seen clearly in this figure by comparing the mean of $\lambda_k^{(1)}$ to the mean of λ_k, for each value of k. By examining Figure 8(a) and Figure 1(a) together, we see that the parameter means in Figure 8(b) decrease when the (normalized) lagged credit spreads are negative and they increase when the lagged credit spreads are positive. The diagnostic plots in Figure 7 show that convergence has been obtained.

Selecting a time lag and economic indicator for Model 3.9

Since the information is given by $\mathcal{D}_I^\star$ (recall (3.5)), the posterior mean of the credibility weight parameter ρ is

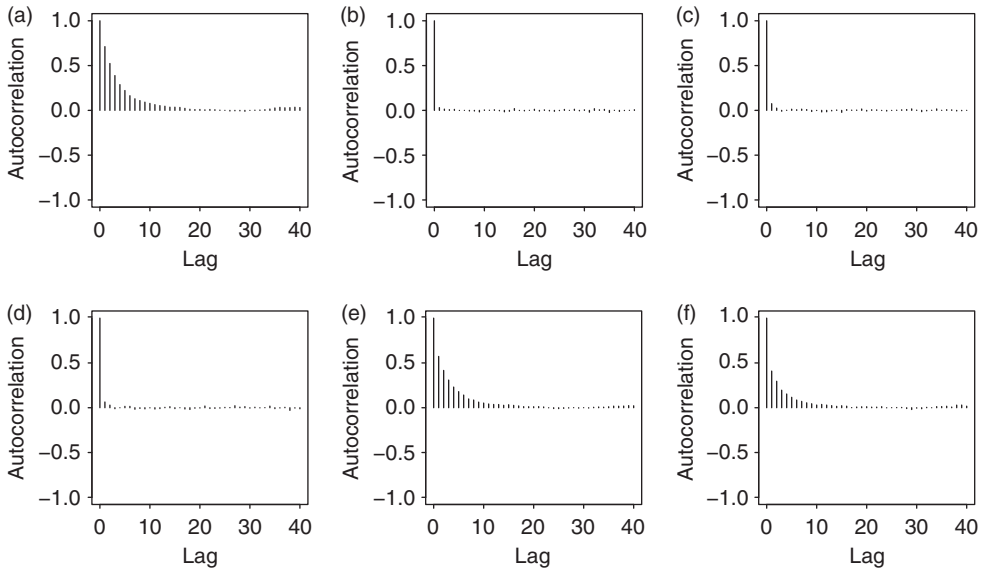$$\bar{\rho} := \mathbb{E}\left(\rho|\mathcal{D}_I^\star\right).$$

Figure 6. Autocorrelation plots for a selection of the parameters of Model 3.9, using credit spreads as the economic indicator, with a time lag $\Delta = 5$ quarter years. (a) ρ (b) π_1 (c) γ_3 (d) γ_4 (e) $\log(\lambda_{25}^{(1)})$ (f) $\log(\lambda_{48}^{(1)})$.
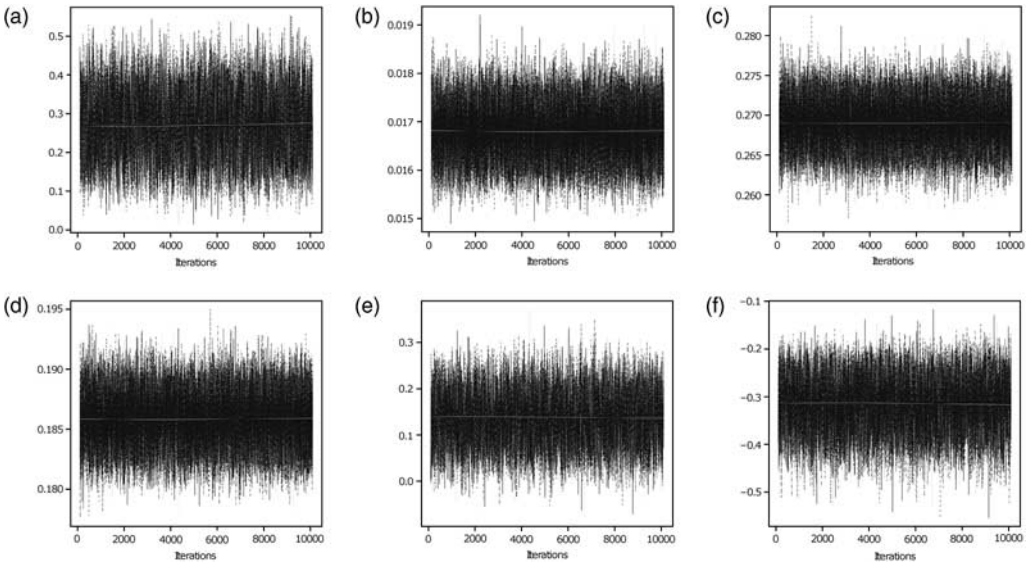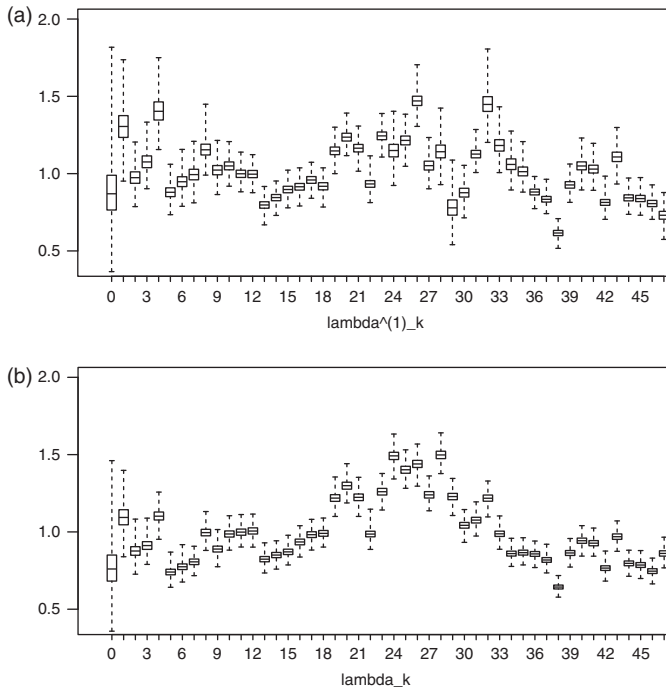

Figure 7. Traceplots for a selection of the parameters of Model 3.9, using credit spreads as the economic indicator, with a time lag $\Delta = 5$ quarter years. (a) ρ (b) π_1 (c) γ_3 (d) γ_4 (e) $\log(\lambda_{25}^{(1)})$ (f) $\log(\lambda_{48}^{(1)})$.

In Figure 9(a) the posterior mean $\bar{\rho}$ of the credibility weight parameter is plotted against the time lag Δ when we use credit spreads as the economic indicator; this is the solid line. The dashed and dotted lines show a 50% and 95% credible interval about the posterior mean, respectively. The plot shows

394

Figure 8. Boxplots showing the posterior distribution of the parameters $(\lambda_k^{(1)})_{k \leq I}$ and $(\lambda_k)_{k \leq I}$ of Model 3.9 using credit spreads as the economic indicator, with a time lag $\Delta = 5$ quarter years. (a) Posterior distribution of the parameters $\lambda_0^{(1)}, \lambda_1^{(1)}, \ldots, \lambda_{47}^{(1)}$ (b) Posterior distribution of the parameters $\lambda_0, \lambda_1, \ldots, \lambda_{47}$.

that the highest posterior mean $\bar{\rho} = 0.270$ is attained at time lag $\Delta = 5$. The time lag $\Delta = 5$ is consistent with the results when Model 3.5 is applied to the same data.

Figure 9(b) shows the results when we use the unemployment rate as an economic indicator. For this latter plot, the highest posterior mean $\bar{\rho} = 0.007$ is attained at time lag $\Delta = 0$. This means that not only is the unemployment rate not useful for prediction, but also that the impact of the unemployment rate on the calendar year development factors is very small.

In summary, the analysis of the claims data using Model 3.9 suggests choosing credit spreads as the economic indicator with time lag $\Delta = 5$. This is consistent with the results in Subsection 3.4.1 and König *et al.* (2011). The unemployment rate is relatively unhelpful, both in its predictive ability and its impact on the claim numbers. This also means that it does not appear worthwhile to extend the model to include both economic indicators.

4 Improving disability prediction

We have considered two Bayesian models which incorporate an economic indicator. Our analysis of these models with the claims data summarized by Table 1 suggests that Models 3.5 and 3.9, with the credit spreads as the economic indicator and a time lag $\Delta = 5$, would improve claims prediction. Here we compare the claims prediction of the latter model against that of Model 3.1, which does not incorporate any economic indicator.

395
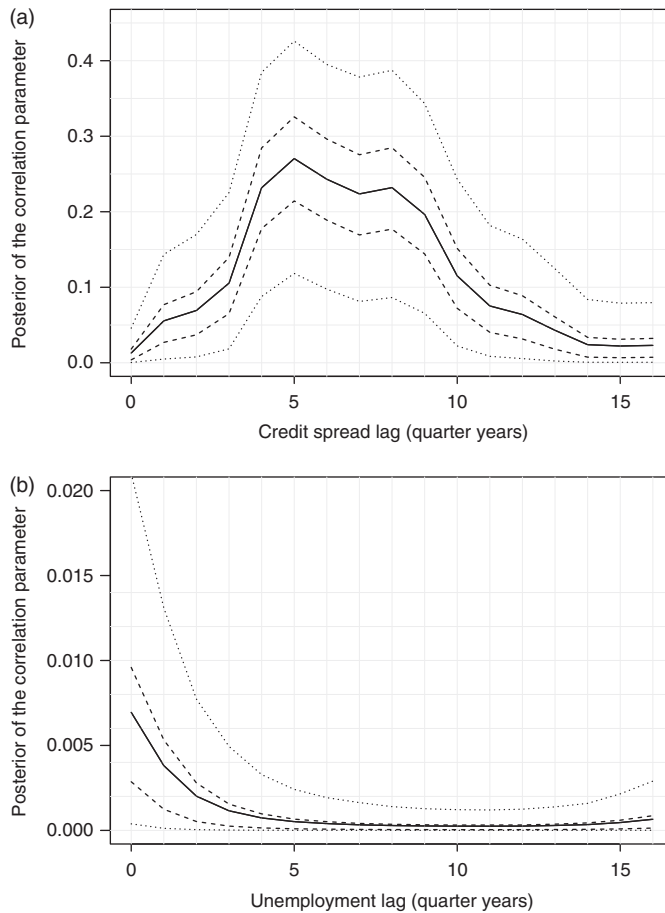
Figure 9. Posterior mean of the credibility weight parameter against time lag, with the dashed lines showing a 50% and 95% credible interval about the mean. (a) Posterior mean of the credibility weight parameters against credit spread lag (b) Posterior mean of the credibility weight parameters against unemployment lag.

Posterior distribution from MCMC

For Models 3.1 and 3.9 applied to the quarterly claims data, we used MCMC techniques as outlined before to obtain 10 000 samples from each of the posterior distributions of the parameters in Θ and $\tilde{\Theta} = (\rho, \Theta)$, respectively. From these samples, we can calculate the empirical densities of the parameters. For example, the posterior density $g(\rho|\mathcal{D}_I^*)$ of the credibility weight parameter ρ is plotted in Figure 10. The plot is approximately symmetrical about the mean $\bar{\rho} = 0.270$, with the density tending to zero at values of ρ near zero and 0.5.

The predictive distribution

A considerable advantage of using the Bayesian approach to claims prediction is that we obtain the predictive distribution of every entry $N_{i,j}$ in the lower triangle. With the predictive distribution at hand, one can do various analyses of the total number of future claims. For example, one can
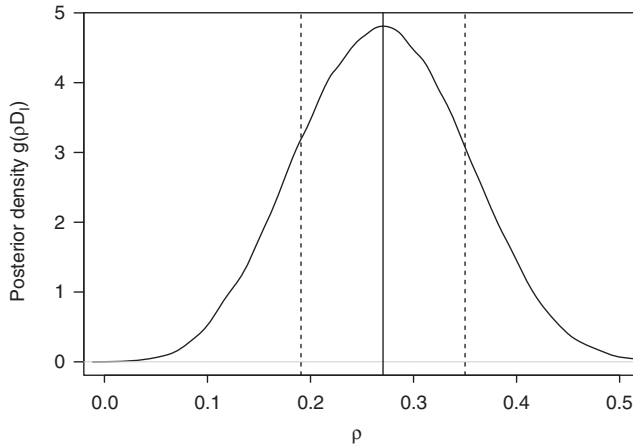
Figure 10. The posterior density plot for the credibility weight parameter ρ given $\mathcal{D}_I^\star$ using credit spreads as the economic indicator with a time lag $\Delta = 5$ in Model 3.9. The solid vertical line shows the posterior mean and the vertical dashed lines show one posterior standard deviation about the posterior mean.

calculate any risk measure and thus perform a tail event analysis, which is very important for any solvency consideration.

Using the relationship

$$f\left(N_{i,j} \mid \mathcal{D}_I^\star\right) = \int f(N_{i,j}|\tilde{\theta})\, g(\tilde{\theta} \mid \mathcal{D}_I^\star)\, d\tilde{\theta}, \;\; \text{for each } i+j > I,$$

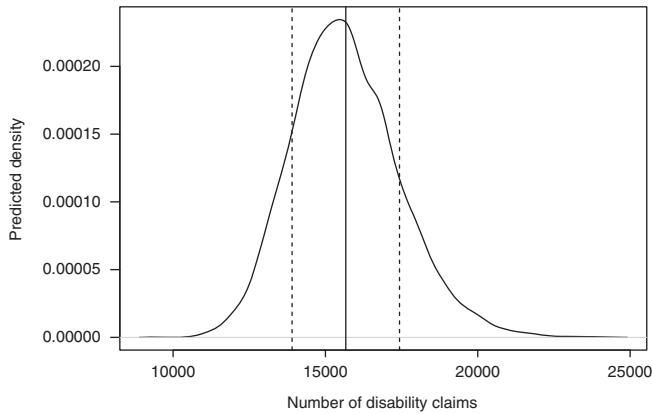we can sample from the predictive densities $f(N_{i,j}|\mathcal{D}_I^\star)$ as follows. Let

$$\hat{\theta} = (\hat{\rho}, \hat{\pi}_0, \ldots, \hat{\pi}_I, \hat{\gamma}_0, \ldots, \hat{\gamma}_I, \hat{\lambda}_0, \ldots, \hat{\lambda}_{2I})$$

denote a sample from the empirical posterior joint density $g(\tilde{\Theta}|\mathcal{D}_I^\star)$ of the parameters, obtained by the MCMC method. We use this sample to simulate from $N_{i,j}|\hat{\theta}$, which by assumption is Poisson distributed with mean $s_i \hat{\pi}_i \hat{\gamma}_j \hat{\lambda}_{i+j}$. The result is a sample from $N_{i,j}|\mathcal{D}_I^\star$. We do this for each of the 10 000 samples from the posterior joint density $g(\tilde{\Theta}|\mathcal{D}_I^\star)$ to obtain 10 000 samples from the predictive density $f(N_{i,j}|\mathcal{D}_I^\star)$. Repeating this procedure simultaneously for one sample $\hat{\theta}$ and all the cells in the lower triangle, we use the resulting samples to calculate empirically the predictive density

$$f\left(\sum_{i=1}^{I} \sum_{j=I-i+1}^{I} N_{i,j}|\mathcal{D}_I^\star\right)$$

of the total number of future claims (that is, the sum of the cells in the lower triangle) for Model 3.9. This empirical predictive density of the lower triangle is plotted in Figure 11. The vertical solid line shows the empirical posterior mean

$$\mathbb{E}\left(\sum_{i=1}^{I} \sum_{j=I-i+1}^{I} N_{i,j}|\mathcal{D}_I^\star\right) = \sum_{i=1}^{I} \sum_{j=I-i+1}^{I} s_i \mathbb{E}\left(\pi_i \gamma_j \lambda_{i+j}|\mathcal{D}_I^\star\right) \tag{4.1}$$

397

Figure 11. Empirical density function of the predicted total number of claims using credit spreads as the economic indicator with a time lag $\Delta = 5$ in Model 3.9. The solid vertical line shows the posterior mean and the vertical dashed lines show one posterior standard deviation about the posterior mean.

and the vertical dashed lines show one posterior standard deviation about the posterior mean. Notice from Figure 11 that we have positive skewness in the predictive density of the total number of future claims.

Remark 4.1 In (4.1) we can no longer decouple λ_{i+j} as in (3.2) since λ_{i+j} now depends on the observed credit spread data through the random walk which projects the credit spread data beyond time period I (assumption (e')).

The main result

As the disability frequency is an important quantity to measure the disability risk of a portfolio, we examine how the prediction of the disability frequency varies between the two models. For Model 3.1, it is given by (3.3) with $\mathbb{E}(\lambda_{i+j}) = 1$. For Model 3.9, the posterior mean predicted disability frequency is given by

$$\hat{\bar{p}}_i = \frac{1}{s_i} \sum_{j=0}^{I-i} N_{i,j} + \sum_{j=I-i+1}^{I} \mathbb{E}\left(\pi_i \gamma_j \lambda_{i+j} \mid \mathcal{D}_I^\star\right), \quad \text{for } i > 1.$$

For Model 3.9, the posterior mean predicted disability frequency is shown in Figure 12 as a solid line, with the dashed lines indicating a 95% credible interval. The posterior mean predicted disability frequency which results from adopting Model 3.1 is also shown; it is the dashed-dotted line, also with a 95% credible interval indicated by dotted lines. The effect of incorporating the credit spread data is clear; the posterior mean predicted disability frequency is much higher for Model 3.9 than for Model 3.1, especially at the later incurred periods where most of the claims numbers are predicted. For example, for claims which were incurred in period 47 (corresponding to the last quarter of 2008), Model 3.9 gives the posterior mean predicted disability frequency $\bar{p}_{47} = 0.0228$, whereas Model 3.1 gives $\bar{p}_{47} = 0.0145$. The significantly higher predicted disability frequency for Model 3.9 reflects the increase in credit spreads observed from calendar period 43 onwards (see Figure 1(a)). The increasing credit spreads result in the corresponding calendar period

398
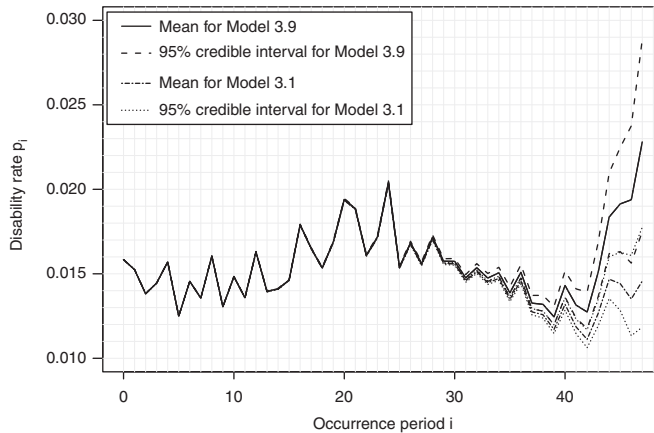
Figure 12. The mean predicted disability rate for Model 3.9, with credit spreads as the economic indicator and time lag $\Delta = 5$ quarter years, and for Model 3.1. The dashed lines show a 95% credible interval about each mean.

development factors increasing, which means that the predicted number of claims increase too. Indeed, credit spreads were at unusually high levels around calendar periods 47 and 48 (corresponding to the last six months of 2008), due to a breakdown in trust between banks which led to a severe liquidity crisis. If we believe that the crisis was a temporary phenomenon which was highly unusual and had a less severe impact on insured lives than on banks, then we may decide to adjust the credit spread data downwards. However, this is a matter of professional judgment and we have not done any such adjustments in our analysis.

Examining the credible intervals in Figure 12, it is clear that they are wider for Model 3.9 than for Model 3.1. This is due to the greater variation in the calendar year development factors of Model 3.9; they are a function not only of the calendar year development factors of Model 3.1 (that is, $\lambda_k^{(1)}$), but also of the credibility weight parameter ρ and the economic indicator data.

In summary, the incorporation of credit spread data into the model has resulted in a marked difference in the estimation of the disability risk. Due to the increasing credit spreads which were observed in 2008, the disability risk for claims which were incurred in calendar year 2008 is significantly higher under Model 3.9 than under Model 3.1, which does not incorporate economic data. The results suggest that ignoring economic indicators when predicting claims can result in a significant mis-estimation of the disability risk because these economic indicators correlate with disability rates.

5 Summary

In this paper, we propose a Bayesian model which incorporates an economic indicator. The motivation is to include economic effects which affect the development of the number of disability claims and hence improve the claims prediction. We examined in detail two possible economic indicators: credit spread and the unemployment rate. For the disability claims data we analyzed, there was evidence that credit spreads are useful indicators for claims development, but there was no compelling evidence for incorporating the unemployment rate.

To illustrate the impact of incorporating economic indicators into the Bayesian model, we focused on the disability frequency, which is a measure of the disability risk in a portfolio. Due to the current financial crisis, the mean predicted disability frequency increased sharply when using credit spreads as an economic indicator, as opposed to not using any economic indicator. The results demonstrate how the incorporation of an economic indicator can significantly alter the prediction of the disability risk.

While we found credit spreads to be a reasonable economic indicator, clearly this depended on the data we analyzed. For other datasets, not only may other economic indicators be relevant, but multiple economic indicators could be incorporated into the Bayesian model. However, many economists use credit spreads as a time-lagged indicator of the state of an economy.

It would be interesting to apply a Bayesian model with economic indicators to the amounts of the disability claims, and not only to the number of claims, as well as the effect of economic conditions on the duration of income disability insurance. Unfortunately, we did not have the data to perform these analyses.

Acknowledgments

References

de Alba, E. (2002). Bayesian estimation of outstanding claims reserves. *North American Actuarial Journal*, **6**(4), 1–20.

de Alba, E. (2006). Claims reserving when there are negative values in the runoff triangle: Bayesian analysis using the three-parameter log-normal distribution. *North American Actuarial Journal*, **10**(3), 45–59.

de Alba, E. & Nieto-Barajas, L. (2002). Claims reserving: A correlated Bayesian model. *North American Actuarial Journal*, **6**(4), 1–20.

Denuit, M., Marechal, X., Pitrebois, S. & Walhin, J.-F. (2007). *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems*. John Wiley & Sons, England.

England, P. & Verrall, R. (2006). Predictive distributions of outstanding liabilities in general insurance. *Annals of Actuarial Science*, **1**(2), 221–270.

König, B., Weber, F. & Wüthrich, M. (2011). Prediction of disability frequencies in life insurance. *Zavarovalniski horizonti*, **7**(3), 5–23.

Ntzoufras, I. & Dellaportas, P. (2002). Bayesian modeling of outstanding liabilities incorporating claim count uncertainty. *North American Actuarial Journal*, **6**(1), 113–128.

Peters, G., Shevchenko, P. & Wüthrich, M. (2009). Model uncertainty in claims reserving within Tweedie's compound Poisson model. *ASTIN Bulletin*, **39**(1), 1–34.

Schriek, K. & Lewis, P. 2010. The link between disability experience and economic conditions in South Africa. Presented at the International Congress of Actuaries 2010 Conference. Download from http://www.ica2010.com/docs/02_final_paper_Schriek,_Lewis.pdf

Scollnik, D. (2001). Actuarial modeling with MCMC and BUGS. *North American Actuarial Journal*, **5**(2), 96–124.

Spiegelhalter, D., Best, N., Carlin, B. & van der Linde, A. (2002). Bayesian measures of complexity and fit. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, **64**(4), 583–639.

Verrall, R. (2004). A Bayesian generalized linear model for the Bornhuetter-Ferguson method of claims reserving. *North American Actuarial Journal*, **8**(3), 67–89.