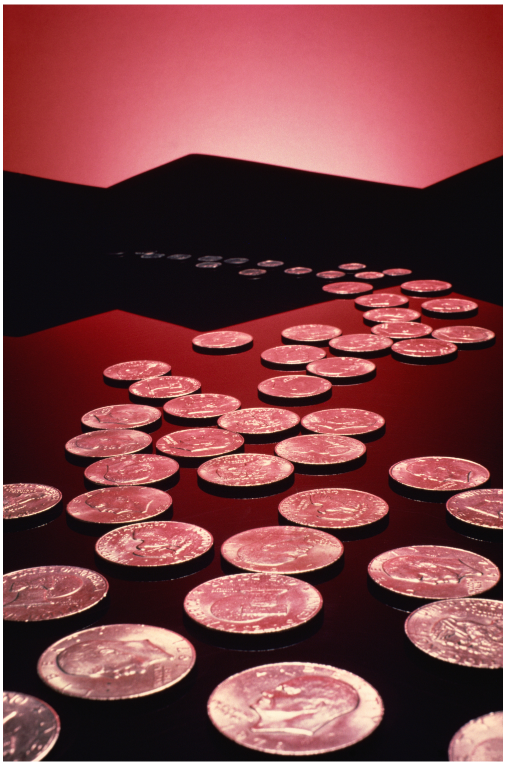
Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Note

Bayesian Modelling of Credit Risk Using Integrated Nested Laplace Approximations



Note no SAMBA/47/08

Author Mathilde Wilhelmsen
Xeni K. Dimakos
Tore Husebø
Marit Fiskaaen

Date February 4, 2009

Norwegian Computing Center

Norsk Regnesentral (Norwegian Computing Center, NR) is a private, independent, non-profit foundation established in 1952. NR carries out contract research and development projects in the areas of information and communication technology and applied statistical modeling. The clients are a broad range of industrial, commercial and public service organizations in the national as well as the international market. Our scientific and technical capabilities are further developed in co-operation with The Research Council of Norway and key customers. The results of our projects may take the form of reports, software, prototypes, and short courses. A proof of the confidence and appreciation our clients have for us is given by the fact that most of our new contracts are signed with previous customers.

Norsk Regnesentral
Norwegian Computing Center
Postboks 114, Blindern
NO-0314 Oslo, Norway

Besøksadresse
Office address
Gaustadalléen 23
NO-0373 Oslo, Norway

Telefon · telephone
(+47) 22 85 25 00
Telefaks · telefax
(+47) 22 69 76 60

Internett · internet
www.nr.no
E-post · e-mail
nr@nr.no

Title	**Bayesian Modelling of Credit Risk Using Integrated Nested Laplace Approximations**

Author	**Mathilde Wilhelmsen** <Mathilde.Wilhelmsen@nr.no> **Xeni K. Dimakos** <Xeni.Dimakos@nr.no> **Tore Husebø** <tore.husebo@sparebank1.no> **Marit Fiskaaen** <marit.fiskaaen@sparebank1.no>

Date	February 4, 2009
Publication number	SAMBA/47/08

Abstract

We apply the method of Integrated Nested Laplace Approximation (INLA) to Bayesian modeling of credit risk. INLA was developed by Rue et al. (2009) for inference in the class of latent Gaussian models. This is a wide class of models that includes spatial and spatio-temporal models, log-Gaussian Cox-processes and geostatistical models, but also Generalized linear models (GLM) which is the benchmark for credit default modeling. We demonstrate how INLA provides fast and accurate inference in the Bayesian credit risk model and compare its performance and results to those of the commonly applied Markov Chain Monte Carlo (MCMC) approach.

Keywords	Bayesian logistic regression, Credit risk, Integrated Nested Laplace Approximations, Markov Chain Monte Carlo, MCMCpack
Target group	International journals
Availability	Open
Project	SIP-MASC, Finance
Project number	220363
Research field	Finance, insurance and commodity markets
Number of pages	25

1 Introduction

For many financial institutions and banks, credit risk is by far the most important risk type. Estimation of default probabilities is fundamental in both credit scoring and credit risk models. As the new Basel capital accord (Basel II, 2006) allows banks to use their own default probability estimations, also the calculation of regulatory capital is influenced by the strategy applied for estimating default probabilities.

The study of bankruptcy prediction goes back to Beaver (1966) and Altman (1968). Two classes of models dominate the literature. The market, or structural models, are based on the value of the firm as set by a market, and often approximated by stock prices. The KMV model (Crosbie and Bohn, 2002) which is widely used in industry is an example of a structural model. Accounting based models, on the other hand, use available financial indicators such as annual financial statements.

This paper focuses on the accounting based models, and in particular on the logistic regression model (McCullagh and Nelder, 1989). The literature includes applications of neural networks, linear discriminant models and general additive models for the same purpose, see for instance Atiya (2001) and Berg (2007). Studies that compare these models for the purpose of default prediction show, not surprisingly, that the behavior and suitability depend on the data at hand. For many practitioners the logistic regression model remains the benchmark model as it is easy to interpret and available on many software platforms.

Defaults are rare and the data sets of recorded defaults tend to be moderate, in particular for smaller banks. Moreover, additional information on the risks will often be available through expert risk assessments, credit ratings from other sources, regulators or publicly available data. Hence, it makes sense to combine the data available with the additional information that can be gathered, meaning that the Bayesian approach might be particularly useful for default probability estimation.

The Bayesian approach to credit risk modeling is not new. Löffler et al. (2005) propose an empirical Bayes approach for banks with small credit default data sets and suggest that prior information may be retrieved from academic literature or regulators. Other applications of Bayesian models to credit risk modeling is Mc Neil and Wedin (2007) and Ando (2006).

Inference in the Bayesian logistic regression model is done using MCMC (Gilks et al., 1996). With software such as OpenBugs (Spiegelhalter et al., 2007) and MCMCpack (Martin et al., 2008), MCMC algorithms are available to everyone in user-friendly environments. However, the fundamental challenge applying MCMC in practice remains: To determine convergence! The practitioner faces the task of choosing between several different measures of convergence, which might give contradictive answers to the question of convergence. Closely related is the issue of computational speed. Many applications require that the MCMC chain is run for hours or even days to obtain satisfactory convergence. Mira and Tenconi (2004) show how to speed up the convergence in a credit risk application. However, this requires implementation outside the scope of the above mentioned packages.

The main contribution of this paper is to introduce Integrated Nested Laplace Approximation (INLA) as an alternative to MCMC for Bayesian credit risk modeling. INLA was developed by Rue et al. (2009) as an efficient method for inference in complex models where the problems of convergence and computational time make MCMC unsuitable, or even infeasible. The examples and models considered in Rue et al. (2009) are more complex than the Bayesian credit risk model we consider. However, even in our relatively simple model, determination of convergence can be non-trivial.

The INLA approach for approximating the posterior marginals is computed in three steps. The first step approximates the posterior marginal distribution of the so-called hyperparameters, that is the parameters of the prior distribution. The second step computes an approximation of the posterior distribution of the model parameters, given the observed response variables and the hyperparameters. The third step combines the previous two steps using numerical integration to obtain the posterior distribution of the regression parameters.

Applying the Bayesian approach requires a method for articulating one's prior information. The Bayesian logistic regression model usually has a prior distribution on the regression coefficients. For regulators and risk managers it is inherently easier to express opinions in terms of default probabilities than regression coefficients. The second contribution of this paper is a simple simulation based method to convert beta distributed priors on default probabilities to priors on the regression coefficients.

We show that INLA is practically exact in the sense that one would have to run MCMC for a very long time to detect any indication of error in the approximate results. With our data, MCMC convergence is also fast. However, this may not always be the case.

The data set used in this paper is collected from four of the savings banks in the SpareBank 1-alliance in Norway. All four banks use the internal ratings-based approach (IRB) for calculating the capital requirements for credit risk. The data set consists of 7080 small and medium sized firms. For each firm a set of financial and non-financial variables are used to estimate the probability of default.

The remainder of this paper is organized as follows: Section 2 introduces the data set. Section 3 presents the Bayesian logistic regression model for credit risk and the latent Gaussian models for which INLA was developed. Section 4 gives an introduction to INLA and the software used in this paper. Section 5 presents our numerical results, and finally Section 6 concludes.

2 Data

In this paper we consider a Sparebank 1-alliance portfolio from 2006/2007 which consists of 7080 customers and their associated accounting variables. The portofolio consists of non-financial firms from many industries, including manufacturing, building and construction, fishery, trade etc. Firms in the industries shipping, shipyards and property development are not included in the portofolio considered here.

There are 126 recorded defaults in the data set, which corresponds to an average default probability of 1.8%. Table 1 shows the 13 explanatory variables the Sparebank 1-alliance considers to be the most important in describing defaults among customers. All variables are transformed into values between 0 and 10. Except for the variable age, 10 is also considered to be the best value a customer can get. In other words, high values are associated with behaviour that reduces the probability of default. The variables can be divided into 4 groups; income, consumption, behaviour and age. Figure 1 shows the histograms of all the explanatory variables in the data set.

3 Model

3.1 Logistic regression

The logistic regression model belongs to the class of Generalised Linear Models (GLM), McCullagh and Nelder (1989). The response variable in a logistic regression model is binomial and the expectation is related to the linear predictor through the logit function.

In our case we introduce an indicator of default. Let $Y_i = 1$ if customer i defaults,

Shortening	*Name*	*Values*	*Category*
PROFIT	Operating profit/Financial costs	[0,10]	Income
LOSS	Operating loss	[0,10]	Income
CPINV	Circulation pace on inventory	[0,10]	Income
DP	Degree of profit	[0,10]	Income
EQPER	Equity percent	[0,10]	Consumption
MPAY	Means of payment	[0,10]	Consumption
DLIQ	Degree of liquidity	[0,10]	Consumption
ODFAC	Overdraft facility	[0,10]	Consumption
REPHIST	History of company's reprimands	[0,10]	Behaviour
REPKF	Reprimands of payment for key figures	[0,10]	Behaviour
REPAC	Reprimands from accountants	[0,10]	Behaviour
AD	Submitted accounts delayed	[0,10]	Behaviour
AGE	Age	[0,10]	Age

Table 1. Overview of the explanatory variables.

and $Y_i = 0$ otherwise. Then,

$$Y_i \sim \text{Binomial}(p_i), i = 1, \ldots, N,$$

and the linear predictor

$$\eta_i = \beta_0 + \sum_{j=1}^{M} \beta_j x_{ij} + \epsilon_i, \tag{1}$$

is linked to the default probability through the logit function,

$$\text{logit}(p_i) = \log\{\frac{p_i}{1-p_i}\} = \eta_i.$$

Here, the explanatory variables x_{ij} are customer characteristics. It follows that the probability of default is given by

$$p_i = \frac{\exp(\beta_0 + \sum_{j=1}^{M} \beta_j x_{ij})}{1 + \exp(\beta_0 + \sum_{j=1}^{M} \beta_j x_{ij})}. \tag{2}$$

3.2 Bayesian formulation

Our Bayesian model is formulated by specifying prior distributions on the regression coefficients in (1),

$$\beta_j \sim \pi(\beta_j | \boldsymbol{\theta}_j), j = 0, \ldots, M.$$

The prior distribution $\pi(\cdot|\boldsymbol{\theta}_j)$ may be any proper probability density function and $\boldsymbol{\theta}_j$ may be a scalar or a parameter vector. The interpretation of the model parameters depends on the choice of the distribution, but may for instance include a

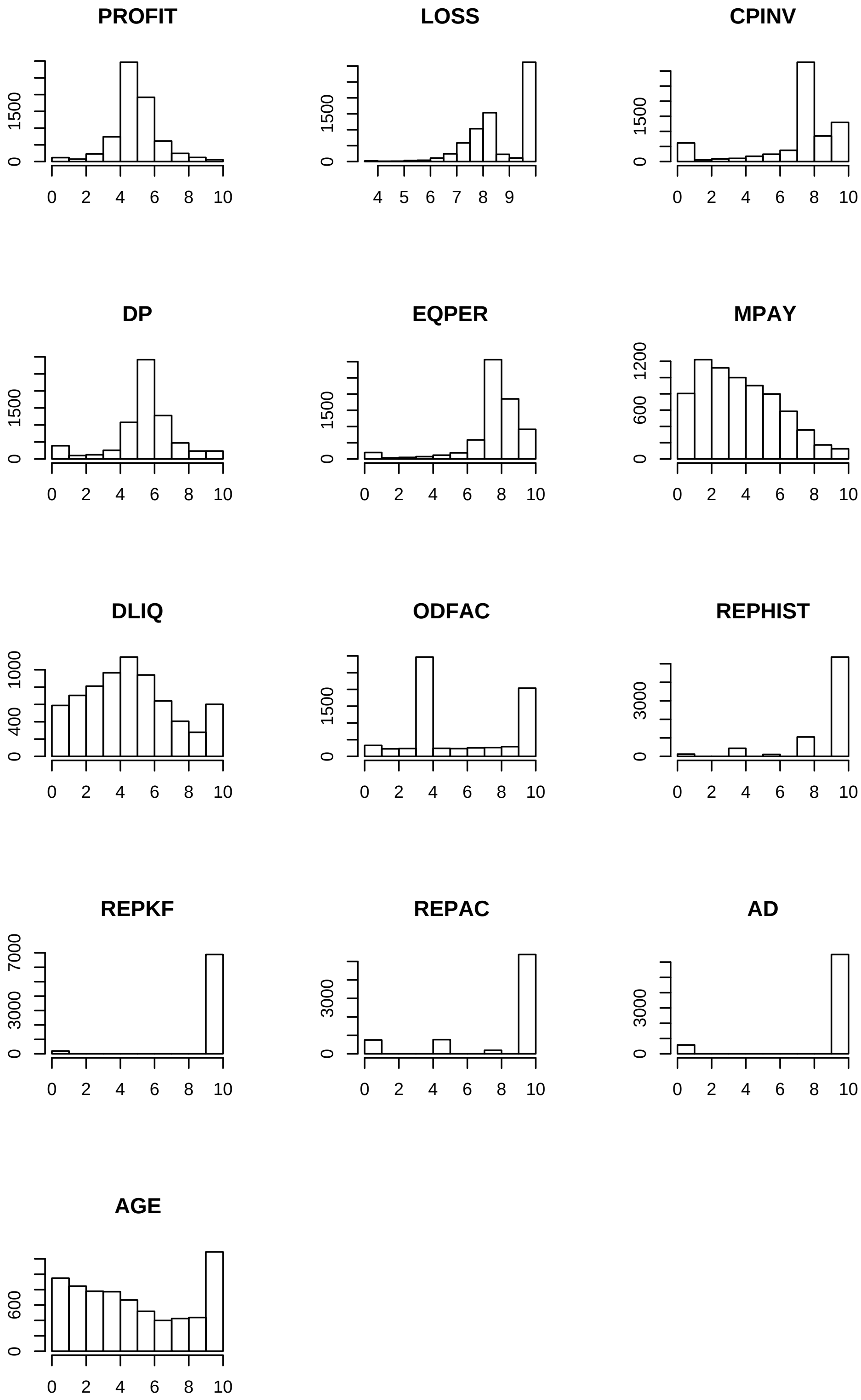


Figure 1. Histograms of the explanatory variables.

measure of the centre and spread of the prior. By specifying different values of $\boldsymbol{\theta}_j$, the regression coefficients may have very different priors even if the distribution function $\pi(\cdot)$ is the same.

The model parameters of the prior distribution in a Bayesian model are often called *hyperparameters*. The hyperparameters may also be assigned priors, in which case we obtain a hierarchical Bayesian model. The INLA methodology was developed for such hierarchical models. However in our application, the hyperparameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M)$ are non-stochastic, fixed values set by the risk manager or regulators.

Bayesian inference is based on the posterior distribution. The posterior is the distribution of the regression coefficients given the observed defaults, that is

$$\pi(\beta_j|\mathbf{y}) = \pi(\beta_j, \mathbf{y})/\pi(\mathbf{y}) \propto \pi(\mathbf{y}|\beta_j)\pi(\beta_j). \tag{3}$$

As the default probabilities are given by (2), their distribution is easily obtained from the posterior values of the regression coefficients.

3.2.1 Choice of prior

We will use normal priors for all the regression coefficients. In principle, any proper probability density function may be used as a prior for the regression coefficients. However, using other distributions, such as skewed or fat-tailed ones, requires specific knowledge or opinions about the distribution of the regression coefficients. In our opinion, this kind of information will rarely be available. Also, the INLA method only allows normal priors.

3.2.2 Prior of the default probabilities

The task of eliciting prior information about the regression coefficients is extremely difficult, in particular because the interpretation of the coefficients depends on the choice of link function.

Rather than defining a prior distribution on the regression coefficients β_j, we suggest that one defines prior distributions on the default probabilities p_i. Assume that the expected values and standard deviations of the default probabilities are specified, and that the default probability p is assumed to be beta distributed. We then have that the relationship between the parameters a and b of the beta distribution and its expected value and variance is given by

$$\mathrm{E}[p] = \frac{a}{a+b}$$

and

$$\mathrm{Var}[p] = \frac{ab}{(a+b)^2(a+b+1)},$$

where $a > 0$ and $b > 0$. For each company i, $i = 1, ..., N$, we can sample a default probability p_i from the beta distribution and thereafter an observation $Y_i \sim \text{Binomial}(p_i)$. Further, we can do a GLM regression with the drawn Y_i variables and the observed account variables, which gives a set of regression parameters $\beta_1, \ldots, \beta_M$. This procedure is repeated many times to obtain realisations of $\beta_1, \ldots, \beta_M$ which belong to $p_i, i = 1, \ldots, N$. Finally we can estimate the expectation and variance of the coefficients by calculating the empirical mean and variance of the realisations.

These values can thereafter be used as prior expectation and variance of the regression coefficients in a Bayesian logistic regression. Further, by studying the histograms of the realisations we have seen that the normal distribution seems to fit well.

One apparently big drawback with this method is that there are many default probabilities if the credit portfolio is large. However, we may have prior runs from other models that makes it possible to specify all these priors automatically.

Our approach is a special case of the general conditional means priors framework for GLMs proposed by Bedrik et al. (1996). However, rather than using simulation to convert a prior on the default probabilities, they derive the induced prior analytically.

3.3 Latent Gaussian models

In the class of structured additive regression models, the response variables are, as in GLM, assumed to belong to an exponential family. The linear predictor takes the more general form

$$\eta_i = \beta_0 + \sum_{k=1}^{K} f^{(k)}(u_{ik}) + \sum_{j=1}^{M} \beta_j x_{ij}. \tag{4}$$

The $\{f^{(k)}(\cdot)\}$'s are unknown finite-dimensional functions of a set of covariates $\mathbf{u}$. As in the GLM (1) the $\{\beta_j\}$'s represent the linear effect of the covariates $\mathbf{x}$. Latent Gaussian models is a subset of Bayesian additive models with the linear predictor on the form (4) which have Gaussian priors for the regression functions- and parameters, $\{f^{(k)}(\cdot)\}$ and $\{\beta_j\}$. Hence, the Bayesian logistic regression model is a special case of the latent Gaussian model (4) with only linear terms in the linear predictor, that is, without the terms $\{f^{(k)}(\cdot)\}$.

4 Methods

4.1 INLA

This section outlines the INLA method. Readers who wish to go into the details of the approximations and numerical issues of INLA are referred to Rue et al. (2009).

In order to use a notation similar to that of Rue et al. (2009), we assume for now that the hyperparameters are stochastic and that the observations are independent conditional on the regression coefficients and the hyperparameters. The posterior (3) may be rewritten as

$$\begin{aligned}\pi(\boldsymbol{\beta},\boldsymbol{\theta}|\mathbf{y}) = \pi(\boldsymbol{\beta},\boldsymbol{\theta},\mathbf{y})/\pi(\mathbf{y}) &\propto \pi(\mathbf{y}|\boldsymbol{\beta},\boldsymbol{\theta})\pi(\boldsymbol{\beta}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \\ &= \prod_i \pi(y_i|\boldsymbol{\beta},\boldsymbol{\theta})\pi(\boldsymbol{\beta}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}).\end{aligned}$$

INLA is based on a nested expression of the posterior marginal distributions of β_j,

$$\pi(\beta_j|\mathbf{y}) = \int \pi(\beta_j,\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta} = \int \pi(\beta_j|\boldsymbol{\theta},\mathbf{y})\pi(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}.$$

The key feature of INLA is to approximate $\pi(\beta_j|\mathbf{y})$ by

$$\widehat{\pi}(\beta_j|\mathbf{y}) = \int \widetilde{\pi}(\beta_j|\boldsymbol{\theta},\mathbf{y})\widetilde{\pi}(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}, \tag{5}$$

where $\widetilde{\pi}(\beta_j|\boldsymbol{\theta},\mathbf{y})$ and $\widetilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ are approximations of the conditional densities. The integration is performed using numerical integration with respect to $\boldsymbol{\theta}$. One obtains the approximate posterior by

$$\widetilde{\pi}(\beta_j|\mathbf{y}) = \sum_r \widetilde{\pi}(\beta_j|\boldsymbol{\theta}_r,\mathbf{y})\widetilde{\pi}(\boldsymbol{\theta}_r|\mathbf{y})\Delta_r, \tag{6}$$

where the sum is over values of $\boldsymbol{\theta}$ with area weights Δ_r. Rue et al. (2009) discuss several approximations of the conditional distributions in the sum (6) (and the integral (5)).

Having derived the above expressions, the INLA approach for approximating the posterior marginals is computed in three steps. The first step approximates the posterior marginal distribution of the hyperparameters using the relationship

$$\pi(\boldsymbol{\theta}|\mathbf{y}) \propto \frac{\pi(\boldsymbol{\beta},\boldsymbol{\theta},\mathbf{y})}{\pi(\boldsymbol{\beta}|\boldsymbol{\theta},\mathbf{y})}.$$

Rue et al. (2009) suggest the approximation

$$\widetilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) \propto \left. \frac{\pi(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{y})}{\widetilde{\pi}_G(\boldsymbol{\beta}|\boldsymbol{\theta}, \mathbf{y})} \right|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*(\boldsymbol{\theta})},$$

where the denominator is the Gaussian approximation to the full conditional of $\boldsymbol{\beta}$ evaluated in the mode for a given value of the hyperparameters.

The second step of INLA is to compute an approximation of the posterior distribution of the regression parameters, given the observed response variables and the hyperparameters. Rue et al. (2009) suggest several different approximations for $\pi(\beta_j|\boldsymbol{\theta}, \mathbf{y})$. They show that in some applications a Gaussian approximation may be sufficient, and that the Laplace or simplified Laplace approximation works well in most cases, see Rue et al. (2009) for details.

The last step of INLA combines the previous two steps using the numerical integration (6) to obtain the posterior distribution of the regression parameters given the observed response variables.

4.1.1 INLA approximation error

Rue et al. (2009) provide several examples of applications of the INLA approach and comparisons with results obtained from intensive MCMC runs. The examples include using simulated data, generalised linear mixed model for longitudinal data, a stochastic volatility model and two spatial models.

The approximation error of INLA is inherited from the error of the two approximations involved and the error of the numerical integration. The only way to assess the error with certainty is to run an MCMC for infinite time. Rue et al. (2009) propose two strategies to assess the approximation error. For each of the approximations, they follow Spiegelhalter et al. (2002) and compare the effective number of parameters to the size of the data set. In addition they compare the difference between the approximations and their Taylor expansions to the size of the data set. The second strategy is to compute more and more accurate approximations, that is, use the Gaussian, simplified Laplace and the Laplace, and compare the corresponding symmetric Kullback-Leibler divergences (Kullback, 1987; Kullback and Leibler, 1951). A small divergence is taken as a sign of acceptable approximation error, otherwise the approximation is labelled “problematic”. Rue et al. (2009) state that they have yet not come across examples of the latter.

The examples of Rue et al. (2009) confirm that INLA provides fast and accurate inference compared to MCMC. In our case, we compare the INLA results to MCMC runs. We focus on an MCMC run with an appropriate number of simulations, but we have also run MCMC for a very long ("infinite") time.

4.2 MCMC

It is not the purpose of this paper to provide a deep understanding of MCMC. For a review of MCMC methods, see Gilks et al. (1996). Applied to the Bayesian logistic model of Section 3.2, the idea of MCMC is to run an ergodic Markov chain which has (3) as stationary distribution. Starting from a set of given initial values, in our case the parameter estimates obtained from standard (non-Bayesian) GLM, the chain is run until it is believed to have reached equilibrium. Thereafter, the chain produces a sequence of dependent samples from the posterior distribution.

4.2.1 MCMC convergence

Assessing the convergence of a sequence of realisations of an MCMC run, is a non-trivial task. A variety of diagnostic tools exist, see Brooks and Roberts (1999) for a review. A collection of commonly applied convergence measures is implemented in the CODA package, Plummer et al. (2006), Cowles and Carlin (1996). We have chosen two of these measures, namely the methods of Geweke (1992) and Raftery and Lewis (1992).

The method of Geweke (1992) compares averages of simulations from the first and last part of the chain. The test statistic is the ratio of the difference of these averages to the standard error of the difference. As the chain converges the difference will tend to be small compared to the standard error, and Geweke (1992) shows that a normal distribution may be used to test for significance.

The method of Raftery and Lewis (1992) and Raftery and Lewis (1995) examines convergence for a given quantile, precision and probability of achieving the specified precision. The method finds the minimum number of iterations that should be run, and a recommended number based on an assessment of the dependency between subsequent simulations.

4.3 Software

All our computations were performed in the R (`www.r-project.org`) environment for statistical computations and graphics. The computations were run on a Dell desktop PC with a Core 2 Duo 2.13 GHz processor and 2GB RAM.

The INLA-software developed by Rue & Martino is down-loadable at the homepage of Håvard Rue (http://www.math.ntnu.no/∼hrue), and it will shortly be released as a package of R.

We ran MCMC using MCMCpack (`http://mcmcpack.wustl.edu/`) and the function `MCMClogit`. The sampling engine of MCMCpack is a Metropolis-Hastings algorithm (Metropolis et al., 1953). To maximise computational efficiency, the actual sampling for each model is done in compiled C++ using the Scythe Statistical Library, which makes the simulations very fast.

The posterior samples returned by MCMCpack are so-called “mcmc objects”, which can easily be summarised and manipulated using the CODA package (Cowles and Carlin, 1996; Plummer et al., 2006).

5 Results

5.1 Results with a vague prior

In order to study the impact of having prior information on the default probabilities, we first present the results obtained when no prior is incorporated (or in the Bayesian framework; when the prior is vague or non-informative). Tables 2 and 3 show the regression coefficients obtained with a vague prior (prior mean 0 and prior variance 1000) using MCMC and INLA, respectively. We observe that the coefficients obtained with the two models are practically the same.

Variable	*Mean*	*SD*	*Lower*	*Upper*
(Intercept)	3.9594	0.6142	2.6644	5.0837
PROFIT	-0.0490	0.0959	-0.2315	0.1295
LOSS	-0.1733	0.0924	-0.3534	0.0149
CPINV	-0.0804	0.0314	-0.1362	-0.0174
DP	-0.1124	0.0487	-0.2065	-0.0101
EQPER	-0.0500	0.0471	-0.1355	0.0433
MPAY	0.0624	0.0464	-0.0276	0.1528
DLIQ	-0.0798	0.0459	-0.1715	0.0254
ODFAC	-0.2241	0.0460	-0.3270	-0.1301
REPHIST	-0.2251	0.0299	-0.2851	-0.1677
REPKF	-0.1284	0.0264	-0.1789	-0.0671
REPAC	-0.0002	0.0256	-0.0507	0.0513
AD	-0.0579	0.0255	-0.1117	-0.0088
AGE	-0.0945	0.0387	-0.1709	-0.0207

Table 2. Posterior mean, standard deviation, lower quantile (2.5%) and upper quantile (97.5%) using MCMC with 10000 iterations and a vague prior.

Variable	*Mean*	*SD*	*Lower*	*Upper*
(Intercept)	3.9263	0.6606	2.6238	5.2318
PROFIT	-0.0539	0.0907	-0.2302	0.1284
LOSS	-0.1696	0.0974	-0.3628	0.0218
CPINV	-0.0798	0.0307	-0.1396	-0.0181
DP	-0.1167	0.0499	-0.2145	-0.0174
EQPER	-0.0504	0.0476	-0.1432	0.0448
MPAY	0.0626	0.0474	-0.0319	0.1552
DLIQ	-0.0771	0.0457	-0.1694	0.0111
ODFAC	-0.2195	0.0441	-0.3086	-0.1342
REPHIST	-0.2277	0.0298	-0.2863	-0.1686
REPKF	-0.1252	0.0268	-0.1770	-0.0709
REPAC	-0.0028	0.0264	-0.0545	0.0498
AD	-0.0615	0.0247	-0.1095	-0.0118
AGE	-0.0882	0.0357	-0.1603	-0.0190

Table 3. Posterior mean, standard deviation, lower quantile (2.5%) and upper quantile (97.5%) using INLA with a vague prior.

5.2 Results with a specific prior

Now we use prior information given by the Sparebank 1-alliance, based on customer information in the time period 1994–2000. The prior mean and standard deviation of the coefficients are shown in Table 4. INLA does not let the user specify any prior on the intercept, so in order to compare the MCMC results with INLA, we use the default prior of INLA also in the MCMC algorithm. This prior uses mean 0 and standard deviation $\sqrt{1000}$ for all the coefficients. It is often difficult for risk managers to interpret and give expert opinions on the intercept, so applying a vague prior on the intercept makes sense.

Variable	*Prior mean*	*Prior standard deviation*
PROFIT	-0.096	0.042
LOSS	-0.127	0.104
CPINV	-0.045	0.022
DP	-0.069	0.024
EQPER	-0.102	0.023
MPAY	-0.103	0.023
DLIQ	-0.032	0.019
ODFAC	-0.078	0.048
REPHIST	-0.268	0.021
REPKF	-0.102	0.017
REPAC	-0.105	0.017
AD	-0.068	0.015
AGE	-0.062	0.017

Table 4. Prior mean and standard deviation of the explanatory variables.

5.2.1 MCMC

Table 5 shows the posterior mean, standard deviation and the 2.5% and 97.5% quantiles of the MCMC run using 10000 iterations. Compared to the vague prior results in Table 2, we see that the specific prior given in Table 4 influences the MCMC results significantly. Figure 2 shows the prior and the posterior distributions of the coefficients, using MCMC with 10000 iterations. The posterior distribution is presented as a density plot of all 10000 realisations of the posterior, hence the unsmooth shape. The MCMC run with an infinite number of simulations represents the "true" posterior. We ran MCMC with 10^6 iterations. The posterior obtained from this run did not differ significantly from the result obtained using 10000 iterations.

Variable	*Mean*	*SD*	*Lower*	*Upper*
(Intercept)	3.5773	0.4951	2.5888	4.4994
PROFIT	-0.0775	0.0369	-0.1555	-0.0077
LOSS	-0.0930	0.0625	-0.2087	0.0344
CPINV	-0.0590	0.0178	-0.0945	-0.0284
DP	-0.0817	0.0203	-0.1245	-0.0459
EQPER	-0.0903	0.0188	-0.1241	-0.0515
MPAY	-0.0721	0.0220	-0.1152	-0.0258
DLIQ	-0.0322	0.0175	-0.0664	-0.0008
ODFAC	-0.1362	0.0339	-0.1992	-0.0696
REPHIST	-0.2442	0.0174	-0.2789	-0.2121
REPKF	-0.1047	0.0140	-0.1292	-0.0772
REPAC	-0.0673	0.0151	-0.0943	-0.0356
AD	-0.0600	0.0126	-0.0860	-0.0341
AGE	-0.0683	0.0150	-0.0964	-0.0421

Table 5. Posterior mean, standard deviation, lower quantile (2.5%) and upper quantile (97.5%) using MCMC with 10000 iterations and a specific prior.

5.2.2 INLA

Table 6 shows the posterior distributions obtained using INLA. Compared to the results obtained using the vague prior (Table 3), the posteriors are here clearly affected by the prior. Figure 3 shows the prior and the posterior distributions of the coefficients. The posterior density distributions are drawn from the normal distribution with the achieved approximations of the posterior means and standard deviations.

5.2.3 Comparison

Figure 4 shows the posterior distributions obtained with the specific prior, using INLA and MCMC with 10000 iterations. Figure 5 also compares the results of INLA and MCMC, but also includes the posterior results obtained using a vague prior, see Section 5.1. The results obtained with informative and vague priors differ considerably for most of the variables. When the prior and the empirical data have diverging information, the posterior is mixture of the two, as given by the Bayesian approach.

We see that the posterior distributions of MCMC and INLA are similar. Since the MCMC posterior probability is a density plot of the realisations from the MCMC iterations, it is unsmooth. The more iterations, the smoother the density plot is. If we compare the posterior means, standard deviations and quantiles in Table 5 and 6 we see that MCMC and INLA give very similar results.

Variable	*Mean*	*SD*	*Lower*	*Upper*
(Intercept)	3.6248	0.4996	2.6366	4.6089
PROFIT	-0.0847	0.0371	-0.1581	-0.0114
LOSS	-0.0945	0.0617	-0.2166	0.0271
CPINV	-0.0593	0.0179	-0.0945	-0.0240
DP	-0.0799	0.0211	-0.1216	-0.0384
EQPER	-0.0903	0.0199	-0.1297	-0.0510
MPAY	-0.0730	0.0207	-0.1139	-0.0322
DLIQ	-0.0307	0.0171	-0.0646	0.0029
ODFAC	-0.1369	0.0308	-0.1983	-0.0768
REPHIST	-0.2445	0.0167	-0.2775	-0.2115
REPKF	-0.1039	0.0143	-0.1320	-0.0757
REPAC	-0.0686	0.0137	-0.0957	-0.0415
AD	-0.0598	0.0125	-0.0844	-0.0352
AGE	-0.0705	0.0152	-0.1005	-0.0405

Table 6. Posterior mean, standard deviation, lower quantile (2.5%) and upper quantile (97.5%) using INLA with a specific prior.

With the size of our data set, running the MCMC simulation with 10000 iterations is done in a short amount of time. There are many methods to check for convergence of an MCMC chain. We have focused on the methods of Geweke (1992) and Raftery and Lewis (1992), see Section 4.2.1. According to the convergence methods some of the coefficients have not converged after 10000 iterations, even though it looks like they have converged from the trace plots of the MCMC run. This illustrates how difficult it is to determine when an MCMC run has converged. If we measure convergence by eye, MCMC and INLA give approximately the same results during approximately the same amount of time.

If the prior information, unlike the case in our study, differs dramatically from what the data alone implies, MCMC will converge very slowly. In such cases, INLA is favoured as it will be computationally faster. Also, a great advantage of INLA is that the user does not have to worry about the issue of convergence.

6 Conclusions

The main contribution of this paper is to introduce INLA as an alternative to MCMC for Bayesian credit risk modelling. We have compared the two methods in a Bayesian logistic regression setting. We have modelled the credit default risk of a dataset of real default data provided by the Norwegian Sparebank 1-alliance.

The Bayesian approach makes is possible to incorporate prior knowledge on the regression coefficients. In our case, specific prior information on the default probabilities was given by the Sparebank 1-alliance and converted to prior information on the regression coefficients in the logistic regression model.

We find that INLA and MCMC give approximately the same posterior results. In our case, the MCMC algorithm converges quickly, and hence it is not obvious that INLA is the best choice. However, the main advantage with INLA for this purpose is that the user does not have to worry about convergence issues. Also, INLA gives the same posterior for each data set, and hence the approximation error is easy to deal with. The MCMC approach gives different posteriors from one run to another, and the MCMC error is therefore more difficult to track. The only "true" posterior is obtained by running the MCMC for an infinitely long time. However, as “infinitely long” must be quantified in terms of a given number of simulations, one can never be certain that the true equilibrium is actually reached.

Acknowledgments

This work was supported by the Research council of Norway, through grant number 186951 (I30, Modern Application driven Statistical Challenges). The authors thank Sara Martino and Håvard Rue for providing the INLA software and guidance on using it. We thank the Sparebank 1-alliance, KFK (Centre of Excellence Credit Risk Modelling) for kindly providing credit default data.

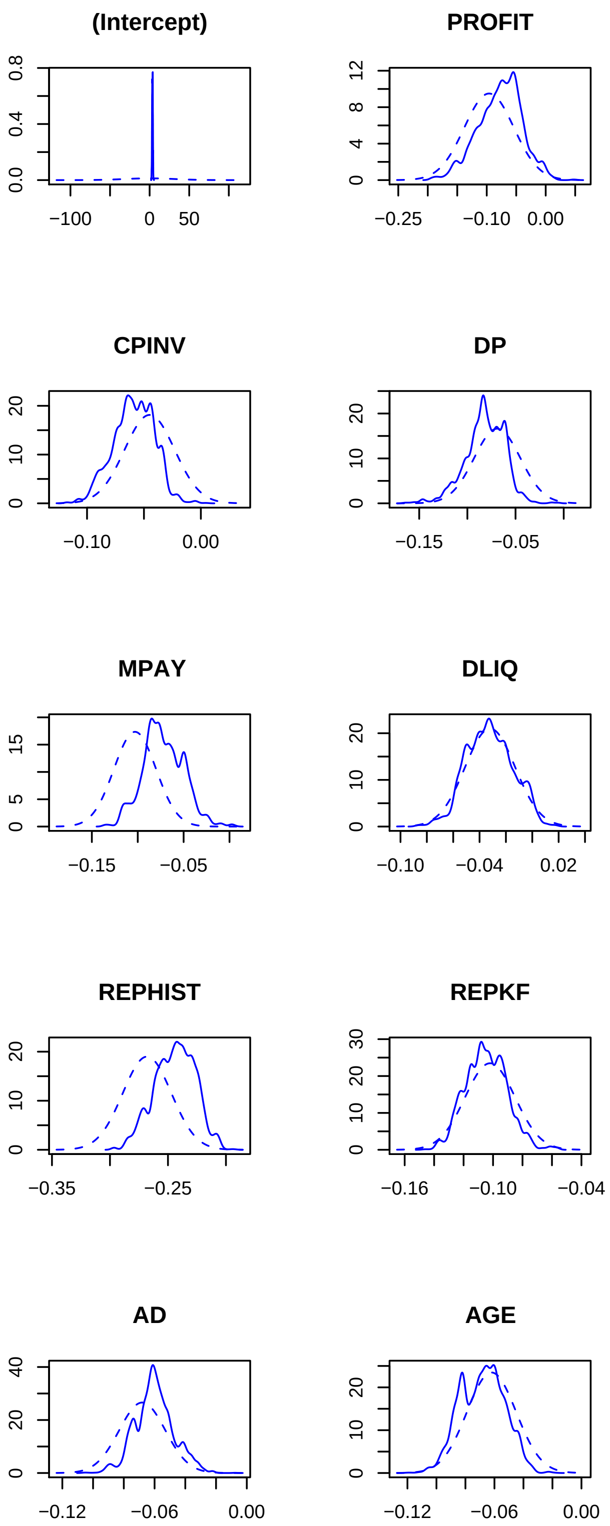


Figure 2. The prior distribution (dashed line) and the posterior distribution (continuous line) of the coefficients, using MCMC with 10000 iterations and a specific prior.

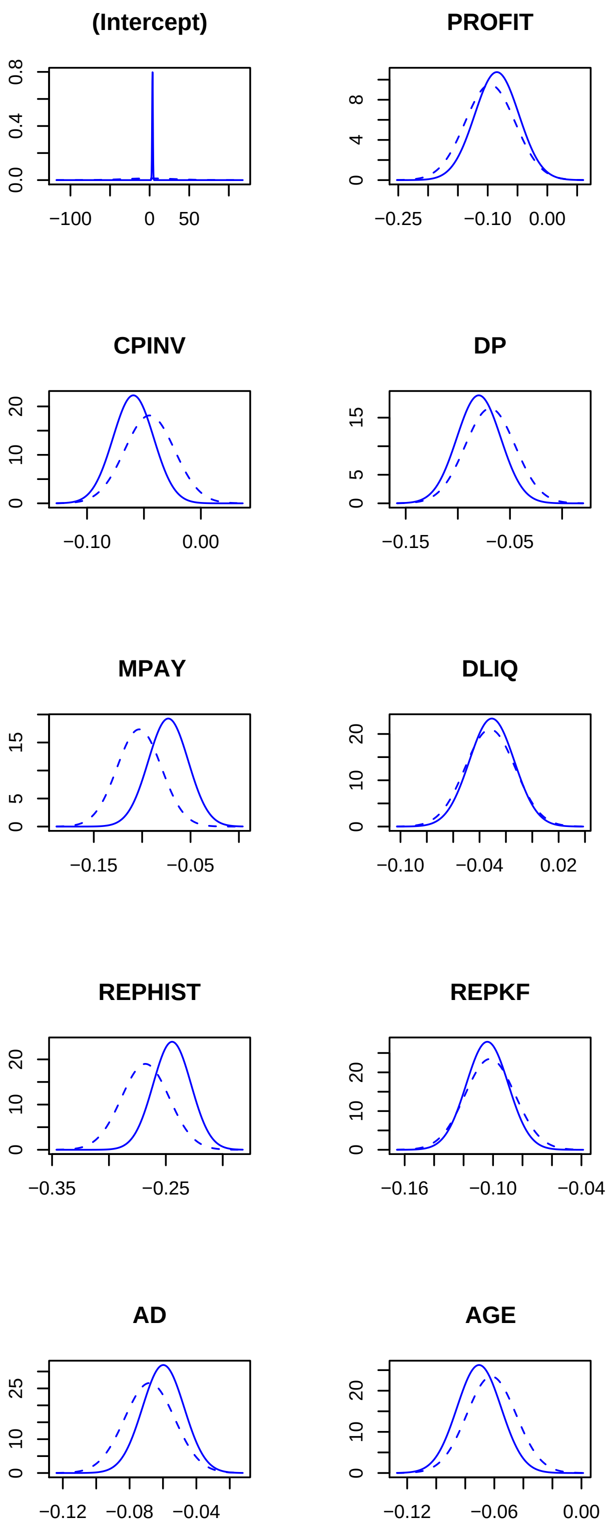


Figure 3. The prior distribution (dashed line) and the posterior distribution (continuous line) of the coefficients, using INLA with a specific prior.

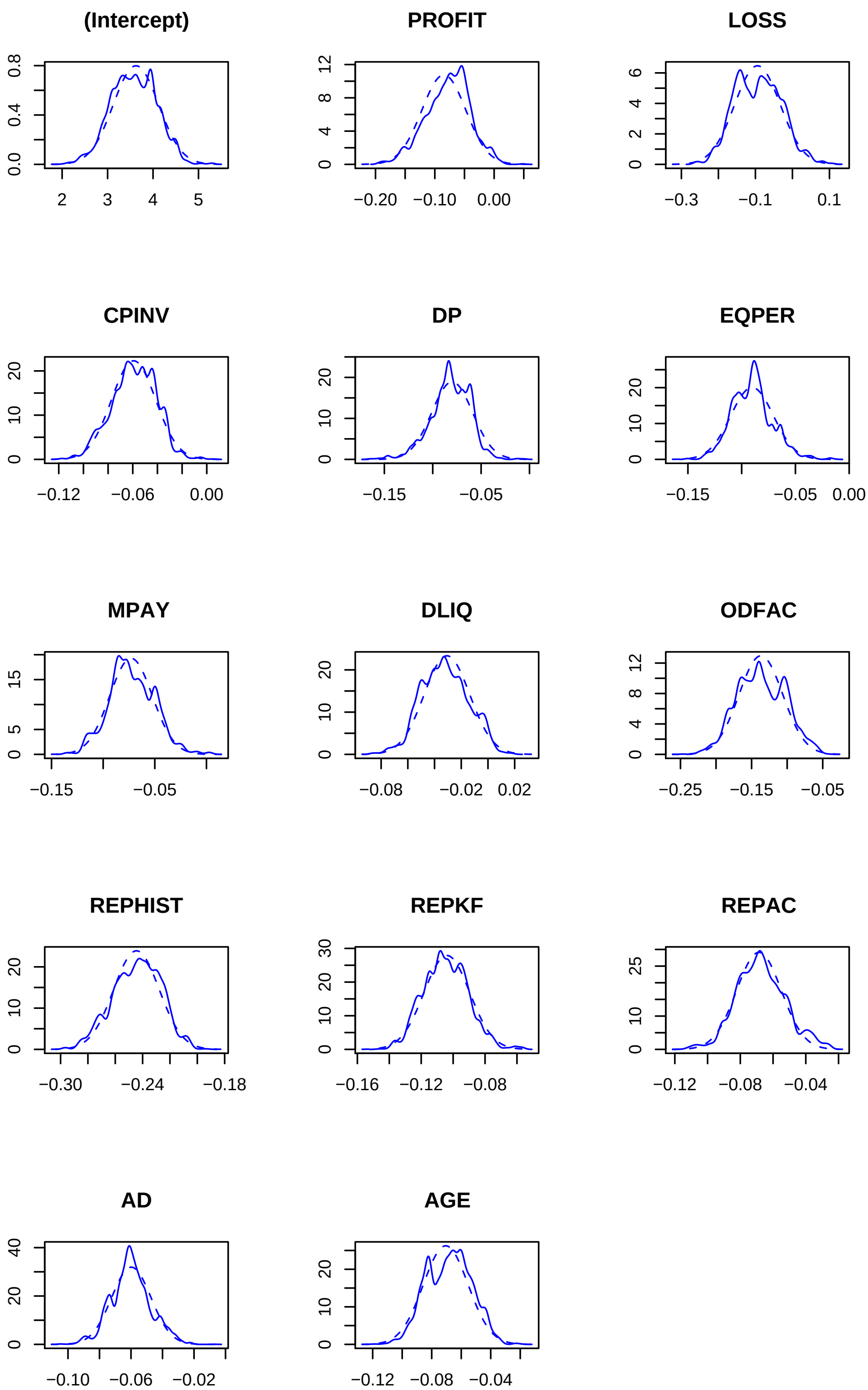


Figure 4. The posterior distributions obtained with the specific prior, using INLA (dashed line) and MCMC with 10000 iterations (continuous line).

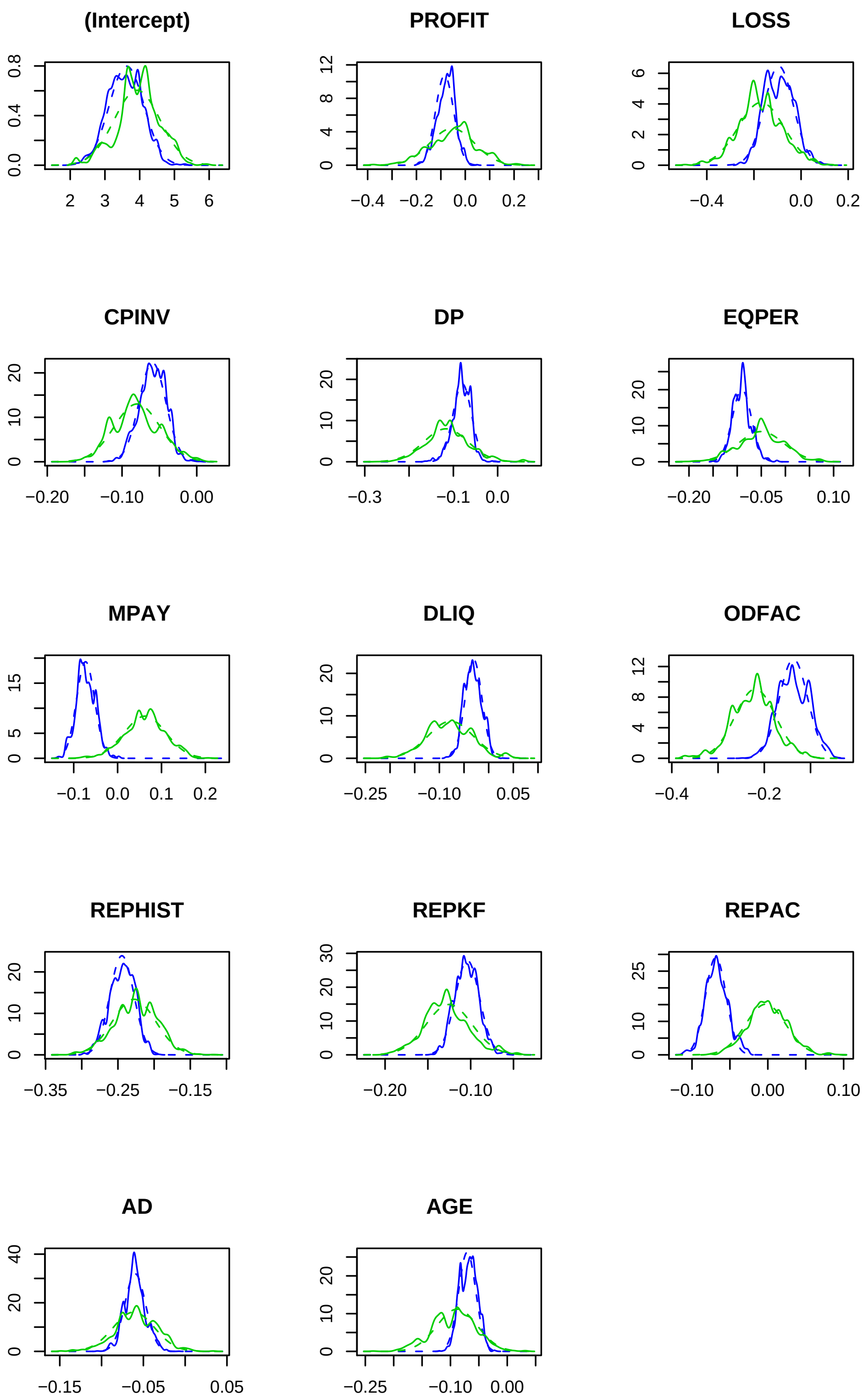


Figure 5. The posterior distribution obtained with INLA, using a specific and a vague prior (blue- and green dashed lines, respectively). The posterior distribution obtained using MCMC with 10000 iterations, using a specific and a vague prior (blue- and green continuous line, respectively).

References

Altman, E. I. (1968). Financial rations, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*, 23(4):589–609.

Ando, T. (2006). Bayesian credit rating analysis based on ordered probit model with functional predictor. In *Prooceding of the third IASTED Int. conference on financial engeneering and applications*, pages 69–76, Cambridge, USA.

Atiya, A. F. (2001). Bankruptcy prediction for neural networks: A survey and new results. *IEEE Transactions on neural networks*, 12(4):929–935.

Basel II (2006). Basel II: International convergence of capital measurements and captial standards: A revised framework, comprehensive version. Technical report, Bank for international settlements.

Beaver, W. H. (1966). Financial rations as predictors of failure. *Journal fo accounting research*, 4:71–111.

Bedrik, E. J., Christensen, R., and Johnson, W. (1996). A new percepcitve on priors for generalized linear models. *Journal of the American Statistical Association*, 91(436):1450–1460.

Berg, D. (2007). Bankrupcty prediction by generlized additive models. *Applied stochasstic models in business and industry*, 23(2):129–143.

Brooks, S. and Roberts, G. (1999). Assessing convergence of Markov chain Monte Carlo algorithms. *Statist. Comput.*, 8:319–335.

Cowles, M. K. and Carlin, B. P. (1996). Markov Chain Monte Carlo diagnostics: A comparative review. *Journal of the American Statistical Association*, 91:883–904.

Crosbie, P. J. and Bohn, J. R. (2002). Modeling default risk. KMV working paper. Available from www.kmv.com.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. *Bayesian Statistics 4*.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.

Kullback, S. (1987). The kullback-leibler distance. *The American statistician*, 41:340–341.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22:79–86.

Löffler, G., Prosch, P. N., and Schöne, C. (2005). Bayesian methods for improving credit scoring models. Technical report, Department of Finance, University of Ulm, Germany.

Martin, A. D., Quinn, K. M., and Park, J. H. (2008). Package 'MCMC pack'. Technical report, Washington University in St.Louis.

Mc Neil, A. J. and Wedin, J. P. (2007). Bayesian inference for generlized linear mixed models of portfolio credit risk. *Journal of empricial finance*, 14(2):131–149.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman and Hall, London, second edition.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equations of state calculations by fast computing machine. *Journal of chemical physics*, 21:1087–1091.

Mira, A. and Tenconi, P. (2004). Bayesian estimate of credit risk via MCMC with delayed rejection. In *Seminar on Stochastic Analysis, Random Fields and Applications IV, Centro Stefano Franscini, Ascona, May 2002*, pages 277–291. Birkhauser Verlag, Basel.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, 6(1):7–11. Available from: `http://CRAN.R-project.org/doc/Rnews/`.

Raftery, A. and Lewis, S. (1992). One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo. *Statistical Science*, 7:493–497.

Raftery, A. and Lewis, S. (1995). *The number of iterations, convergence diagnostics and generic Metropolis algorithms.* In *Practical Markoc Chain Monte Carlo*. Chapman & Hall, London, U.K.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate bayesian inference for latent gaussian models using integrated nested Laplace approximations. *J.R.Statistics Soc B*, 71(2):1–35.

Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2007). OpenBUGS user manual. Technical report, MRC Biostatistics unit, Cambridge, http://mathstat.helsinki.fi/openbugs/.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model compexity and fit (with discussion). *Journal of the Royal Statistical Society*, 64(2):583–639.