The Bayesian Approach to Default Risk Analysis and the Prediction of Default Rates

Michael Jacobs Jr.[1]

Nicholas M. Kiefer[2]

Office of the Comptroller of the Currency and Cornell University[3]

June 11, 2011

[1]US Department of the Treasury, Office of the Comptroller of the Currency, Credit Risk Analysis Division.

[2]Cornell University, Departments of Economics and Statistical Science, 490 Uris Hall, Ithaca, NY 14853-7601, US. email:nicholas.kiefer@cornell.edu; US Department of the Treasury, Office of the Comptroller of the Currency.

[3]*Disclaimer: The statements made and views expressed herein are solely those of the author, and do not necessarily represent official policies, statements or views of the Office of the Comptroller of the Currency or its staff.*

Abstract

A Bayesian approach to default rate estimation is used to predict default rates on the basis of information from data and experienced industry experts. The principle advantage of the Bayesian approach is the potential for coherent incorporation of expert information - crucial when data are scarce or unreliable. A secondary advantage is access to efficient computational methods such as Markov Chain Monte Carlo. The power of this approach is illustrated using annual default rate data from Moody's (1999-2009) for two risk buckets and priors elicited from industry experts. Three structural credit models in the asymptotic single risk factor (ASRF) class underlying the Basel II framework (Generalized Linear and Generalized Linear Mixed Models), are analyzed using a Markov Chain Monte Carlo technique. The predictive distributions for defaults are obtained.

Keywords: Basel II, risk management, prior elicitation, maximum entropy, MCMC.

1 Introduction

Estimation and prediction of default rates for groups of homogeneous assets (portfolio "buckets") is essential for determining adequate capital. The Basel II (B2) framework (Basel Committee on Banking Supervision (2006)) for calculating minimum regulatory capital requirements provides for banks to use models to assess credit (and other) risks. Of course, specification of a model requires definition of parameters or quantities of interest, specification of the parameter space, identification of relevant data, and decisions on how these all fit together. Each step requires judgment and is subject to criticism. Strong justification is required. Indeed this justification is an important part of the validation procedure expected of financial institutions (OCC (2000) has been extremely influential; there is new guidance in OCC (2011)).In response to the credit crisis, the Basel Committee has stressed the continuing importance of quantitative risk management, see Basel Committee on Banking Supervision (2009). Our emphasis is on the incorporation of nondata information and the prediction of defaults, so we focus on elicitation and representation of expert information and then on the Bayesian approach to inference in the context of a series of default-risk models. The first two models are consistent with the asymptotic single-factor model underlying B2. The third adds temporal correlation in asset values, generalizing B2. Uncertainty about the default probability should be modeled the same way as uncertainty about defaults – represented in a probability distribution. A future default either occurs or doesn't, given the definition. Since it is not known in advance whether default occurs or not, it is conventional to the point of never being questioned to model this uncertain event with a probability distribution. Similarly, the default probability is unknown. But there is information available about the default rate in addition to the data information. The simple fact that loans are priced and grouped into risk buckets shows that some risk assessment is occurring. This information should be organized and incorporated in the analysis in a coherent way, specifically represented in a probability distribution. We discuss elicitation briefly in section 2, then run through the steps of a formal Bayesian analysis. A step-by-step guide to Bayesian analysis in the default setting, including details on elicitation of expert information, is available in Jacobs and Kiefer (2010). We then turn to the predictive distributions of defaults. There is temporal variation in default rates implied by the single-factor model and there is also some prediction uncertainty due to the fact that Bayesians do not condition on particular values of unknown parameters, but marginalize over them, taking account of this uncertainty instead of ignoring it. We apply the approach to two data sets, one typical of a mid-portfolio bucket of commercial loans, the other consisting of relatively safe loans (hence fewer defaults and less data information on default rates). We find that our procedures give sensible, useful results. We also find that the asset correlations specified by B2 are higher than the data warrant (though they may be the values that give the appropriate capital levels).

In section 2 we consider elicitation of expert information and its representation in a probability distribution. We favor the maximum entropy (ME) representation. In section 3 we describe the sequence of statistical models for the default rates. Section 4 treats the mid-portfolio application and section 5 the safe portfolio bucket. Section 6 concludes.

2 Elicitation and Representation of Expert Information

Elicitation of prior distributions is an area that has attracted attention. General discussions of the elicitation of prior distributions are given by Garthwaite, Kadane, and O'Hagan (2005), O'Hagan, Buck, Daneshkhah, Eiser, Garthwaite, Jenkinson, Oakley, and Rakow (2006), Cooke (1991), Kadane and Wolfson (1998), and Jacobs and Kiefer (2010). A stylized representation of elicitation consists of four stages. First, we prepare for the elicitation by identifying the expert, training the expert to think in terms of probabilities (in this application, the experts are generally comfortable with probabilities), and identifying what aspects of the problem to elicit. Second, we elicit specific summaries of the experts' distributions for those aspects. Third, we fit a probability distribution to those summaries elicited in the second step. Finally, we note that elicitation is in almost all cases an iterative process, so that the final stage is an assessment of the adequacy of the elicitation, which leaves open the possibility of a return to earlier stages in order to gather more summaries from the expert. For example, the fitted prior distribution of the PD parameter may be presented to the expert, and if the expert is not comfortable with the shape for whatever reason, we may try to gather more quantiles, re-fit and return later to make further assessments. We merely sketch the elicitation and representation of the priors here: details on the mid-portfolio prior are given in a previous application, Kiefer (2010) and on the low-default prior in Kiefer (2009b).

Step 3, representing elicitated information to a statistical distribution, is often done by specifying a functional form for the statistical distribution and choosing parameters to match the elicited information. For example, the Beta distribution might be chosen to represent prior information on a default rate. We have used this approach but have come to prefer the ME approach. ME provides a method to specify the distribution that meets the expert specifications and imposes as little additional information as possible. Thus, we maximize the entropy (minimize the information) in the distribution subject to the constraints (indexed by k given by the assessments). Entropy (differential) is

$$H(p) = -\int \log(p(x))dP$$

Entropy is a widely used measure of the information in an observation (or an experiment). Further discussion from the information theory viewpoint can be found in Cover and Thomas (1991). The general framework is to solve for the distribution p

$$\max_p \{-\int p \ln(p(x))dx\} \tag{1}$$

$$\begin{aligned} s.t. \int p(x)c_k(x)dx = 0 \ for \ k = 1, ..., K \\ and \int p(x)dx = 1 \end{aligned}$$

In our application the assessed information consists of quantiles. The constraints are written in terms of indicator functions for the α_k quantiles q_k; for example the median constraint corresponds to $c(x) = I(x < median) - 0.5$. To solve this maximization problem, form the Lagrangian with multipliers λ_k and μ and differentiate with respect to $p(x)$ for each x. Solving the resulting first-order conditions gives

$$p_{ME}(\theta) = \kappa \exp\{\sum_k \lambda_k(I(\theta < q_k) - \alpha_k)\} \tag{2}$$

The multipliers are chosen so that the constraints are satisfied. For details see Cover and Thomas (1991) or for an approach not using the Lagrangian Csiszar (1975).

The discontinuities in $p_{ME}(\theta)$ due to the indicator functions in the exponent are perhaps unlikely to reflect characteristics of expert information and indeed this was the view of the expert. Smoothing was accomplished using the Epanechnikov kernel with several bandwidths h chosen to offer the expert choices on smoothing level (including no smoothing). Specifically, with $p_S(\theta)$ the smoothed distribution with bandwidth h we have

$$p_S(\theta) = \int_{-1}^{1} K(u) p_{ME}(\theta + u/h) du \tag{3}$$

with $K(u) = 3(1 - u^2)/4$ for $-1 < u < 1$. Since the density $p_{ME}(\theta)$ is defined on bounded support there is an endpoint or boundary "problem" in calculating the kernel-smoothed density estimator. Specifically, $p_S(\theta)$ as defined in (3) has larger support than $p_{ME}(\theta)$, moving both endpoints out by a distance $1/h$. We adjust for this using reflection, $p_{SM}(\theta) = p_S(\theta) + p_S(a - \theta)$ for $a \leq \theta < a + 1/h$, $p_{SM}(\theta) = p_S(\theta)$ for $a + 1/h \leq \theta < b - 1/h$, and $p_{SM}(\theta) = p_S(\theta) + p_S(2b - \theta)$ for $b - 1/h \leq \theta \leq b$. See Schuster (1985).

3 Models for Defaults: Likelihood Functions

The simplest probability model for defaults of assets in a homogeneous segment of a portfolio is the Binomial, in which the defaults are assumed independent across assets and over time, and occur with common probability $\theta \in [0, 1]$. This is invariably the starting model for default analysis, and in many cases it is also the final model. Large institutions will typically go farther. Suppose the value of the ith asset in time t is

$$v_{it} = \epsilon_{it}$$

where ϵ_{it} is the time and asset specific shock (idiosyncratic risk) and default occurs if $v_{it} < T^*$, a default threshold value. A mean of zero is attainable through translation without loss of generality. We assume the shock is standard normal with distribution function $\Phi(\cdot)$. Let d_i indicate whether the ith observation was a default ($d_i = 1$) or not ($d_i = 0$). The distribution of d_i is Bernoulli $p(d_i|\theta) = \theta^{d_i}(1-\theta)^{1-d_i}$, where $\theta = \Phi(T^*)$. Let $D = \{d_i\}_{i=1}^n$ denote the whole data set and $r = r(D) = \sum_i d_i$ the count of defaults. Then the joint distribution of the data is

$$p(D|\theta) = \prod_{i=1}^{n} \theta^{d_i}(1-\theta)^{1-d_i} \tag{4}$$
$$= \theta^r(1-\theta)^{n-r}$$

Since this distribution depends on the data D only through r (n is regarded as fixed), the sufficiency principle implies that we can concentrate attention on the Binomial(n,θ) distribution of r

$$p(r|\theta) = \binom{n}{r}\theta^r(1-\theta)^{n-r} \tag{5}$$

This is Model I.

Basel II suggests there may be heterogeneity due to systematic temporal changes in asset characteristics or to changing macroeconomic conditions. There is some evidence from other markets that default probabilities vary over the cycle. See Nickell, Perraudin, and Varotto (2000) and Das, Duffie, Kapadia, and Saita (2007). The B2 capital requirements are based on a one-factor model due to Gordy (2003) that accommodates systematic temporal variation in asset values and hence in default probabilities. This model can be used as the basis of a model that allows temporal variation in the default probabilities, and hence correlated defaults within years. The value of the ith asset in time t is

$$v_{it} = \rho^{1/2}x_t + (1-\rho)^{1/2}\epsilon_{it} \tag{6}$$

where ϵ_{it} is the time and asset specific shock (as above) and x_t is a common time shock, inducing correlation $\rho \in [0,1]$ across asset values within a period. The random variables x_t are assumed to be standard normal and independent of each other and of the ϵ_{it}. The overall or marginal default rate is $\theta = \Phi(T^*)$. However, in each period the default rate θ_t depends on the systematic factor x_t. The model implies a distribution for θ_t. Specifically, the distribution of v_{it} conditional on x_t is $N(\rho^{1/2}x_t, 1-\rho)$. Hence the period t default probability (also referred to as the conditional default probability) is

$$\theta_t = \Phi[(T^* - \rho^{1/2}x_t)/(1-\rho)^{1/2}]. \tag{7}$$

The distribution function for $\theta_t \in [0,1]$ is given by

$$\begin{aligned} \Pr(\theta_t \leq A) &= \Pr(\Phi[(T^* - \rho^{1/2}x_t)/(1-\rho)^{1/2}] \leq A) \\ &= \Phi[((1-\rho)^{1/2}\Phi^{-1}[A] - \Phi^{-1}[\theta])/\rho^{1/2}] \end{aligned} \tag{8}$$

using the standard normal distribution of x_t and writing $\theta = \Phi(T^*)$. Differentiating gives the density $p(\theta_t|\theta,\rho)$. This is the Vasicek distribution, see e.g. Bluhm, Overbeck, and Wagner (2003) Section 2.5, for details. The parameters are θ, the marginal or mean default probability and the asset correlation ρ. The conditional distribution of the number of defaults in each period is (from (5))

$$p(r_t|\theta_t) = \binom{n_t}{r_t}\theta_t^{r_t}(1-\theta_t)^{n_t-r_t} \tag{9}$$

from which we obtain the distribution conditional on the underlying parameters

$$p(r_t|\theta,\rho) = \int p(r_t|\theta_t)p(\theta_t|\theta,\rho)d\theta_t$$

Since different time periods are independent, the distribution for $R = (r_1, ...r_T)$ is

$$p(R|\theta, \rho) = \prod_{t=1}^{T} p(r_t|\theta, \rho) \tag{10}$$

where we condition on $(n_1, ..., n_T)$, i.e. they are considered to be known. Regarded as a function of (θ, ρ) for fixed R, (10) is the likelihood function. This is Model II.

Model II allows clumping of defaults within time periods, but not correlation across time periods. This is the next natural extension. It goes beyond models considered in B2. Specifically, let the systematic risk factor x_t follow an $AR(1)$ process

$$x_t = \tau x_{t-1} + \eta_t$$

with η_t iid standard normal and $\tau \in [-1, 1]$. Now the formula for θ_t (7) still holds but the likelihood calculation is different and cannot be broken up into the period-by-period calculation, cf. (7). Write using (9)

$$p(R|\theta_1, ...\theta_T) = \prod_{t=1}^{T} p(r_t|\theta_t(x_t, \theta, \rho))$$

emphasizing the functional dependence of θ_t on x_t as well as θ and ρ. Now we can calculate the desired unconditional distribution

$$p(R|\theta, \rho, \tau) = \int \cdots \int \prod_{t=1}^{T} p(r_t|\theta_t(x_t, \theta, \rho))p(x_1, ..., x_T|\tau)dx_1...dx_T \tag{11}$$

where $p(x_1, ..., x_T|\tau)$ is the density of a zero-mean random variable following an $AR(1)$ process with parameter τ. Regarded as a function of (θ, ρ, τ) for fixed R, (11) is the likelihood function. This is Model III.

Model I is a very simple example of a Generalized Linear Model (GLM) (McCullagh and Nelder (1989)). Models II and III are in the form of the General

Linear Mixed Model (GLMM), a parametric mixture generalization of the popular GLM class. These models were analyzed using MCMC in the default application by McNeil and Wendin (2007) using convenience priors and focussing on default rate estimation, and by Kiefer (2009a) using an elicited prior and focussing on predictability of default rates.

4 The Mid-Portfolio Bucket Example

4.1 The Prior

We have asked an expert to consider a portfolio bucket consisting of loans that might be in the middle of a bank's portfolio. These are typically commercial loans to unrated companies. If rated, these might be about Moody's Ba-Baa or S&P BB-BBB. Our expert is an experienced industry (banking) professional with responsibilities in risk management and other aspects of business analytics. He has seen many portfolios of this type in different institutions. The expert found it easier to think in terms of the probabilities directly than in terms of defaults in a hypothetical sample. This is not uncommon in this technical area, as practitioners are accustomed to working with probabilities. We focussed on the elicitation of quantiles, as experience shows these are much easier to think about than moments (small changes in tails can change moments dramatically). The minimum value for the default probability was 0.0001 (one basis point). The expert reported that a value above 0.035 would occur with probability less than 10%, and an absolute upper bound was 0.3. The upper bound was discussed: the expert thought probabilities in the upper tail of his distribution were extremely unlikely, but he did not want to rule out the possibility that the rates were much higher than anticipated (prudence?). Quartiles were assessed by asking the expert to consider the value at which larger or smaller values would be equiprobable given the value was less than the median, then given the value was more than the median. The median value was 0.01. The former, the .25 quartile, was 0.0075. The latter, the .75 quartile, was assessed at .0125. The expert, who has long experience with this category of assets, seemed to be thinking of a distribution with a long and thin upper tail but otherwise symmetric. After reviewing the implications, the expert added a .99 quantile at 0.02, splitting up the long upper tail. These were fitted and smoothed as described in Section 2.

The prior distribution for θ is shown in Figure 1.

Model 2 requires a prior on the asset correlation ρ. For this portfolio bucket, B2 recommends a value of approximately 0.20. We did not assess further details from an expert on this parameter. There appears to be little experience with correlation, relative to expert information available on default rates. There is agreement that the correlation is positive (as it has to be asymptotically if there are many assets). Consequently, we choose a Beta prior with mean equal to 0.20 for ρ. Since the B2 procedure is to fix ρ at that value, any weakening of this constraint is a generalization of the model. We choose a Beta(12.6, 50.4) distribution, with a standard deviation of 0.05. This prior is illustrated in Figure 2. Thus, the prior specifications on the parameters for which we have no expert information beyond that given in the B2 guidelines reflect the guidelines as means and little else. The joint prior for θ and ρ is obtained as the product, which is the maximum-entropy combination of the given marginals. Here, it does not seem to make sense to impose correlation structure in the absence of expert information.

As to τ, here we have little guidance. We take the prior to be uniform on [-1,1]. It might be argued that τ is more likely to be positive than negative, and this could certainly be done. Further, some guidance might be obtained from the literature on asset prices, though this usually considers less homogeneous portfolios. Here we

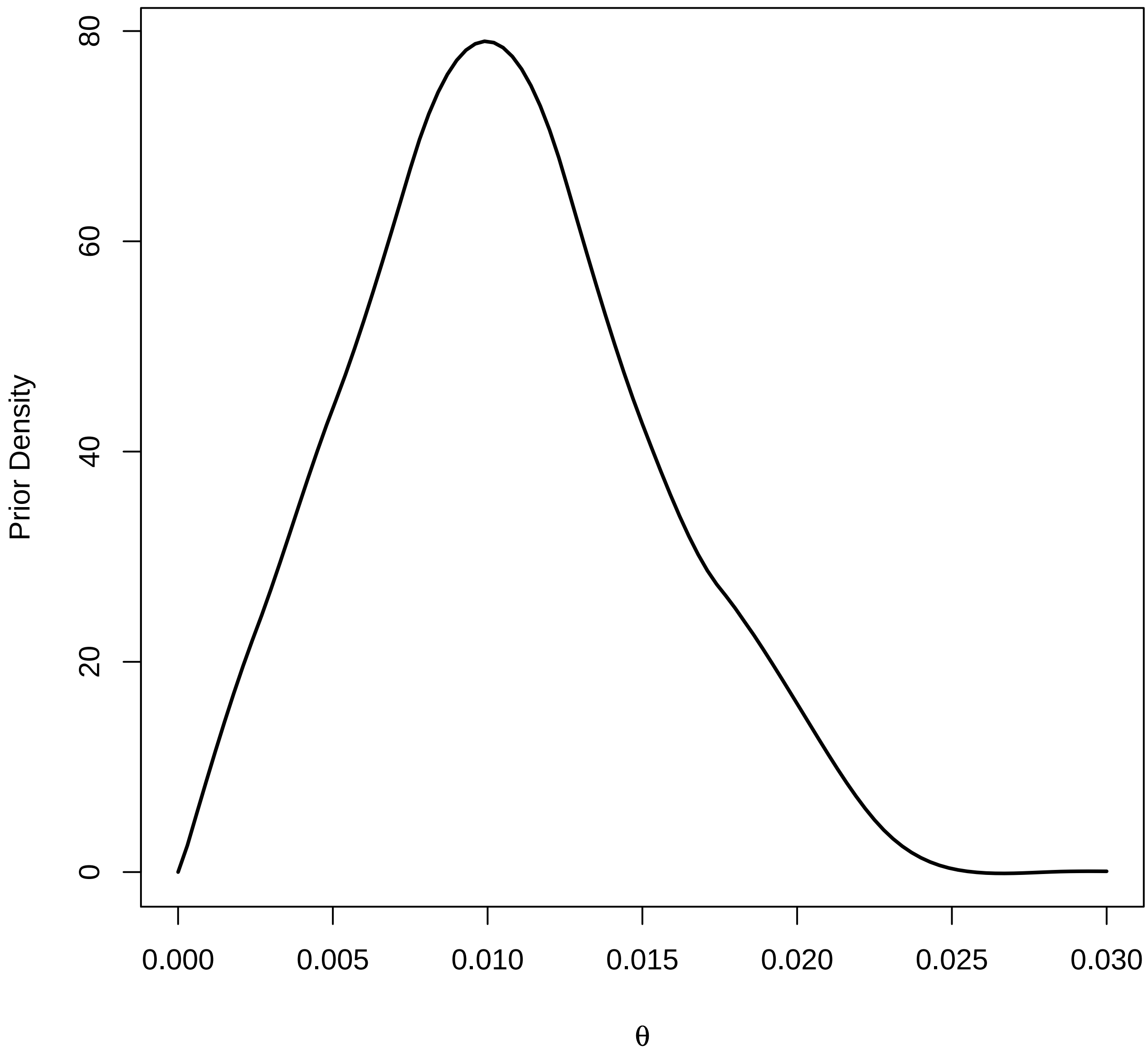

Figure 1: Prior on the long-run default probability θ

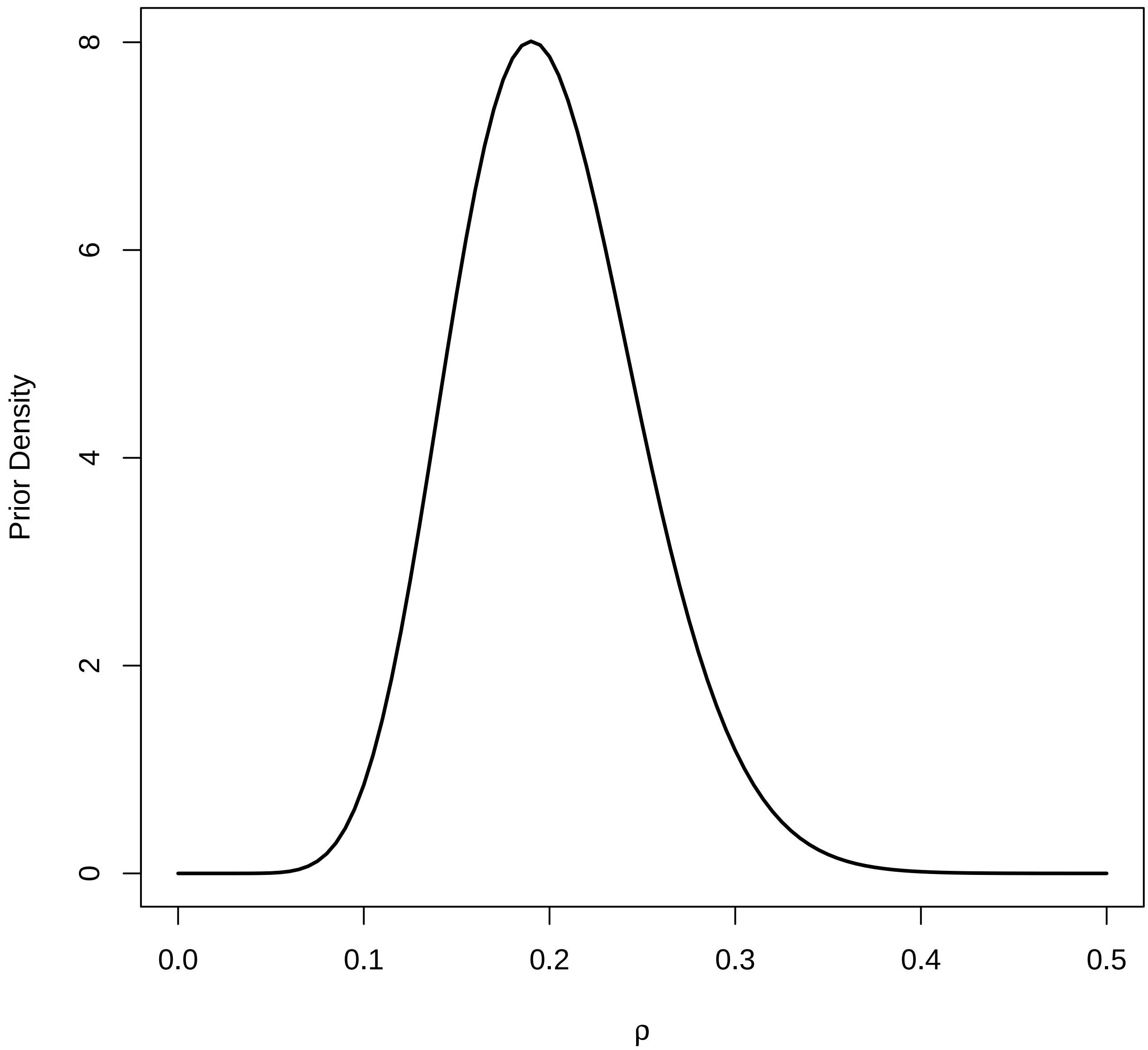

Figure 2: Prior on the asset correlation ρ

choose a specification that has the standard B2 model at its mean value, so that allowing for nonzero τ is a strict generalization of existing practice.

4.2 Inference

Writing the likelihood function generically as $p(R|\phi)$ with $\phi \in \{\theta, (\theta, \rho), (\theta, \rho, \tau)\}$ depending on whether we are referring to the likelihood function (5), (10), or (11), and the corresponding prior $p(\phi)$, inference is a straightforward application of Bayes rule. The joint distribution of the data R and the parameter ϕ is

$$p(R, \phi) = p(R|\phi)p(\phi)$$

from which we obtain the marginal (predictive) distribution of R,

$$p(R) = \int p(R, \phi)d\phi \tag{12}$$

and divide to obtain the conditional (posterior) distribution of the parameter ϕ :

$$p(\phi|R) = p(R|\phi)p(\phi)/p(R) \tag{13}$$

Given the distribution $p(\phi|R)$, we might ask for a summary statistic, a suitable estimator for plugging into the required capital formulas as envisioned by Basel Committee on Banking Supervision (2006). A natural value to use is the posterior expectation, $\overline{\phi} = E(\phi|R)$. The expectation is an optimal estimator under quadratic loss and is asymptotically an optimal estimator under bowl-shaped loss functions.

In many applications the distribution $p(\phi|R)$ can be difficult to calculate due to the potential difficulty of calculating $p(R)$ which requires an integration over a possibly high dimensional parameter. Here, the dimensions in models 1, 2, and 3 are 1,2, and 3. The first model can be reliably integrated by direct numerical integration, as can model 2 (requiring rather more time). Model 3 becomes very difficult and simulation methods are more efficient. Since many applications will require simulation and efficient simulation methods are available, and since these methods can replace direct numerical integration in the simpler models as well, we describe the simulation approach. Here we describe the Markov Chain Monte Carlo concept briefly and give details specific to our application. For MCMC details see Robert and Casella (2004).

Markov Chain Monte Carlo methods are a class of procedures for calculating posterior distributions, or more generally sampling from a distribution when the normalizing constant is unknown. We consider here a simple case, the Metropolis method. The idea is to construct a sampling method generating a sample of draws $\phi^0, \phi^1, ..., \phi^N$ from $p(\phi|R)$, when $p(\phi|R)$ is only known up to a constant. The key insight is to note that it is easy to construct a Markov Chain whose equilibrium (invariant, stationary) distribution is $p(\phi|R)$. Begin with a proposal distribution $q(\phi'|\phi)$ giving a new value of ϕ depending stochastically on the current value. Assume (for simplicity - this assumption is easily dropped) that $q(\phi'|\phi) = q(\phi|\phi')$. This distribution should be easy to sample from and in fact is often taken to be normal: $\phi' = \phi + \epsilon$ where ϵ is normally distributed with mean zero and covariance matrix diagonal with elements chosen shrewdly to make the algorithm work. Then, construct a sample in which ϕ^{n+1} is calculated from ϕ^n by first drawing ϕ' from $q(\phi'|\phi^n)$ then defining $\alpha(\phi', \phi^n) = p(R, \phi')/p(R, \phi^n) \wedge 1$ and defining $\phi^{n+1} = \phi'$ with probability $\alpha(\phi', \phi^n)$ or ϕ^n with probability $(1 - \alpha(\phi', \phi^n))$. Note that $p(R, \phi)$ is easy to calculate (the product of the likelihood and prior). Further, the ratio $p(R, \phi')/p(R, \phi^n) = p(\phi'|R)/p(\phi^n|R)$ since the normalizing constant $p(R)$ cancels.

The resulting sample $\phi^0, \phi^1, ..., \phi^N$ is a sample from a Markov Chain with equilibrium distribution $p(\phi|R)$. Eventually (in N) the chain will settle down and the sequence will approximate a sequence of draws from $p(\phi|R)$. Thus the posterior distribution can be plotted, moments calculated and expectations of functions of ϕ can be easily calculated by sample means. Calculation of standard errors should take into account that the data are not independent draws. Software to do these calculations with a user-supplied $p(R, \phi)$ exists. We use the the mcmc package (Geyer (2009)) in R (R Development Core Team (2009)). Guidance and associated warnings are available on the website noted in the package documentation. Generally, an acceptance ratio of about 25% is good (see Roberts, Gelman, and Gilks (1997)). The acceptance rate is tuned by adjusting the variances of ϵ. Long runs are better than short. There is essentially no way to prove that convergence has occurred, though nonconvergence is often obvious from time-series plots. For our illustrative application M samples from the joint posterior distribution were taken after a 5000-sample burnin. Scaling of the proposal distribution allowed an acceptance rate between 22 and 25 percent. Calculation of posterior distributions of the parameters and the predictive distributions of default rates are based on these samples.

We construct a segment of upper tier high-yield corporate bonds, from firms rated Ba by Moody's Investors Service, in the Moody's Default Risk Service^{TM} (DRS^{TM}) database (release date 1-8-2010). These are restricted to U.S. domiciled, non-financial and non-sovereign entities. Default rates were computed for annual cohorts of firms starting in January 1999 and running through January 2009. In total there are 2642 firm/years of data and 24 defaults, for an overall empirical rate of 0.00908. The data are shown in Figure 3.

The posterior distributions for this prior and data set were obtained and de-

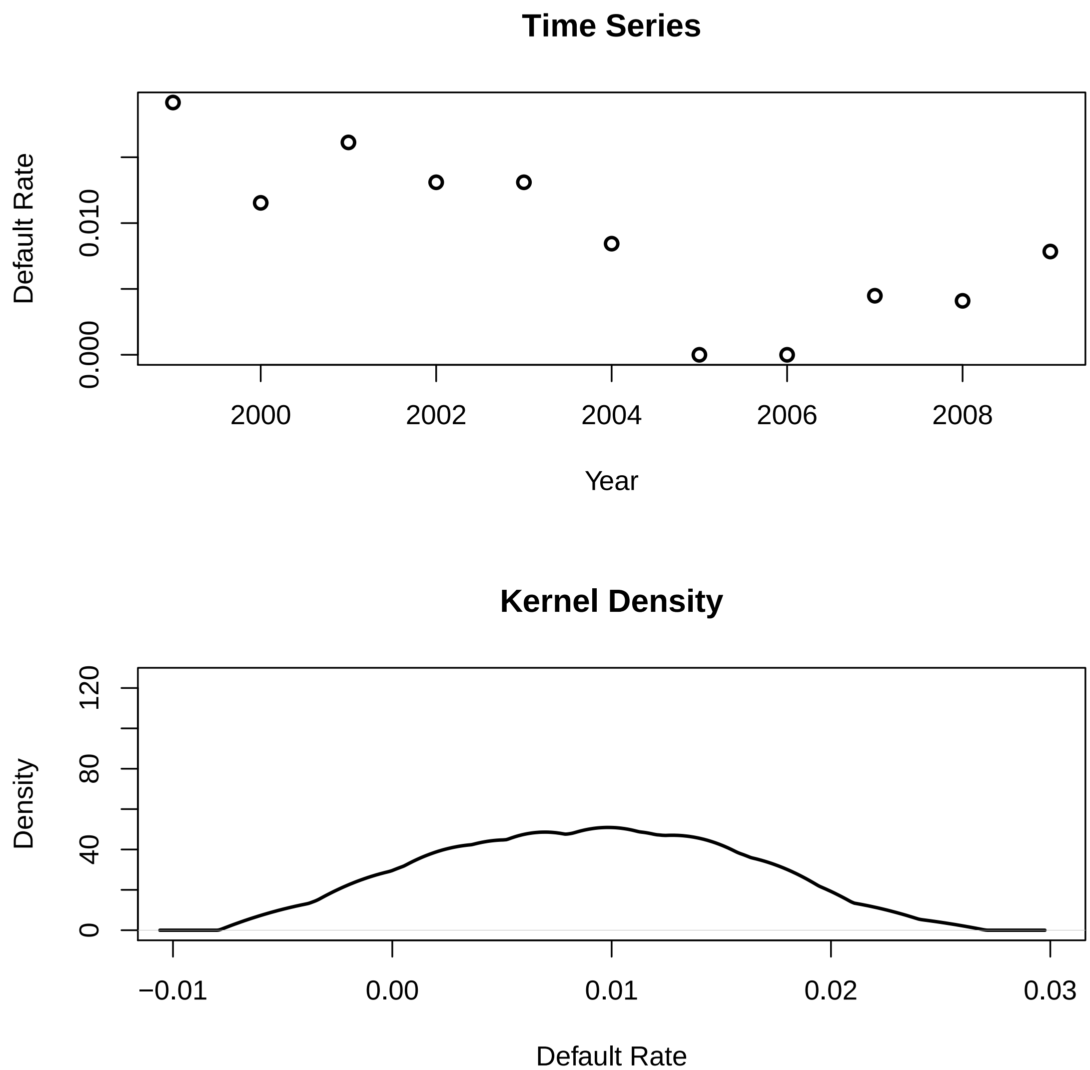

Figure 3: Default Rates: Mid-Portfolio application.

scribed in detail in Jacobs and Kiefer (2010), so we review these calculations briefly here and turn to the analysis of prediction, not addressed in the earlier paper. The analysis of the binomial model is straightforward using direct calculations involving numerical integration to calculate the predictive distribution and various moments (recall we are not in a conjugate-updating framework due to the flexible form of the prior representation).

The posterior distribution for the binomial model is shown in Figure 4.

This density has $E(\theta|R = r = 24) = 0.0098$ and $\sigma_\theta = 0.00174$.

Model II has asset value correlation within periods, allowing for heterogeneity in the default rate over time (but not correlated over time) and clumping of defaults. The marginal posterior distributions are shown in Figures 5 and 6.

This density has $E(\theta|R) = 0.0105$ and $\sigma_\theta = 0.00175$. The 95% credible interval for θ is (0.0073, 0.0140).

This density has $E(\rho|R) = 0.0770$ and $\sigma_\rho = 0.0194$. Note that the prior mean (0.2) is well outside the posterior 95% confidence interval for ρ. Analysis of the Vasicek distribution shows that the data information on ρ comes through the year-to-year variation in the default rates. At $\theta = 0.01$ and $\rho = 0.2$ the Vasicek distribution implies an intertemporal standard deviation in default rates of 0.015. With $\rho = 0.077$, the posterior mean, the implied standard deviation is 0.008. In our sample, the sample standard deviation is 0.0063. This is the aspect of the data which is moving the posterior to the left of the prior.

The marginal posterior distributions for Model III are shown in Figures 7-9.

This density has $E(\theta|R) = 0.0100$ and $\sigma_\theta = 0.00176$, $E(\rho|R) = 0.0812$ and $\sigma_\rho = 0.0185$, and $E(\tau|R) = 0.162$ and $\sigma_\tau = 0.0732$. Thus, there is some evidence that the intertemporal correlation parameter tau is positive but not large, broadly in line with asset pricing theory and evidence.

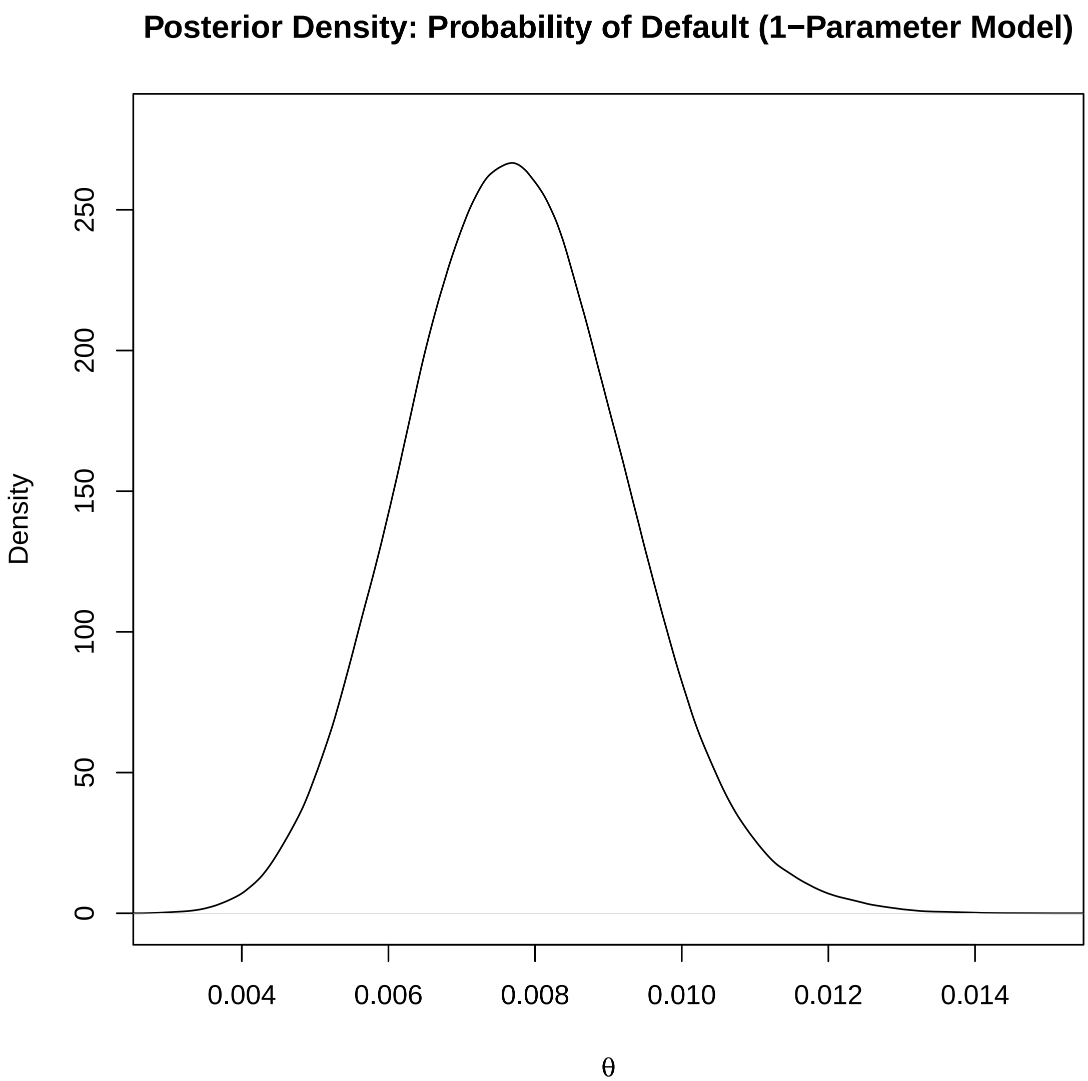

Figure 4: Model I, $p(\theta|R)$

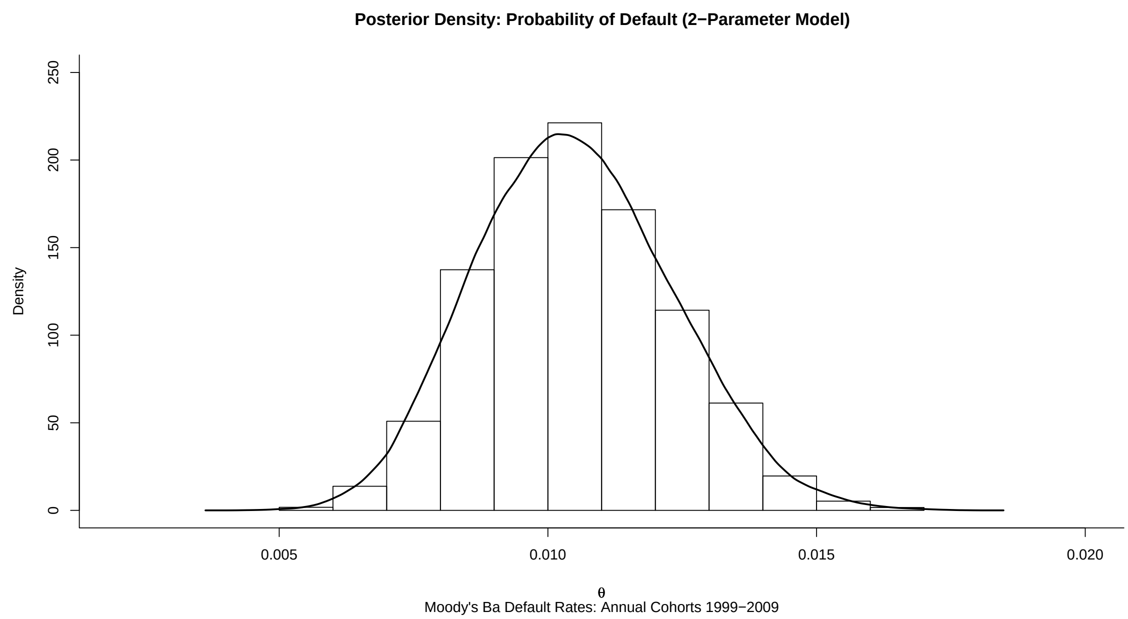

Figure 5: Model II, $p(\theta|R)$

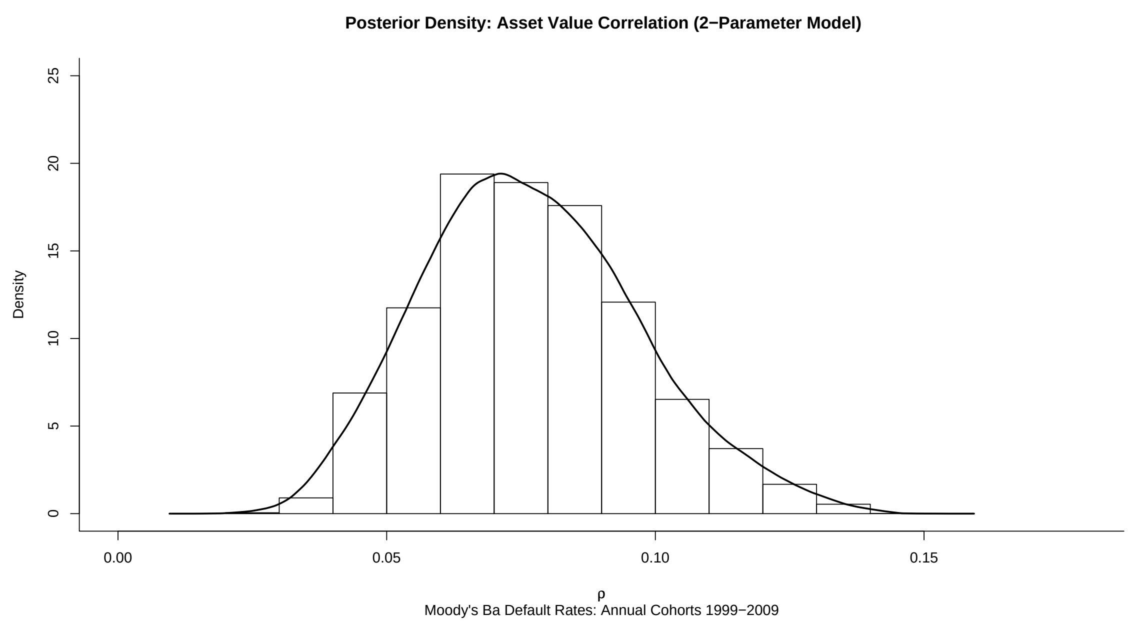

Figure 6: Model II, $p(\rho|R)$

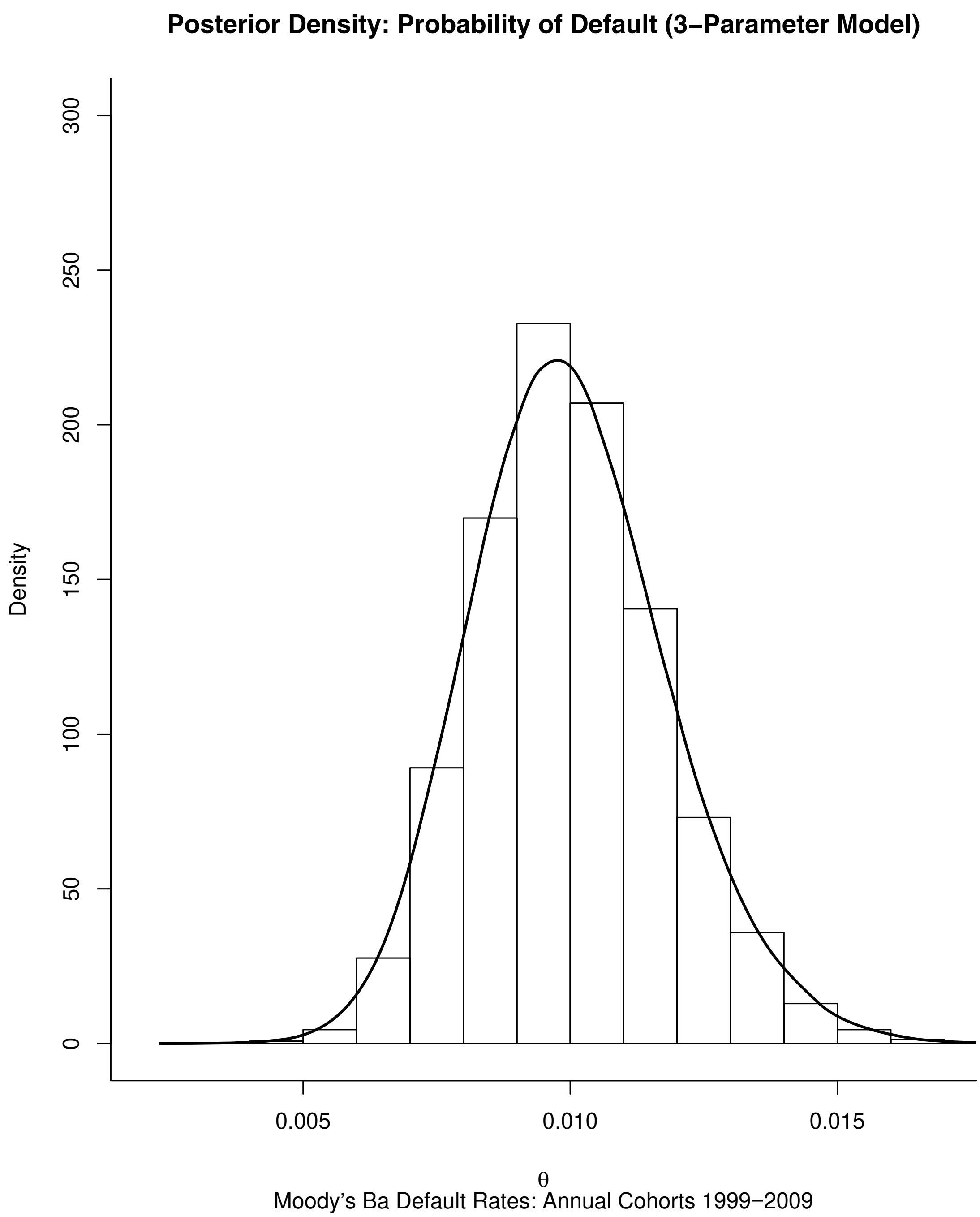

Figure 7: Model III $p(\theta|R)$

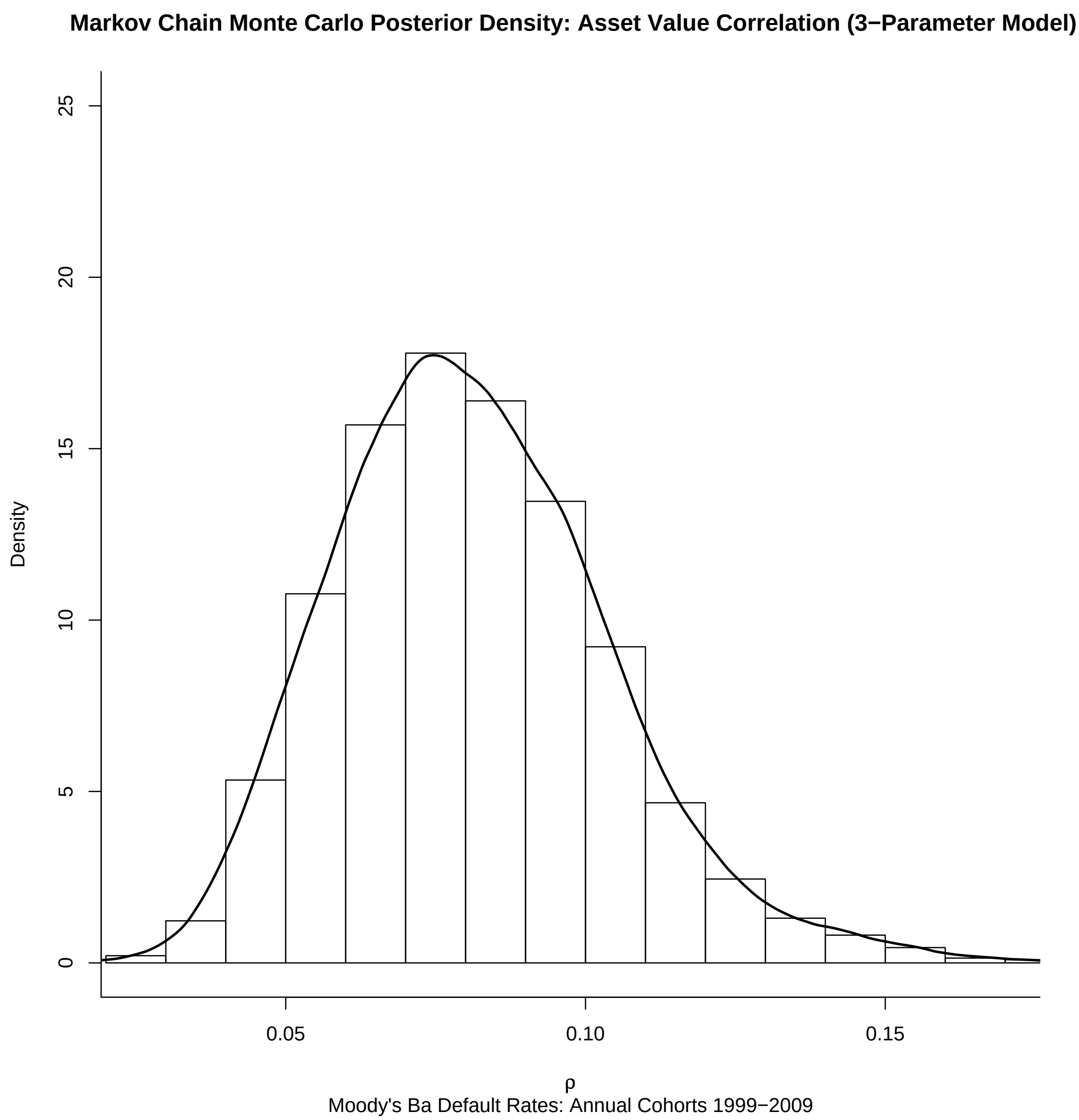

Figure 8: Model III, $p(\rho|R)$

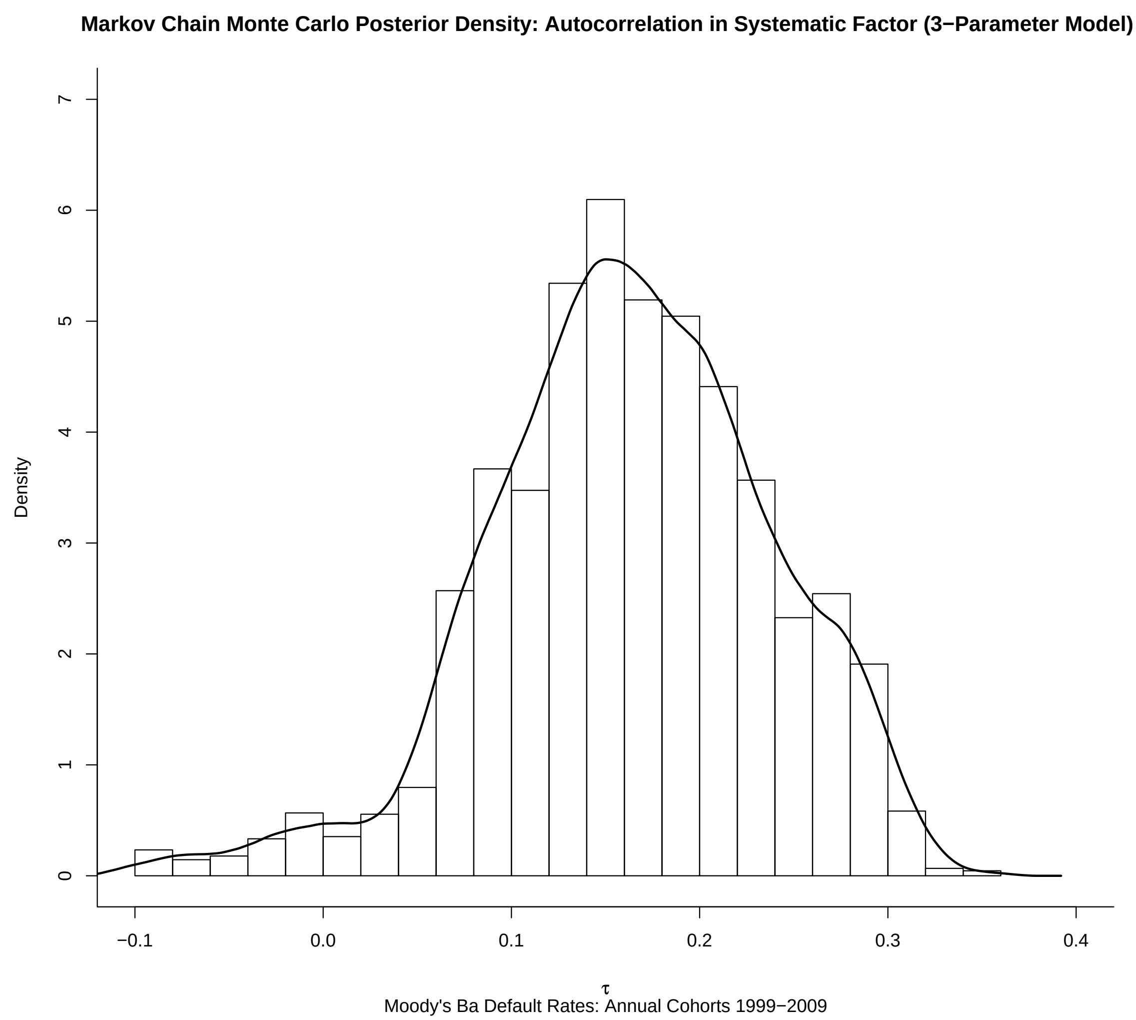

Figure 9: Model III, $p(\tau|R)$

4.3 Prediction

We obtain the predictive distributions for θ_{T+1} in 2 steps: first, we calculate the distribution for given values of the parameters - here there is variation due to the stochastic nature of the model in Models 2 and 3. Then, we marginalize with respect to the unknown parameters. In Model 1, the Binomial, the default rate is constant over time so, conditional on parameters, the $T+1$ forecast default rate θ^F_{T+1} is simply θ, the known long-run default rate.

In Model 2, the distribution of the default rate θ_{T+1} conditional on $\eta = (\theta, \tau)$ is from 8

$$\Pr(\theta_{T+1} < A) = \Phi[((1-\rho)^{1/2}\Phi^{-1}[A] - \Phi^{-1}[\theta])/\rho^{1/2}]$$

and the density $p(\theta_{T+1}|\theta, \rho)$ is obtained by differentiating.

In Model 3 the realized period T default rate is useful in predicting θ_{T+1} because of the dynamics of the systemic factor x. From 7 we can write

$$\begin{aligned} x_t &= (T^* - (1-\rho)^{1/2}\Phi^{-1}(\theta_t))\rho^{-1/2} \\ &= \tau(T^* - (1-\rho)^{1/2}\Phi^{-1}(\theta_{t-1}))\rho^{-1/2} + \epsilon_t \end{aligned} \tag{14}$$

Hence

$$\begin{aligned} &\Pr(\theta_t \leq A|\theta_{t-1}) \\ &= \Pr(\Phi[(T^* - \rho^{1/2}(\tau(T^* - (1-\rho)^{1/2}\Phi^{-1}(\theta_{t-1}))\rho^{-1/2} + \epsilon_t))/(1-\rho)^{1/2}] \leq A) \\ &= \Pr(\epsilon_t < \rho^{-1/2}(\Phi^{-1}(A) - T^* + \tau(T^* - (1-\rho)^{1/2}\Phi^{-1}(\theta_{t-1}))) \end{aligned} \tag{15}$$

using the fact that ϵ_t is symmetric around zero. This is just a standard normal integral and the density $p(\theta_{T+1}|\theta_T, \theta, \rho, \tau)$is obtained by differentiation.

Of course, the predictive distribution for θ_{T+1} from Model 1 is simply the posterior distribution of θ given in Figure 3. Turning to Model 2, the relevant density is

$$p(\theta_{T+1}|R) = \iint p(\theta_{T+1}|\theta, \rho)p(\theta, \rho|R)d\theta d\rho$$

where the definite integrals are over the supports of θ and ρ. This density is shown in Figure 10. It has $E(\theta_{T+1}|R)$ =0.010 and $\sigma_{\theta_{T+1}}$ =0.010. Thus, even accounting for parametric uncertainty, incorporating the variation predicted by the one-factor model increases the prediction standard error relative to the Binomial model by a factor of 8.

For Model 3 the conditional density (on lagged defaults) is

$$p(\theta_{T+1}|\theta_T, R) = \int\int\int p(\theta_{T+1}|\theta_T, \theta, \rho, \tau)p(\theta, \rho, \tau|R)d\theta d\rho d\tau$$

where the integrals are definite. This density for the two trial values of lagged θ, namely 0.004 and 0.015, are graphed in Figures 11 and 12. Summary statistics are $E(\theta_{T+1}|\theta_T = 0.004, \theta, \rho, R) = 0.006$ with $\sigma_{\theta_{T+1}} = 0.004$, and $E(\theta_{T+1}|\theta_T = 0.004, \theta, \rho, R) = 0.014$ with $\sigma_{\theta_{T+1}} = 0.010$.

5 The Low-Default Portfolio Bucket Example

5.1 The Prior

The minimum value for the default probability was 0.0001 (one basis point). The expert gave quantiles of the distribution of the default probability. A useful device here is to think about equiprobable events, leading naturally to assessment of the median value, and then conditionally equiprobable events, leading to the quartiles. Finer quantiles are a little more difficult, though risk managers are used to thinking about tail events. After some discussion, the expert reported 0, 0.25, 0.50, 0.75, 0.90,

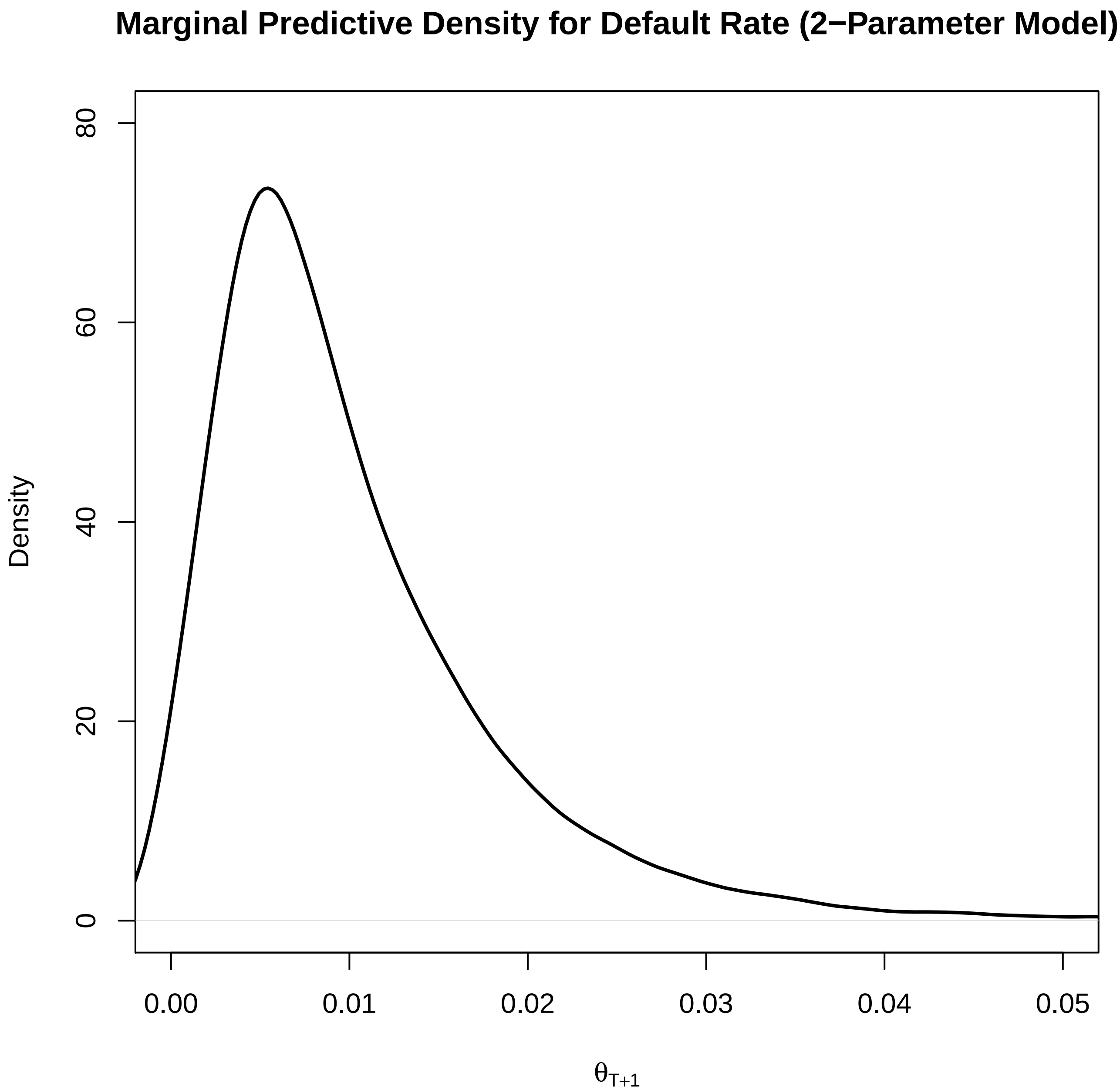

Figure 10: Predictive density $p(\theta_{T+1}|R)$ from Model 2.

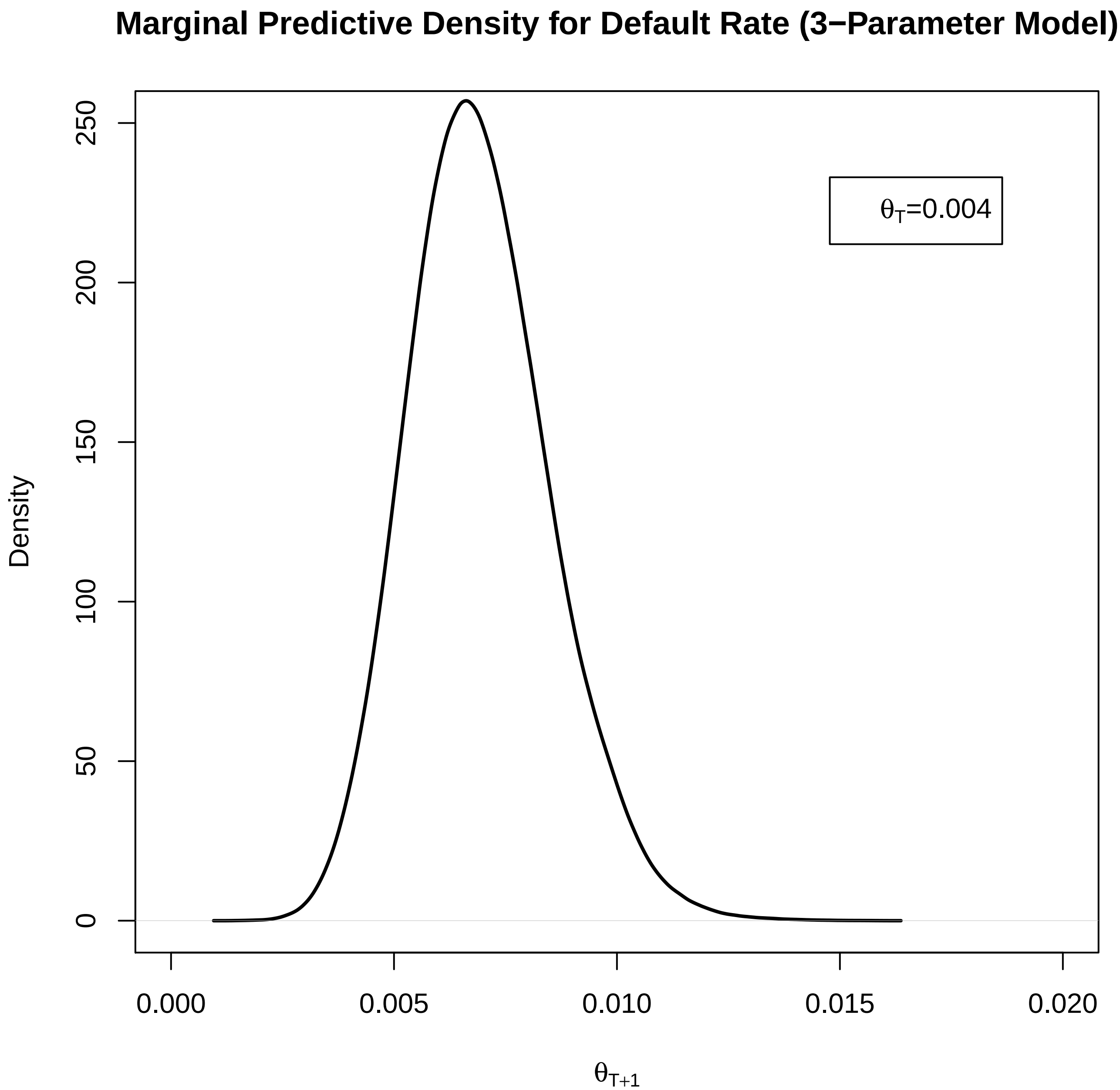

Figure 11: Predictive density $p(\theta_{T+1}|\theta_T = 0.004, R)$ from Model 3.

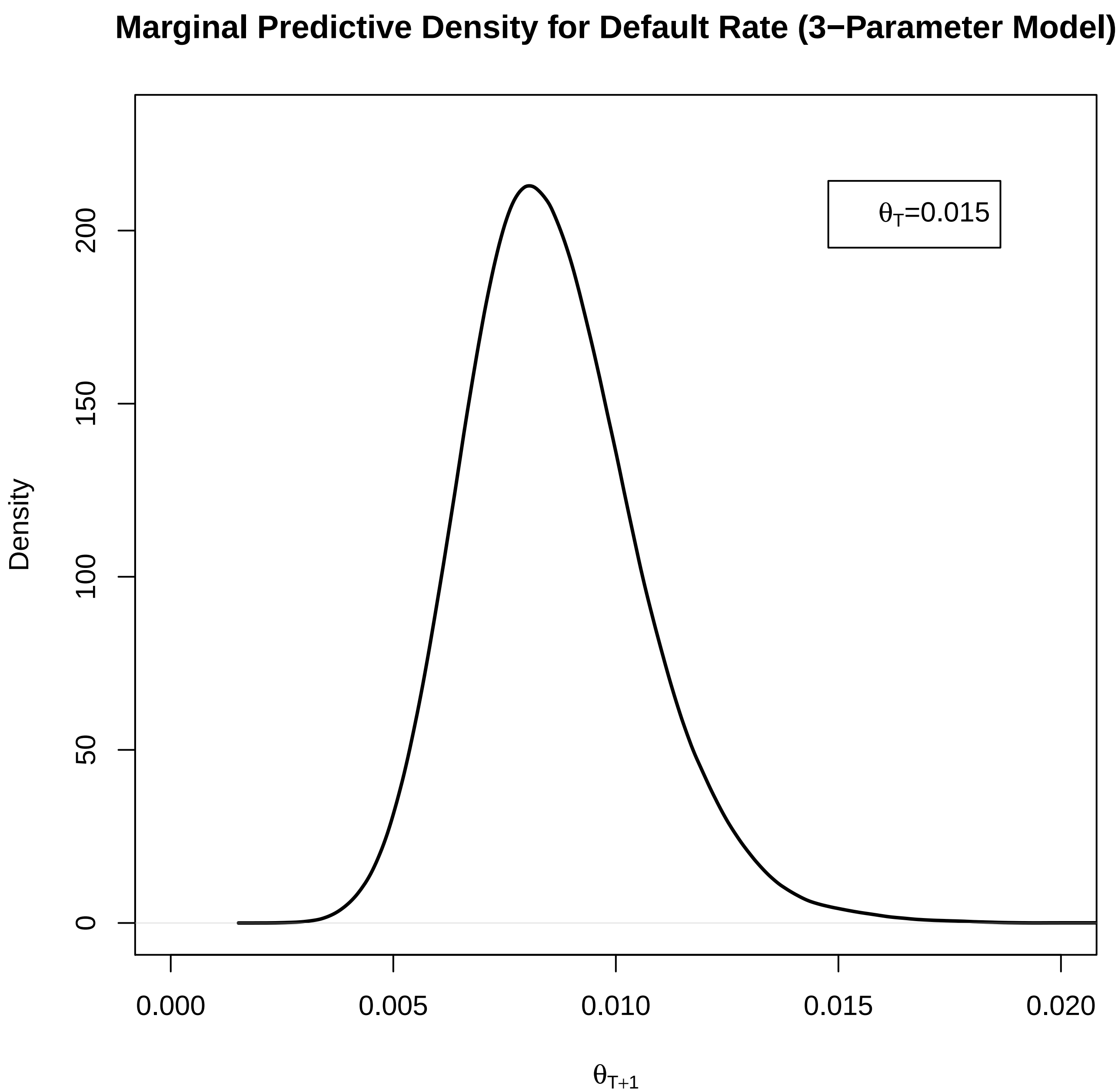

Figure 12: Predictive density $p(\theta_{T+1}|\theta_T = 0.015, R)$ from Model 3.

and 1.0 quantiles as 0.0001, 0.00225, 0.0033, 0,025, 0.035, and 0.05.. Our expert found it much easier to think in terms of quantiles than in terms of moments. As above, we fit and then smooth a maximum entropy representation.

The data for this section comprise a segment of investment grade corporate bonds, from firms rated between Baa and Aaa by Moody's Investors Service, in the Moody's Default Risk Service^{TM} (DRS^{TM}) database (release date 1-8-2010). As with the mid-portfolio segment, these are restricted to U.S. domiciled, non-financial and non-sovereign entities. Default rates were computed for annual cohorts of firms starting in January 1999 and running through January 2009. In total there are 8905 firm/years of data and 17 defaults, for an overall empirical rate of 0.0019.

The priors for ρ and τ, not developed from experts but in accord with the B2 prescriptions, are as in the mid-portfolio example above.

To be brief, we turn immediately to the predictive distributions, noting that in model 1 this is the posterior distribution (Figure 15), and in model 2 this is the marginal posterior distribution for θ (Figure 16).

In model 3 we condition on previous realizations: here we take 10 and 40 basis points (Figures 17 and 18).

Here the expectation and standard error of θ_{T+1} are 0.0020110 and 0.0017.

Here the expectation and standard error of θ_{T+1} are 0.0015 and 0.0015.

6 Conclusion

In summary, the picture on the default probability is pretty clear: it is around 0.01 in all models for the midportfolio example and around 0.0022 ror the low-default portfolio. The asset value correlation is around 0.08 for both samples. This is substantially less than the value specified in B2. This result is coming from the

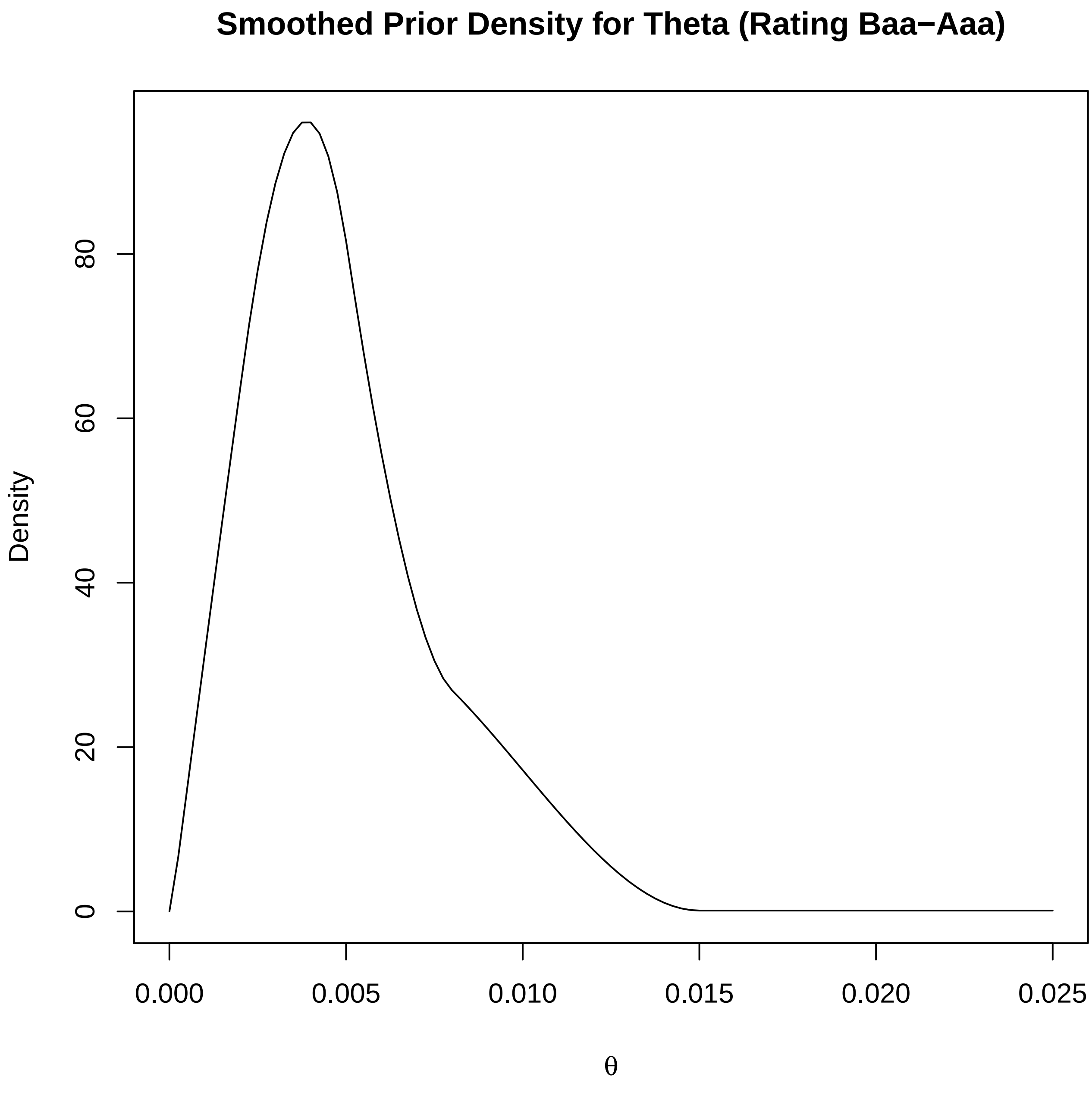

Figure 13: Prior for default rate θ from the low-default portfolio.

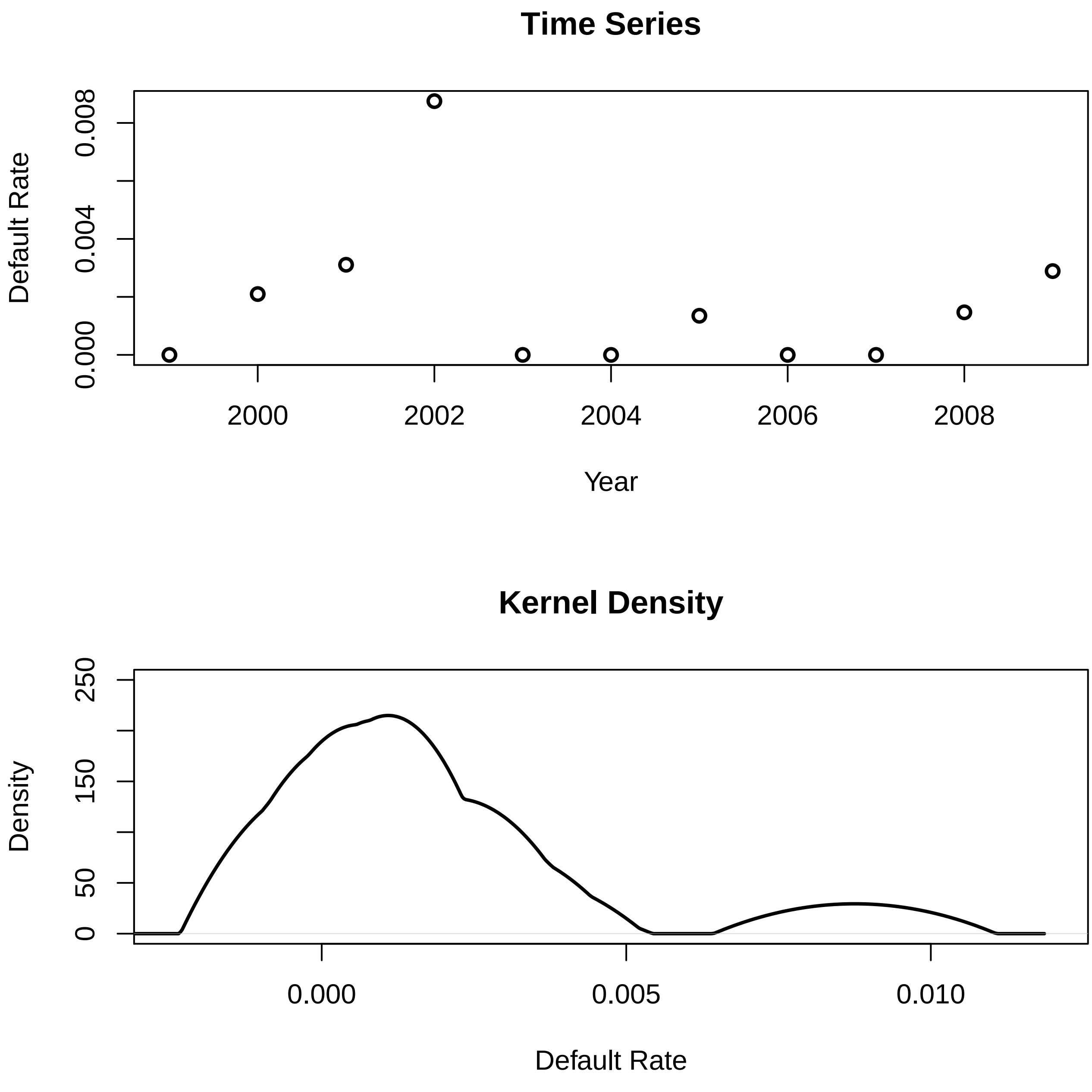

Figure 14: Data: The low-default application.

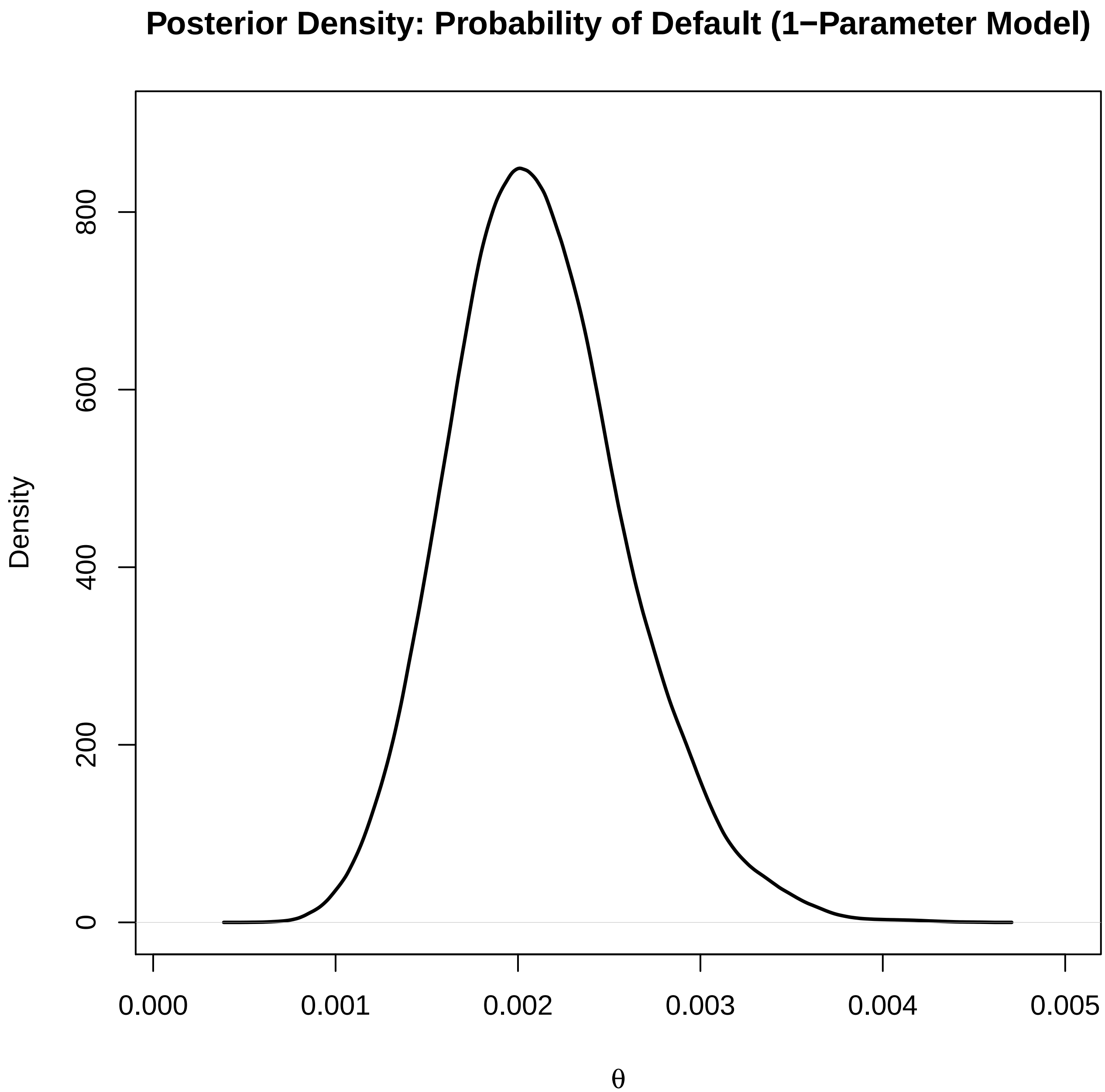

Figure 15: Predictive density $p(\theta|R)$ from low-default portfolio, model 1.

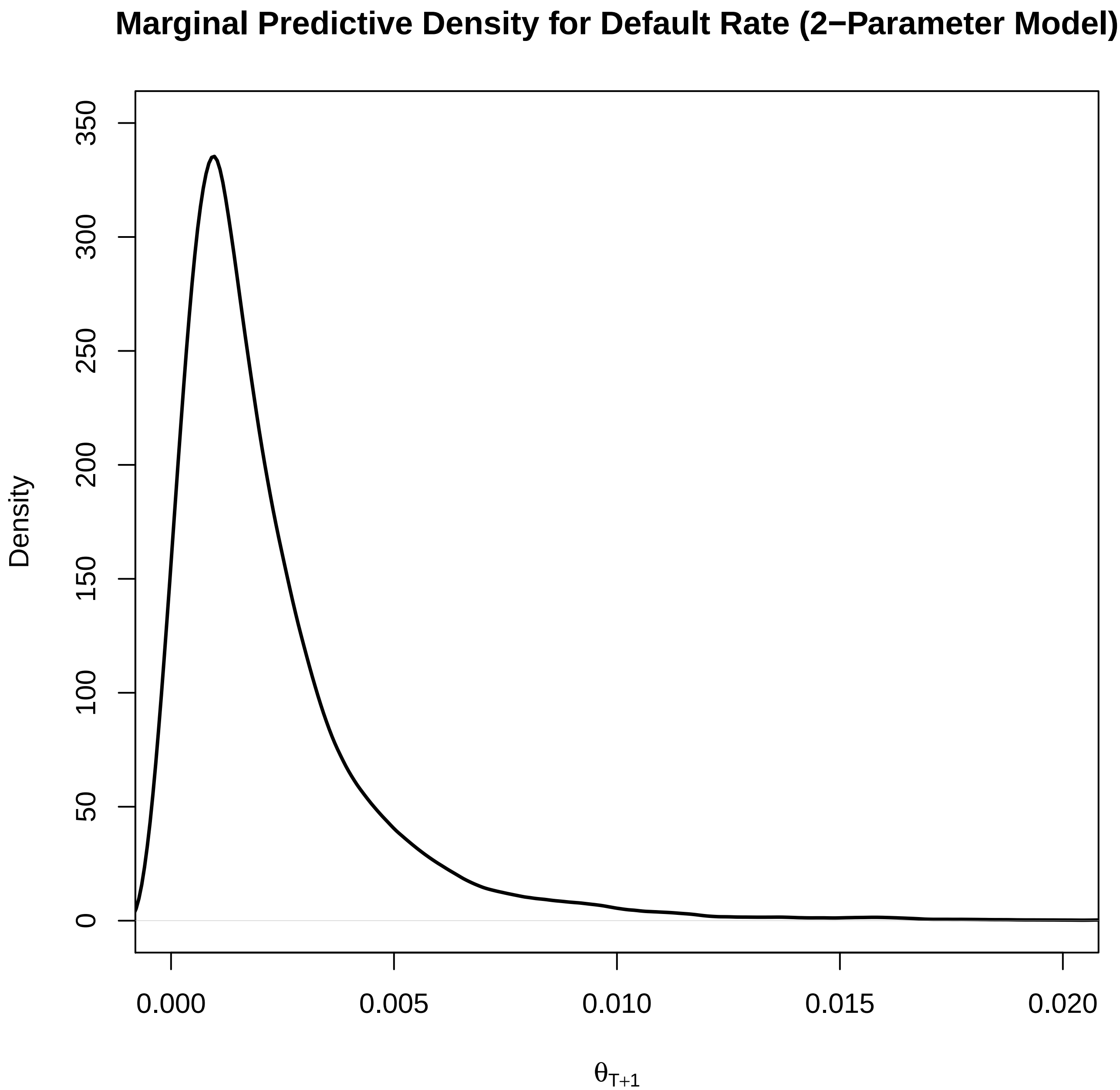

Figure 16: Predictive density $p(\theta|R)$ from low-default portfolio, model 2.

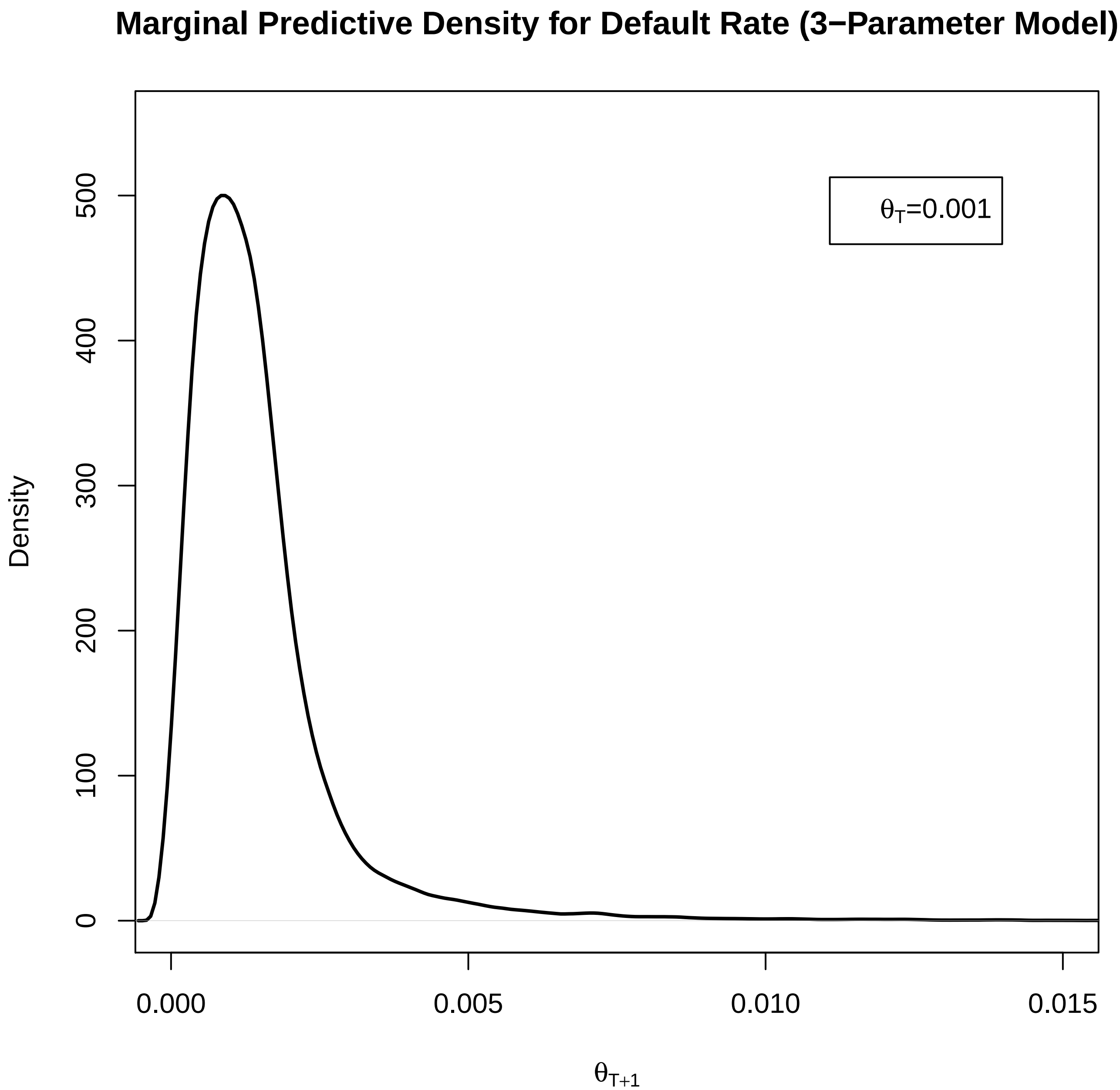

Figure 17: Predictive density $p(\theta_{T+1}|\theta_T = 0.001, R)$ from Model 3.

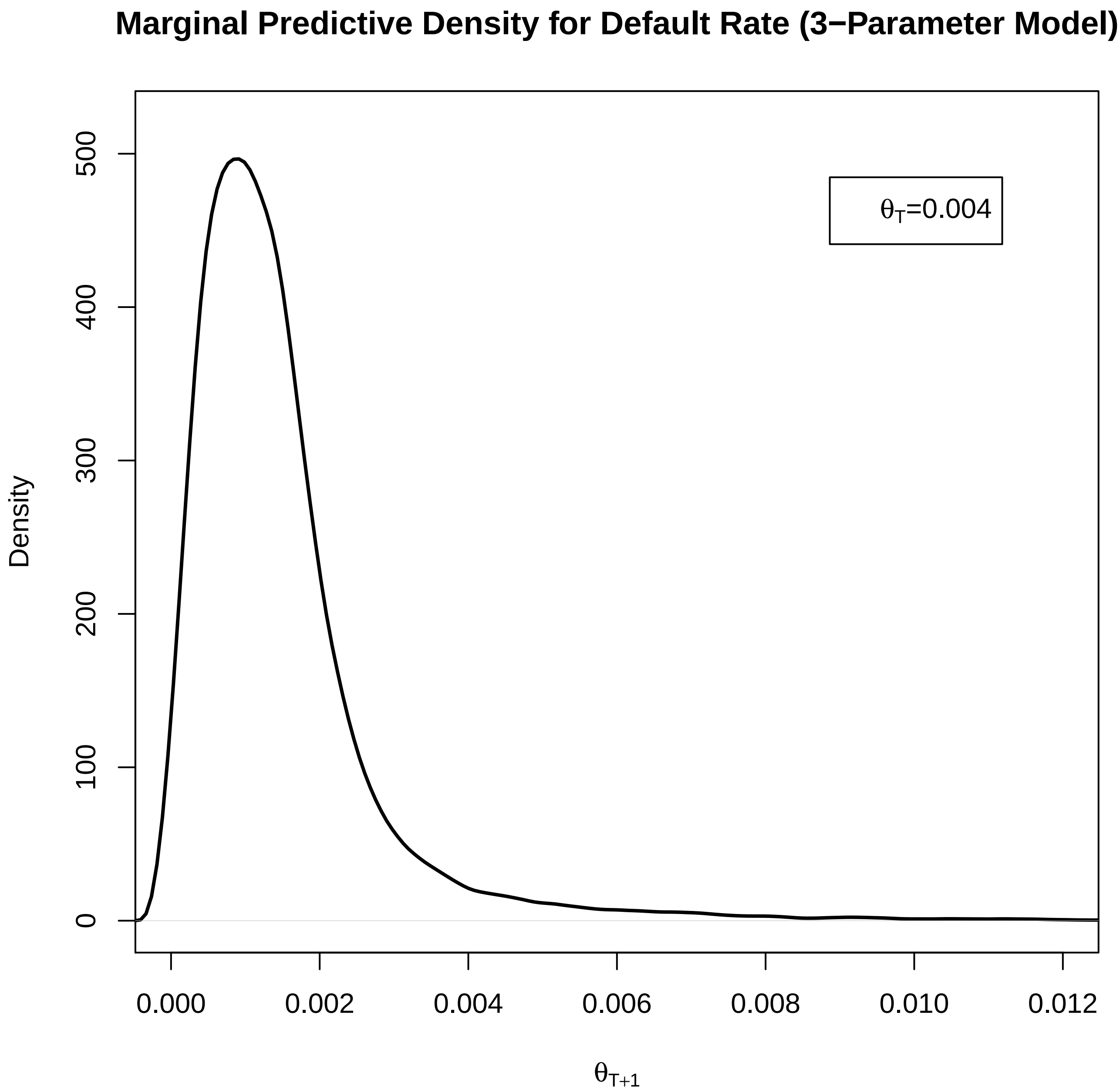

Figure 18: Predictive density $p(\theta_{T+1}|\theta_T = 0.004, R)$ from Model 3.

Estimation of Alternative Bayesian Credit Risk Models								
		E(θ|R)	σ_θ	E(ρ|R)	σ_ρ	E(τ|R)	σ_τ	Acceptance Rate
Binomial (1 Parameter	Ba	0.00977	0.00174					0.245
	Baa-Aaa	0.00212	0.00047					0.2705
Single Risk Factor Basel 2 (2	Ba	0.0105	0.00175	0.0770	0.0194			0.228
	Baa-Aaa	0.002241	0.00052	0.0777	0.01855			0.2688
Autocorrelated Signgle Risk	Ba	0.0100	0.00176	0.0812	0.0185	0.162	0.0732	0.239
	Baa-Aaa	0.0023	0.00050	0.0849	0.02286	-0.278	0.4898	0.2671

Figure 19: Summary statistics: all models.

variation in default rates over time: it is substantially less than that implied by the B2 specification. Default rates seem to be more predictable than contemplated by the asset correlations in B2. We have found this in other datasets as well. Note that this does not mean that the specified value is inappropriate for determining minimum regulatory capital; only that it is not an accurate representation of patterns in the data. The temporal correlation in the systematic factor is only present in model III. The evidence is sparse here (recall there are only 11 years of data and the prior information was as uninformative as possible) but it appears to be slightly and significantly positive in the mid-portfolio group and essentially undetermined in the low-default portfolio (recall this is identified from correlation in defaults over time, and there are few defaults in this group).

In this and related applications the risk modeler and forecaster faces the dual chore of modeling the data distribution with a specification of a statistical distribution and modeling expert information with a statistical distribution. We generate the posterior distributions for the parameters of a nested sequence of models and calculate predictive distributions for future default rates. We consider a nested sequence of models, starting with the binomial, adding asset value correlation as contemplated by Basel 2, and then adding temporal correlation in the systematic shock; not contemplated by B2 but appearing in advanced models. The more general models provide insight into the extent to which default rates over time are predictable, and to the extent to which risk calculations should look ahead over a number of years

References

Basel Committee on Banking Supervision (2006): "International Convergence of Capital Measurement and Capital Standards: A Revised Framework, Comprehensive Version," Bank for International Settlements.

——— (2009): "Proposed enhancements to the Basel II framework: Consultative Document," Discussion paper, Bank for International Settlements.

Bluhm, C., L. Overbeck, and C. K. J. Wagner (2003): *An Introduction to Credit Risk Modeling*. Chapman & Hall-CRC, Financial Mathematics Series, London.

Cooke, R. M. (1991): *Experts' in Uncertainty: Opinion and Subjective Probability in Science*. Oxford University Press.

Cover, T. M., and J. A. Thomas (1991): *Elements of Information Theory*. John Wiley & Sons.

Csiszar, I. (1975): "I-divergence Geometry of Probability Distributions and Minimization Problems," *The Annals of Probability*, 3, 146–158.

Das, S. R., D. Duffie, N. Kapadia, and L. Saita (2007): "Common Failings: How Corporate Defaults are Correlated," *Journal of Finance*, 62, 93–117.

Garthwaite, P. H., J. B. Kadane, and A. O'Hagan (2005): "Statistical Methods for Eliciting Probability Distributions," *Journal of the American Statistical Association*, 100, 780–700.

Geyer, C. J. (2009): *mcmc: Markov Chain Monte Carlo.* R package version 0.6.

Gordy, M. B. (2003): "A Risk-Factor Model Foundation for Ratings-Based Bank Capital Rules," *Journal of Financial Intermediation*, 12, 199–232.

Jacobs, M., and N. M. Kiefer (2010): "The Bayesian Approach to Default Risk: A Guide," in *Rethinking Risk Measurement and Reporting: Vol II*, ed. by K. Bocker, pp. 319–344. Risk Books.

Kadane, J. B., and L. J. Wolfson (1998): "Experiences in Elicitation," *The Statistician*, 47(1), 3–19.

Kiefer, N. M. (2009a): "Correlated Defaults, Temporal Correlation, Expert Information and Predictability of Default Rates," Discussion paper, Cornell University.

Kiefer, N. M. (2009b): "Default Estimation for Low Default Portfolios," *Journal of Empirical Finance*, 16, 164–173.

Kiefer, N. M. (2010): "Default Estimation and Expert Information," *Journal of Business and Economic Statistics*, 28(2), 320–328.

McCullagh, P., and J. Nelder (1989): *Generalized Linear Models, 2nd ed.* Chapman & Hall, London.

McNeil, A. J., and J. P. Wendin (2007): "Bayesian inference for generalized linear mixed models of portfolio credit risk," *Journal of Empirical Finance*, 14, 131149.

Nickell, P., W. Perraudin, and S. Varotto (2000): "Stability of Rating Transitions," *Journal of Banking and Finance*, 24, 203–227.

OCC (2000): "OCC Bulletin 2000-16, Subject: Risk Modeling, Description: Model Validation," Discussion paper, Office of the Comptroller of the Currency.

——— (2011): "OCC Bulletin 2011-12: Supervisory Guidance on Model Risk Management," Discussion paper, Office of the Comptroller of the Currency.

O'Hagan, A., C. E. Buck, A. Daneshkhah, J. R. Eiser, P. Garthwaite, D. J. Jenkinson, J. E. Oakley, and T. Rakow (2006): *Uncertain Judgements: Eliciting Experts' Probabilities*. Chichester: John Wiley & Sons.

R Development Core Team (2009): *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

Robert, C., and G. Casella (2004): *Monte Carlo Statistical Methods (2nd edition)*. New York: Springer-Verlag.

Roberts, G. O., A. Gelman, and W. R. Gilks (1997): "Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms," *The Annals of Applied Probability*, 7(1), 110–120.

Schuster, E. F. (1985): "Incorporating support constraints into nonparametric estimators of densities," *Communications in Statistical Theory and Methods*, 14, 1123–1136.