12

The Bayesian Approach to Default Risk: A Guide

Michael Jacobs Jr; Nicholas M. Kiefer

US Department of the Treasury; Cornell University

All competent statistical analyses involve a subjective or judgmental component. Sometimes the importance of this input is minimised in a quest for objectivity. Nevertheless, it is clear that specification of a model, definition of parameters or quantities of interest, specification of the parameter space and identification of relevant data all require judgment and are subject to criticism and require justification. Indeed, this justification is an important part of the validation procedure expected of financial institutions (OCC 2000). However, estimation of parameters after these judgments are made typically proceeds without regard for potential non-data information about the parameters, again in an attempt to appear completely objective. But subject-matter experts typically have information about parameter values, as well as about model specification, etc. For example, a default rate should lie between 0 and 1 (definition of the parameter space), but if we are considering a default rate for a particular portfolio bucket, we in fact have a better idea of the location of the rate. The Bayesian approach allows formal incorporation of this information, formal combination of the data and non-data information using the rules of probability. A simple example in the case of estimating default rates is sketched in Kiefer (2007).

Word added – OK?

The Bayesian approach is most powerful and useful when used to combine data and non-data information. There is also an advantage in that powerful computational techniques such as the Markov Chain Monte Carlo (MCMC) and related techniques are available. These are widely discussed in the economics literature and have been applied in the default estimation setting. These applications invariably specify a "prior" which is convenient and adds minimal information (there is no such thing as an uninformative prior), allowing

Changes to sentence OK?

computationally efficient data analysis. This approach, while valuable, misses the true power of the Bayesian approach: the coherent incorporation of expert information.

The difficulty in Bayesian analysis is the elicitation and representation of expert information in the form of a probability distribution. This requires thought and effort, rather than mere computational power, and is therefore not commonly done. Further, in "large" samples, data information will typically overwhelm non-dogmatic prior information, so the prior is irrelevant asymptotically, and economists often justify ignoring prior information on this basis. However, there are many settings in which expert information is extremely valuable: in particular, cases in which data may be scarce, costly or of questionable reliability. These issues come up in default estimation, where data may not be available in quantity for low-default assets or for new products, or where structural economic changes may raise doubts about the relevance of historical data.

We go through the steps in a Bayesian analysis of a default rate. Estimation of long-run default rates for groups of homogeneous assets 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. In response to the 2007–9 credit crisis, in a document for comment the Basel Committee has stressed the continuing importance of quantitative risk management (Basel Committee on Banking Supervision 2009). Our emphasis is on the incorporation of non-data information, so we focus on elicitation and representation of expert information and then on the Bayesian approach to inference in the context of a simple model of defaults. Uncertainty about the default probability should be modelled the same way as uncertainty about defaults: represented in a probability distribution. A future default either does or does not occur, given the definition. Since we do not know in advance whether default occurs or not, we 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 made shows that some risk assessment is occurring. This information should be organised and incorporated in the analysis in a sensible way, specifically represented in a probability distribution.

Changes to sentence OK?

Changes to sentence OK?

We discuss elicitation generally in the next section. Then we run through the steps of a formal Bayesian analysis in a particular example. Definition of a parameter of interest requires a model, so we next turn to specification of a sequence of simple models (each generating a likelihood function). The first two models are consistent with the asymptotic single-factor model underlying B2. The third adds temporal correlation in asset values, generalising B2 and is perhaps in line with the validation expectations of **?**, in that accounting for autocorrelation in the systematic factor speaks to how a model should be forward looking. The next section goes through the actual elicitation and representation for a particular expert. We then sketch the MCMC approach to calculating the posterior distribution (the distribution resulting from combining data and expert information coherently using the rules of probability) and illustrate all of the steps using Moody's data on corporate defaults. In the final section we give our conclusions.

Changes to this paragraph OK? Book style is for unnumbered sections.

Word added – OK?

There is no 'Basel2009a' in your bibliography – please give reference details.

To summarise: the steps in a Bayesian analysis are as follows.

1. Specify an economic model precisely defining the quantity of interest and generating a likelihood function for well-defined data (see page 351 onwards).
2. Identify a qualified expert and elicit information about the quantity of interest (see page 354).
3. Quantify this information in a probability distribution (see page 354).
4. Use the rules of probability to combine coherently the likelihood and the prior distribution, generating the posterior distribution of the quantity of interest (see page 358).
5. Analyse the posterior distribution using MCMC (see page 358).

ELICITATION OF EXPERT INFORMATION

A general definition of elicitation that we may offer in this context is a structured process or algorithm for transforming an expert's beliefs regarding an uncertain phenomenon into a probability distribution. In deploying Bayesian statistical technology in the domain of credit risk, elicitation arises as a method for specifying a prior distribution for one or more unknown parameters governing a model of credit risk (ie, a probability of default (PD)), where the expert may be an experienced statistician or a non-quantitatively oriented risk specialist (eg, a loan officer). In this setting, the prior distribution will be combined with a data likelihood through Bayes's Theorem, to derive the posterior distribution of the measure of risk (ie, the distribution of the default rate). While our focus is on formulating a probability distribution for uncertain quantities for the purpose of inference about a parameter, especially when there is no or very limited data, we note here that this is not the only context in which elicitation is important. This situation also arises in decision-making where uncertainty about "states of nature" needs to be expressed as a probability distribution in order to derive and maximise expected utility. Similarly, this methodology arises in the application of mechanistic models built in almost all areas of science and technology to describe, understand, and predict the behaviour of complex physical processes. In that application a model developer will typically propose sensible model parameters in order to obtain outputs in cases where in general there is uncertainty about the inputs' true values. As in our application, this highlights the importance of having a coherent approach to represent that uncertainty.

A useful way to frame the elicitation is to identify the model developer or econometrician as a facilitator, who helps the expert transform the "soft data" (ie, experience or opinion) into a form amenable to statistical inference, which is the process of crafting expert's knowledge into probabilistic form. Elicitation is a complex process that, if done well, calls for a facilitator to be skilled and multi-faceted, as the role of the facilitator is central to the process of elicitation. Therefore, not only is the facilitator an econometrician, but they should also have knowledge of the business of making loans and issues in managing credit risk, as well as being a strong communicator.

Changes to sentence OK?

We may be able to set criteria for the quality of an elicitation. In doing so, we might believe that a meaningful distinction exists between the quality of an expert's knowledge, on the one hand, and the accuracy with which that knowledge is translated into probabilistic form, on the other. Therefore, we say that an elicitation is done well under the condition that the distribution so derived is an accurate representation of the expert's knowledge, no matter what the quality of that knowledge. A good facilitator asking probing questions may also be able to determine whether the expert really does have valuable information. We do not pursue this line here.

We may conclude that accurate elicitation of expert knowledge is by no means a straightforward task. This remains the case even if all we wish to elicit is expert's beliefs regarding only a single of event or hypothesis, an example in credit risk being the proportion of defaulted obligors in a particular rating class (or segment) over a given horizon. Here we seek an assessment of probabilities, but it is entirely possible that the expert may be unfamiliar with the meaning of probabilities, or if he can think intuitively in these terms then it still may be hard for him to articulate precise statements about probabilities. Even in the case where the expert is comfortable with probabilities and their meaning, it is still challenging to accurately assess numerical probabilities with respect to a relatively rare event such as that of default for a set of obligors, especially if they are highly rated and there is a lack of historical performance data on a portfolio of similar credits. Changes to sentence OK?

Let us now consider the task of eliciting a distribution for a continuous parameter θ, the proportion of customers in a given rating class defaulting. How may we proceed with this? One direct approach, which is impractical, involves implicitly eliciting an infinite collection of probabilities regarding this uncertain proportion (itself a probability), which we may write in terms of the distribution function for all of the possible values of θ. Note the symmetry here, as we characterise the uncertainty regarding the unknown probability governing the distribution of the default rate PD itself in terms of probabilities. However, we realise rather early on in this process that it is clearly impossible to do this, as in practice an expert can make only a finite number (and usually a rather limited number) of statements of belief about θ. It is likely that the best that we could hope for is that such statements might take the form of a small set of either individual probabilities, or a few quantiles of the distribution of θ; or possibly this might involve other summaries of the distribution, such as modes. In the case of a joint distribution for a collection of random quantities, for example, default rates in conjunction with loss severities, the elicitation task is much more complex. Words added – OK?

Given the apparent formidable difficulties involved in the elicitation process, a reasonable observer may question if it is worth the effort to even attempt this. The answer to why this is a worthy endeavour lies in the use of elicitations as part of the business of Word added – OK?

decision-making. We often find that a sensible objective for elicitation is to measure salient features of the expert's opinion, so that exact details may not be of the highest relevance with respect to the decision to be reached. For example, in forming a prior for the distribution of the default rate, a general sense of where it is centred (5 basis points, 1% or 10%?), and degree to which the tail is elongated, may be enough to inform the data at hand. Note the similarity to the issue of specification of the likelihood function, where typically an infinity (for continuous data) of probabilities are specified as a function of a small number of parameters. The point is made strongly in the normal case, when the whole set of probabilities are specified as functions of a mean and variance. This can hardly be credible as an exact description of a real data distribution, but nevertheless its usefulness has been proven in countless applications. Similarly, we specify a prior distribution on the basis of a small number of elicited properties. Even for cases in which the decision is somewhat sensitive to the exact shape of the elicited distribution, it may not be the decision but another metric that is of paramount importance, for example, the regulatory capital impact or the expected utility of the supervisor, which in many cases may quite often be robust to details of the expert's opinion.

Changes to sentence OK?

New paragraph started here OK?

Changes to sentence OK?

Another use that supports the importance of elicitation is in statistical inference, particularly in the estimation of posterior distributions or predictive densities. This is a case in which elicitation promotes a careful consideration on the part of both the expert and the facilitator regarding the meaning of the parameters being elicited. This process results in two beneficial effects. First, it results in an analysis that is closer to the application, through requiring attention to the subject of the modelling exercise; in our application, this ensures that the focus is upon a set of plausible observed default rates, over a set horizon, with respect to obligors of a particular credit quality. Second, this discipline is useful in giving rise to a posterior distribution that, when finally calculated, is a meaningful object. By way of illustration, this process produces not only a PD estimate that can be used in a compliance exercise, but a complete predictive distribution of the default rate that is potentially useful in other risk management contexts, such as credit decision-making, account management or portfolio stress testing.

Changes to sentence OK?

A natural interpretation of elicitation is to conceive of it as part of the process of statistical modelling. When statisticians write down a likelihood function for an applied problem, this is nothing more than an informed (we hope) opinion regarding a data-generation process, which is conditional on a parameter set. In hierarchical frameworks (examples being random-effects models or models incorporating latent variables), we have distributions on a sub-set of parameters that are conditional on another set of parameters. Therefore, what we term "elicitation" in this context can be interpreted as nothing more than the final step in such a hierarchy: the statement of the form of the probability distribution of the highest-level parameters. This highlights that we should not lose sight of the fact that all of the usual principles of statistical modelling also apply to elicitation. Changes to punctuation here OK?

A stylised representation of the elicitation process consists of four separate stages. First, in the set-up stage, we prepare for the elicitation by identifying the expert, training the expert and identifying what aspects of the problem to elicit. The second step, clearly the heart of the process, is to elicit specific summaries of the experts' distributions for those aspects. This is followed by the fitting of a probability distribution to those summaries elicited in the second step. Note that in practice there may be overlap between this and the previous phase of the process, in the sense that the choice of what data to elicit often follows from the choice of distributional form that the facilitator prefers to fit. For example, if we prefer a simple parametric distribution such as a beta to describe the prior of the PD, then a few quantiles may suffice, whereas the more data intensive choice of a non-parametric kernel density may require other pieces of information. 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 an iterative 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.

Thus far, we have framed the process of conducting an elicitation as that of formulating in probabilistic terms the beliefs regarding uncertainty from an expert, which we argue is the appropriate way to think about credit risk. However, in this context, who is the expert? There are two aspects here: the qualification of the expert and the basis of their information. For the first, we look at education and experience, particularly at experience in related risk-management situations. For the second, we evaluate the quality of the arguments: would they be convincing to other experts? Are they based on reasoning from particular similar portfolios or configurations of economic conditions? In practice, the choice of expert or experts must be justified. In our context, the experts are not that difficult to identify: they are the individuals who are making risk-management decisions for the relevant portfolio in a successful financial institution.

In summary, we outline suggested criteria for the conduct of elicitations, in the context of formulating beliefs regarding parameters governing default risk. While some of these aspects may be ignored in an informal elicitation, they become considerations of

Word added – OK?

<u>the</u> utmost importance wherever substantive decisions or inferences may depend on the expert's knowledge, such as deriving a distribution of the default rate for either risk management or regulatory capital purposes. First, we should keep in mind that the objective is elicitation of a PD rate distribution, which represents the credit expert's current knowledge on the risk inherent in a portfolio, and it is very useful in this regard to have a summary of what that knowledge is based on (eg, state of the credit cycle, industry condition of the obligors or average features of the portfolio that are drivers of default risk). Second, we must be wary of any financial or personal interest that the credit expert may possess, and any inferences or decisions that will depend on the expert's distribution so elicited should be declared up front (eg, if the credit executive's bonus is a function of the regulatory capital charge on their portfolio). Next, it is also paramount that training should be offered in order to familiarise the expert with the interpretation of probability, as well as whatever other concepts and properties of probability will be required in the elicitation. It may be helpful in this regard to perform a "dry run" through an elicitation exercise with a view toward providing practice in the protocol that the facilitator proposes to use. Finally, the elicitation should be well documented. Ideally, this should set out all the questions asked by the facilitator, the expert's responses to those, and the process by which a probability distribution was fitted to those responses (eg, details of any moment matching or smoothing

performed, such as well-commented computer code). These documentation requirements for expert elicitation fit well in the the supervisory expectations for the documentation of developmental evidence.

STATISTICAL MODELS FOR DEFAULTS

Before elicitation can proceed, the quantities of interest need to be defined precisely. This requires a model. 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 $\theta \in [0, 1]$. The Basel requirements demand an annual default probability, estimated over a sample long enough to cover a full cycle of economic conditions. Thus, the probability should be marginal with respect to external conditions. Perhaps this marginalisation can be achieved within the binomial specification by averaging over the sample period; thus, many discussions of the inference issue have focused on the binomial model and the associated frequency estimator. Suppose the value of the ith asset in time t is Change OK?

$$v_{it} = \varepsilon_{it}$$

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 $p(d_i \mid \theta) = \theta^{d_i}(1 - \theta)^{1-d_i}$, where $\theta = \Phi(T^*)$, our binomial parameter. Let $D = \{d_i,\ i = 1, \dots, n\}$, let $n \in I^+$ denote the whole data set and let Changes to sentence OK?

$$r = r(D) = \sum_i d_i$$

denote the count of defaults. Then the joint distribution of the data is

$$\begin{aligned} p(D \mid \theta) &= \prod_{i=1}^{n} \theta^{d_i}(1 - \theta)^{1-d_i} \\ &= \theta^r (1 - \theta)^{n-r} \end{aligned} \tag{12.1}$$

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

Word added – OK?

$$p(r \mid \theta) = \binom{n}{r} \theta^r (1-\theta)^{n-r} \tag{12.2}$$

Change for consistency with other chapters here – OK?

a $\text{binom}(n, \theta)$ distribution. This is model I. Model I underlies what the ratings agencies assume in their PD estimation.

The Basel II guidance 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 (Das *et al* 2007; Nickell *et al* 2000). 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 modelled as

$$v_{it} = \rho^{1/2} x_t + (1-\rho)^{1/2} \varepsilon_{it} \tag{12.3}$$

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 we are interested in 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

Fractions stacked here and below – OK?

$$\theta_t = \Phi\left[\frac{T^* - \rho^{1/2} x_t}{(1-\rho)^{1/2}}\right] \tag{12.4}$$

Thus, for $\rho \neq 0$ there is random variation in the default probability over time. The distribution function for $A \in [0, 1]$ is given by

$$\begin{aligned} \Pr(\theta_t \leqslant A) &= \Pr\left(\Phi\left[\frac{T^* - \rho^{1/2} x_t}{(1-\rho)^{1/2}}\right] \leqslant A\right) \\ &= \Phi\left[\frac{(1-\rho)^{1/2} \Phi^{-1}[A] - \Phi^{-1}[\theta]}{\rho^{1/2}}\right] \end{aligned} \tag{12.5}$$

using the standard normal distribution of x_t and $\theta = \Phi(T^*)$. Differentiating gives the density $p(\theta_t \mid \theta, \rho)$. This is the Vasicek distribution.[1] 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 Equation 12.2)

Citation moved to endnote – OK? This reference does not appear in your bibliography – please give further details.

$$p(r_t \mid \theta_t) = \binom{n_t}{r_t} \theta_t^{r_t} (1 - \theta_t)^{n_t - r_t} \tag{12.6}$$

from which we obtain the distribution conditional on the underlying parameters

$$p(r_t \mid \theta, \rho) = \int p(r_t \mid \theta_t) p(\theta_t \mid \theta, \rho) \, \mathrm{d}\theta_t$$

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

$$p(R \mid \theta, \rho) = \prod_{t=1}^{T} p(r_t \mid \theta, \rho) \tag{12.7}$$

where we condition on $(n_1, \dots, n_T)$, ie, they are considered to be known. Regarded as a function of (θ, ρ) for fixed R, Equation 12.7 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. Specifically, let the systematic risk factor x_t follow an AR(1) process

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

with η_t independent and identically distributed (iid) standard normal and $\tau \in [-1, 1]$. Now the formula for θ_t (Equation 12.4) still holds but the likelihood calculation is different and cannot be broken up into the period-by-period calculation (cf Equation 12.7). Using Equation 12.6 we write

Changes to sentence OK?

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

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

$$\begin{aligned} &p(R \mid \theta, \rho, \tau) \\ &\quad = \int \cdots \int \prod_{t=1}^{T} p(r_t \mid \theta_t(x_t, \theta, \rho)) p(x_1, \dots, x_T \mid \tau) \, \mathrm{d}x_1 \cdots \mathrm{d}x_T \end{aligned} \tag{12.8}$$

where $p(x_1, \dots, x_T \mid \tau)$ 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, Equation 12.8 is the likelihood function. This is Model III.

Model I is a very simple example of a generalised linear model (GLM) (McCullagh and Nelder 1989).[2] Models II and III are in the form of the general linear mixed model (GLMM), a parametric mixture generalisation of the popular GLM class. These models were analysed using MCMC in the default application by McNeil and Wendin (2007) using convenience priors and focusing on default rate estimation, and by Kiefer (2009) using an elicited prior and focusing on predictability of default rates.

ELICITATION: EXAMPLE

Change OK?

We 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 Standard & Poor's BB–BBB. The elicitation method included a specification of the problem and some specific questions over email, followed by a discussion. Elicitation of prior distributions is an area that has attracted attention. General discussions of the elicitation of prior distributions are given in Part II of this volume and also by Garthwaite *et al* (2005), O'Hagan *et al* (2006) and Kadane and Wolfson (1998). Our expert is an experienced industry (banking) professional with responsibilities in risk management and other aspects of business analytics. They have seen many portfolios of this type in different institutions. The elicitation took place in 2006. 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. The mean value was 0.01. 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 they 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

Change in page range OK?

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 0.25 quartile, was 0.0075. The latter, the 0.75 quartile, was assessed at 0.0125. 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 0.99 quantile at 0.02, splitting up the long upper tail.

At this point a choice must be made on the representation of the elicited information. Of course, without further assumptions, we do not have enough information to specify a probability distribution. In principle that would require an infinity of elicitations. However, choosing a parametric form for a statistical distribution allows determination of the parameters on the basis of the assessed information (assuming standard identification properties; we cannot assess a median alone and uniquely determine a $k > 1$ parameter distribution). This is the most common approach in practice and parallels the usual practice in specifying the data distribution: a parametric form based (we hope) on an economic model, allowing an infinity (or large number in the discrete case) of probabilities to be determined by finitely many parameters. This approach is illustrated in Kiefer (2010), where the elicited information was used to fit a truncated Beta distribution. The disadvantage of this approach is that there is rarely good guidance beyond convenience on the choice of functional form. Thus, this choice can insert information not elicited from the expert nor really intended by the analyst. Based on experience, we prefer a non-parametric approach (really, less parametric): the maximum entropy approach (**?**).

'Kiefer2010' does not appear in your bibliography. Please give further details.

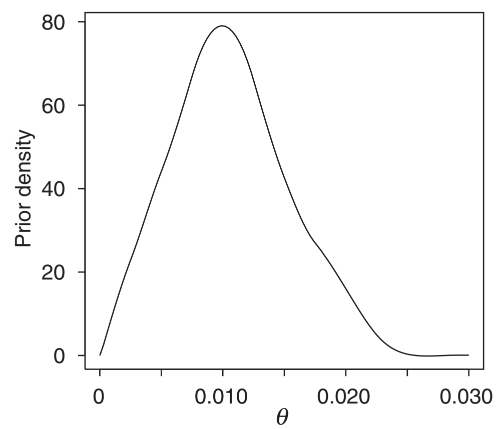
The maximum entropy approach provides a method to specify the distribution that meets the expert specifications and imposes as little additional information as possible. Thus, we maximise the entropy (minimise the information) in the distribution subject to the constraints indexed by k given by the assessments. Entropy is

Abbreviation spelled out in full to avoid confusion with 'moment estimator' elsewhere in the book. OK?

$$H(p) = -\int \log(p(x))\,\mathrm{d}P$$

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 **?**. The general framework is to

Figure 12.1 Prior on the long-run default probability θ



solve for the distribution p

$$\max_p \left\{ - \int p \ln(p(x))\, \mathrm{d}x \right\} \tag{12.9}$$

Change to equation layout OK?

such that

$$\int p(x) c_k(x)\, \mathrm{d}x = 0 \quad \text{for } k = 1, \dots, K$$

and

$$\int p(x)\, \mathrm{d}x = 1$$

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 $c(x) = I(x < \text{median}) - 0.5$. To solve this maximisation 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

Notation OK or '$\mathbf{1}_{\{x<\text{median}\}} \cdots$' and similar in Equation 12.10 for consistency with other chapters?

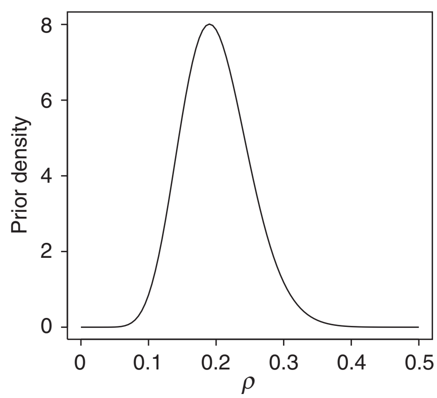
$$p_{\mathrm{ME}}(\theta) = \kappa \exp\left\{ \sum_k \lambda_k (I(\theta < q_k) - \alpha_k) \right\} \tag{12.10}$$

The multipliers are chosen so that the constraints are satisfied.[3]

Citations moved to endnote – OK? References cited here do not appear in the bibliography – please give publication details.

This gives a piecewise uniform distribution for θ. It can be argued that the discontinuities in $p_{\mathrm{ME}}(\theta)$ are unlikely to reflect characteristics of expert information, and indeed this was the view of the

Figure 12.2 Prior on the asset correlation ρ



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 $p_{\mathrm{S}}(\theta)$ the smoothed distribution with bandwidth h we have

I have assumed that suffix 'S' is a label for 'smoothed' and set it upright according to book style. Please mark any that denote variable 'S' and should be italic. Fraction stacked in displayed equation – OK?

$$p_{\mathrm{S}}(\theta) = \int_{-1}^{1} K(u) p_{\mathrm{ME}}\left(\theta + \frac{u}{h}\right) \mathrm{d}u \tag{12.11}$$

with $K(u) = \frac{3}{4}(1 - u^2)$ for $-1 < u < 1$. Since the density $p_{\mathrm{ME}}(\theta)$ is defined on bounded support, there is an end-point or boundary "problem" in calculating the kernel-smoothed density estimator. Specifically, $p_{\mathrm{S}}(\theta)$ as defined in Equation 12.11 has larger support than $p_{\mathrm{ME}}(\theta)$, moving both end points out by a distance $1/h$. We adjust for this using reflection

Again I have assumed that 'ME' and 'SM' are labels – OK?

$$p_{\mathrm{SM}}(\theta) = \begin{cases} p_{\mathrm{S}}(\theta) + p_{\mathrm{S}}(a - \theta) & \text{for } a \leqslant \theta < a + 1/h \\ p_{\mathrm{S}}(\theta) & \text{for } a + 1/h \leqslant \theta < b - 1/h \\ p_{\mathrm{S}}(\theta) + p_{\mathrm{S}}(2b - \theta) & \text{for } b - 1/h \leqslant \theta \leqslant b \end{cases}$$

Change to 'cases' environment here OK?

The resulting smoothed density has support on $[a, b]$ and integrates to 1 (Schuster 1985). The prior distribution for θ is shown in Figure 12.1.

Model II requires a prior on the asset correlation ρ. Here B2 provides guidance. 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

Change OK?

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 generalisation 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 12.2. 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 choose a specification that has the standard B2 model at its mean value, so that allowing for non-zero τ is a strict generalisation of existing practice.

Clarify sentence?

INFERENCE

Writing the likelihood function generically as $p(R \mid \phi)$ with $\phi \in \{\theta, (\theta, \rho), (\theta, \rho, \tau)\}$ depending on whether we are referring to the likelihood function from Equations 12.2, 12.7 or 12.8), 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

Change OK?

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

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

$$p(R) = \int p(R, \phi) \, \mathrm{d}\phi \tag{12.12}$$

Comma OK in integrand or $p(\phi \mid R)$ as elsewhere?

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

$$p(\phi \mid R) = \frac{p(R \mid \phi)p(\phi)}{p(R)} \tag{12.13}$$

Given the distribution $p(\phi \mid R)$, we might ask for a summary statistic, a suitable estimator for plugging into the required capital formulas as envisioned by the Basel Committee on Banking Supervision (2006). A natural value to use is the posterior expectation, $\bar{\phi} = E(\phi \mid R)$. 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 $p(\phi \mid R)$ 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 I–III are 1, 2 and 3, respectively. The first model can be reliably integrated by direct numerical integration, as can model II (requiring rather more time). Model III 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 MCMC concept briefly and give details specific to our application. For a thorough and wide-ranging description see Chapter 2 (in particular, page 53) and Robert and Casella (2004).

Changes to sentence OK?

Change OK?

MCMC methods are a wide class of procedures for calculating posterior distributions, or more generally sampling from a distribution when the normalising 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 $\phi^0, \phi^1, \dots, \phi^N$ from $p(\phi \mid R)$, when $p(\phi \mid R)$ 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 $p(\phi \mid R)$. We begin with a proposal distribution $q(\phi' \mid \phi)$ giving a new value of ϕ depending stochastically on the current value. Assume (for simplicity; this assumption is easily dropped) that $q(\phi' \mid \phi) = q(\phi \mid \phi')$. This distribution should be easy to sample from and in fact is often taken to be normal: $\phi' = \phi + \varepsilon$, 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 $q(\phi' \mid \phi^n)$, then defining

Word added – OK?

$$\alpha(\phi', \phi^n) = p(R, \phi')/p(R, \phi^n) \wedge 1$$

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

$$p(R, \phi')/p(R, \phi^n) = p(\phi' \mid R)/p(\phi^n \mid R)$$

since the normalising constant $p(R)$ cancels. The resulting sample $\phi^0, \phi^1, \dots, \phi^N$ is a sample from a Markov chain with equilibrium distribution $p(\phi \mid R)$. Eventually (in N) the chain will settle down and the sequence will approximate a sequence of draws from $p(\phi \mid R)$. 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) used in "R" (R Development Core Team 2009). Some experimentation with these methods is useful to gain understanding. Valuable guidance and associated warnings are available on the website noted in the package documentation. Generally, an acceptance ratio of about 25% is good (Roberts *et al* 1997). The acceptance rate is tuned by adjusting the variances of ε. Long runs are better than short ones. There is essentially no way to prove that convergence has occurred, though non-convergence is often obvious from time-series plots. For our illustrative application, M samples from the joint posterior distribution were taken after a 5,000-sample burn-in. Scaling of the proposal distribution allowed an acceptance rate of between 22% and 25%. This procedure was used for Model II (M = 10,000) and for Model III (M = 40,000). Calculation of posterior distributions of the parameters and the functions of parameters considered below 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 (DRS) database (release date January 8, 2010). These are restricted to US domiciled, non-financial and non-sovereign entities. Default rates were computed for annual cohorts of firms starting in January 1999 and running through to January 2009. In total there are 2,642 firm/years of data and 24 defaults, for an overall empirical rate of 0.00908. The data is shown in Figure 12.3.

Change OK?

Word added – OK?

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

Figure 12.3 Default rates

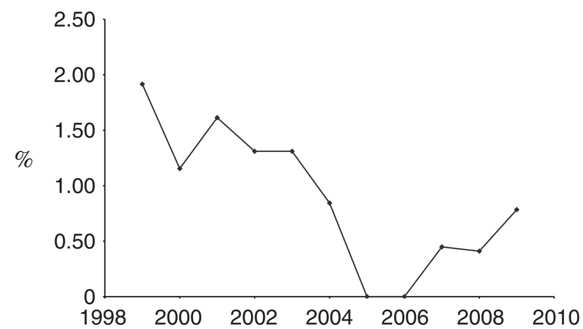
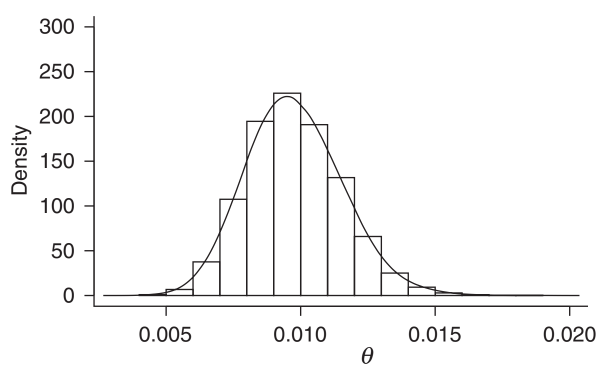


Figure 12.4 Model I, $p(\theta \mid R)$



Horizontal axis shows Moody's BA default rates: annual cohorts 1999–2009.

conjugate-updating framework due to the flexible form of the prior representation).

The posterior distribution for the binomial model is shown in Figure 12.4.

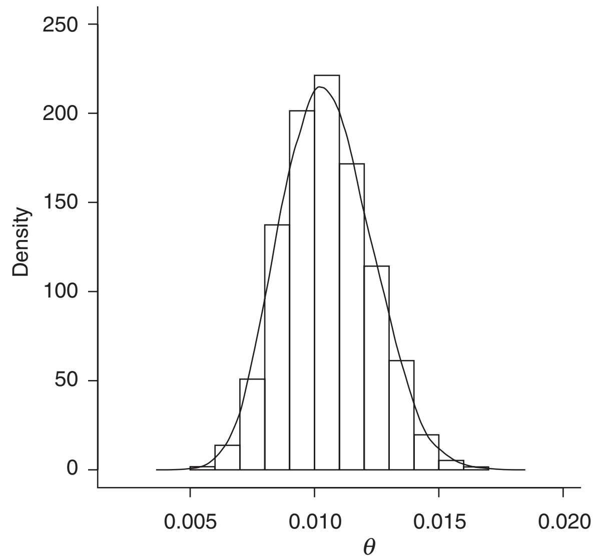
Details from horizontal axis label added to figure notes here and in later figures – OK?

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

Note that this is higher than the empirical default rate of 0.0091. The right-skewness of the distribution is evident, which has flowed through from the prior distribution. The 95% credible interval for θ is $(0.00662, 0.0134)$, which corresponds to a relative uncertainty of about 68% for the estimated PD.

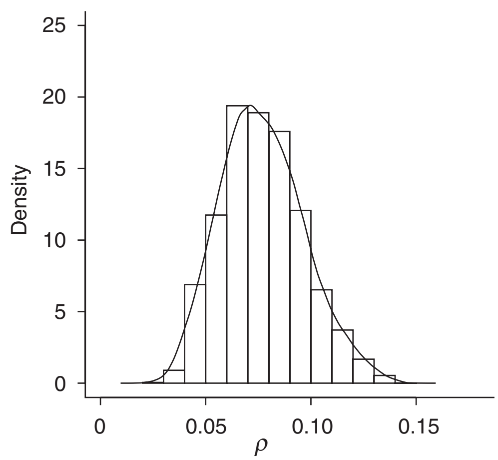
Model II has asset value correlation within periods, allowing for heterogeneity in the default rate over time (but not correlated over

Figure 12.5 Model II, $p(\theta \mid R)$



Horizontal axis shows Moody's BA default rates: annual cohorts 1999–2009.

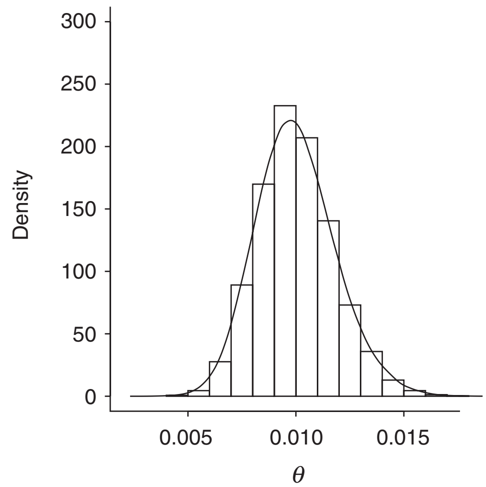
Figure 12.6 Model II, $p(\rho \mid R)$



Horizontal axis shows Moody's BA default rates: annual cohorts 1999–2009.

time) and clumping of defaults. The marginal posterior distributions are shown in Figures 12.5 and 12.6.

Figure 12.7 Model III $p(\theta \mid R)$



Horizontal axis shows Moody's BA default rates: annual cohorts 1999–2009.

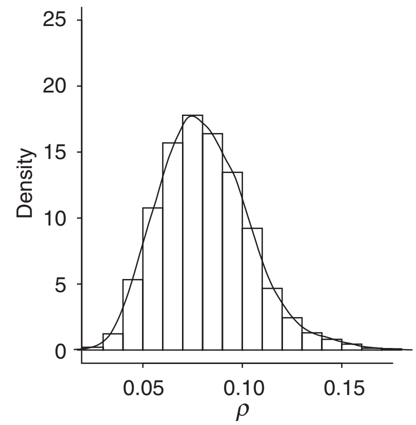
We observe that the estimate of the probability of default in this model is higher than in the one-parameter model, this density having $E(\theta \mid R) = 0.0105$ and $\sigma_\theta = 0.00175$. The 95% credible interval for θ is $(0.0073, 0.0140)$. This density has $E(\rho \mid R) = 0.0770$ and $\sigma_\rho = 0.0194$, so that there is a higher degree of variability relative to the mean in the estimated distribution of the asset value correlation compared with the PD parameter. The 95% credible interval for ρ is $(0.0435, 0.119)$. 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.

Changes to sentence OK?

The marginal posterior distributions for Model III are shown in Figures 12.7–12.9.

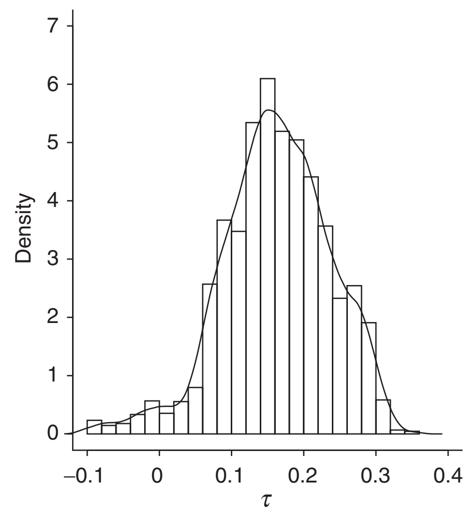
We observe that the estimate of the probability of default in this model is slightly higher than in the one-parameter model, this density having $E(\theta \mid R) = 0.0100$ and $\sigma_\theta = 0.00176$. This density has

Figure 12.8 Model III, $p(\rho \mid R)$



Horizontal axis shows Moody's BA default rates: annual cohorts 1999–2009.

Figure 12.9 Model III, $p(\tau \mid R)$



Horizontal axis shows Moody's BA default rates: annual cohorts 1999–2009.

$E(\rho \mid R) = 0.0812$ and $\sigma_\rho = 0.0185$ with a 95% credible interval of $(0.043, 0.132)$. The density of the autocorrelation parameter in the latent systematic factor has $E(\tau \mid R) = 0.162$ and $\sigma_\tau = 0.0732$. The 95% credible interval is $(-0.006, 0.293)$.

In summary, the picture on the default probability is pretty clear: it is around 0.01 in all models. The asset value correlation is around 0.08, estimated to be slightly higher in model III than in model II. This is substantially less than the value specified in B2. The temporal correlation in the systematic factor is only present in model III. The evidence is sparse here (recall that there are only 11 years of data and the prior information was as uninformative as possible) but it appears to be slightly positive.

Change OK?

CONCLUSION

In this and related applications the econometrician faces the dual chore of modelling the data distribution with a specification of a statistical distribution and modelling expert information with a statistical distribution. Adding the latter task substantially increases the range of applicability of econometric methods. This is clearly an area for further research. Our application has gone through the steps of a formal Bayesian analysis, focusing on the default probability, a key parameter which is required to be estimated under B2 by a large number of institutions worldwide. We concluded our analysis by generating the posterior distributions for the parameters of a nested sequence of models and calculating summary statistics. The mean default probability would be a natural estimator to use for calculating minimum regulatory capital requirements using the formulas provided by B2. In practice, these distributions have many uses, and the analysis would be ongoing. For example, institutions might want to use the entire distribution of the default probability in pricing credit and in setting in-house capital levels. Having an entire distribution for the PD could also be useful in stressing of IRB models, ie, plugging in a high quantile of PD into the capital formula to model an adverse scenario. 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. An analysis of loss given default (LGD) using Bayesian methods would be useful: here there is substantial experience and a joint analysis of LGD and the default probability is likely to be extremely interesting. These and many other possible analyses build on the methods illustrated here.

'Industrial Revenue Bond'? Please spell out, as this abbreviation is not used elsewhere.

This research was funded by the Danish Science Foundation, University of Aarhus, Denmark. The statements made and views expressed herein are solely those of the authors, and do not necessarily represent official policies, statements or views of the Office of the Comptroller of the Currency or its staff.

1 See, for example, **?**, Section 2.5 for details.

2 See also page 10 of Chapter 1.

3 For details, see **?** or, for an approach not using the Lagrangian, **?**.

The '?' in endnotes refer to missing references. Please provide details, as requested in the body of the chapter. Geyer reference typeset as a book – OK?

REFERENCES

Basel Committee on Banking Supervision, 2006, *International Convergence of Capital Measurement and Capital Standards* (Basel: Bank for International Settlements).

Basel Committee on Banking Supervision, 2009, "Proposed Enhancements to the Basel II Framework: Consultative Document", Discussion Paper, Bank for International Settlements.

Das, S. R., D. Duffie, N. Kapadia and L. Saita, 2007, "Common Failings: How Corporate Defaults Are Correlated", *Journal of Finance* 62, pp. 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, pp. 680–700.

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, pp. 199–232.

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

Kiefer, N. M., 2007, "The Probability Approach to Default Probabilities", *Risk Magazine* July, pp. 146–50.

Kiefer, N. M., 2009, "Correlated Defaults, Temporal Correlation, Expert Information and Predictability of Default Rates", Discussion Paper, Cornell University.

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

McCullagh, P., and J. Nelder, 1989, *Generalized Linear Models*, Second Edition (London: Chapman & Hall).

McNeil, A. J., and J. P. Wendin, 2007, "Bayesian Inference for Generalized Linear Mixed Models of Portfolio Credit Risk", *Journal of Empirical Finance* 14, pp. 131–49.

Nickell, P., W. Perraudin and S. Varotto, 2000, "Stability of Rating Transitions", *Journal of Banking and Finance* 24, pp. 203–27.

Office of the Comptroller of the Currency, 2000, "Risk Modeling: Model Validation", Discussion Paper 2000-16, 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* (Vienna,: R Foundation for Statistical Computing).

Robert, C., and G. Casella, 2004, *Monte Carlo Statistical Methods*, Second Edition (Springer).

Roberts, G. O., A. Gelman and W. R. Gilks, 1997, "Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms", *Annals of Applied Probability* 7(1), pp. 110–20.

Schuster, E. F., 1985, "Incorporating Support Constraints into Nonparametric Estimators of Densities", *Communications in Statistical Theory and Methods* 14, pp. 1123–36.