Journal of Empirical Finance 16 (2009) 164–173

Contents lists available at ScienceDirect

Journal of Empirical Finance

journal homepage: www.elsevier.com/locate/jempfin

Default estimation for low-default portfolios ☆, ☆☆

Nicholas M. Kiefer

Cornell University, Departments of Economics and Statistical Science and Center for Analytic Economics, 490 Uris Hall, Ithaca, NY 14853-7601, United States

ARTICLE INFO

Article history:
Received 15 May 2007
Received in revised form 21 March 2008
Accepted 28 March 2008
Available online 6 April 2008

JEL classification:
C11
C13
C44
G18
G32

Keywords:
Bayesian inference
Bayesian estimation
Expert information
Basel II
Risk management

ABSTRACT

Risk managers at financial institutions are concerned with estimating default probabilities for asset groups both for internal risk control procedures and for regulatory compliance. Low-default assets pose an estimation problem that has attracted recent concern. The problem in default probability estimation for low-default portfolios is that there is little relevant historical data information. No amount of data data-processing can fix this problem. More information is required. Incorporating expert opinion formally is an attractive option. The probability (Bayesian) approach is proposed, its feasibility demonstrated, and its relation to supervisory requirements discussed.

1. Introduction

The Basel II framework (Basel Committee on Banking Supervision, 2004) for capital standards provides for banks to use models to assess risks and determine minimum capital requirements. All aspects of the models – specification, estimation, validation – will have to meet the scrutiny of national supervisors.

The presumption is that these models will be the same ones that sophisticated institutions use to manage their loan portfolios. Banks using internal ratings-based (IRB) methods to calculate credit risks must calculate default probabilities (PD), loss given default (LGD), exposure at default (EAD) and effective maturity (M) for groups of homogeneous assets. For very safe assets calculations based on historical data may "not be sufficiently reliable" (Basel Committee on Banking Supervision, 2005) to form a probability of default estimate, since so few defaults are observed. This issue has attracted attention in the trade literature, for exampleBalthazar (2004). Methods which advocate departing from the usual unbiased estimator have been proposed by Pluto and Tasche (2005). A related estimator for PD in low-default portfolios based on the CreditRisk+ model is proposed by Wilde and Jackson (2006). A Bayesian approach is proposed by Dwyer (2006), who uses the Bayesian probability mechanics (an improvement over standard practice) but does not incorporate expert information.

In this paper I argue that uncertainty about the default probability should be modeled the same way as uncertainty about defaults — namely, represented in a probability distribution. A future default either occurs or doesn't (given the definition). Since

☆ US Department of the Treasury, Office of the Comptroller of the Currency, Risk Analysis Division, and CREATES, funded by the Danish Science Foundation, University of Aarhus, Denmark.

☆☆ 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.

E-mail address: nicholas.kiefer@cornell.edu.

0927-5398/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jempfin.2008.03.004

we do not know in advance whether it occurs or not, we model this uncertain event with a probability distribution. This model reflects our partial knowledge of the default mechanism. Similarly, the default probability is unknown. But experts do know something about the latter, and we can represent this knowledge in a probability distribution. Inference should be based on a probability distribution for the default probability. The final distribution should reflect both data and expert information. This combining of information is easy to do using Bayes rule, once the information is represented in probability distributions. The result is an estimator which is different from the unbiased estimator, but which moves the unbiased estimator toward an expert opinion rather than simply bounding it away from zero.

For convenience and ease of exposition I focus here on estimating the default probability θ for a portfolio of safe assets.Section 2 treats the specification of the likelihood function and indicates what might be expected from the likelihood function. This paper is mainly concerned with the incorporation of expert information, so a simple likelihood specification, which is nevertheless compatible with industry practice and the Basel II prescriptions, is sufficient. General comments on the modeling of uncertainty through probabilities, the standard approach to default modeling, are made inSection 3. The approach is applied to expert information about the unknown default probability and how that might be represented. Specifically, it is represented in a probability distribution, for exactly the same reasons that uncertainty about defaults is represented in a probability distribution. Combination of expert and data information is taken up inSection 4, following for example DeGroot (1970). Section 5 considers elicitation of an expert's information and its representation in a probability distribution.Section 6 treats the inferences that could be made on the basis of the expert information and likely data information. It is possible in the low-default case to consider all likely data realizations in particular samples. Section 7 considers additional inference issues and supervisory issues.Section 8 concludes.

2. The likelihood function

Expert judgement is crucial at every step of a statistical analysis. Expert knowledge could be the result of accumulated experience with similar problems and data or simply the result of knowledgeable consideration. Typical data consist of a number of asset/years for a group of similar assets. In each year there is either a default or not. This is a clear simplification of the actual problem in which asset quality can improve or deteriorate and assets are not completely homogeneous. Nevertheless, it is useful to model the problem as one of independent Bernoulli sampling with unknown parameter θ. This is especially appropriate in the low-default case, as models dealing with realistic complications in large datasets with many defaults will not be supported by the data when the sample size is small and the number of defaults is low (typically 0,1, or 2). With large datasets exhibiting a significant default experience the independent Bernoulli model is the simplest but not the only possibility. Certainly independence is a strong assumption and would have to be considered carefully. Note that independence here is conditional independence. The marginal (with respect to θ; see below) distribution of D certainly exhibits dependence. It is through this dependence that the data are informative on the default probability. Second, the assumption that the observations are identically distributed may be unwarranted. Perhaps the default probabilities differ across assets, and the most risky generally default first. In low-default portfolios, these issues do not arise. Note that all of these assumptions must be explicit and are subject to supervisory review. See OCC (2006).

Let $D=\{d_i, i=1,\ldots,n\}$ denote the whole dataset and $r=r(D)=\sum_i d_i$ the count of defaults. Then the joint distribution of the data is

$$\begin{aligned} p(D|\theta) &= \Pi\theta^{d_i}(1-\theta)^{1-d_i} \\ &= \theta^r(1-\theta)^{n-r} \end{aligned}. \tag{1}$$

As a function of θ for given data D, this is the likelihood function $L(\theta|D)$. 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 distribution of r

$$p(r|\theta) = \binom{n}{r}\theta^r(1-\theta)^{n-r}. \tag{2}$$

Regarded as a function of θ for given data, Eq.(2) is the likelihood function $L(\theta|r)$. Since $r(D)$ is a sufficient statistic no other function of the data is informative about θ given $r(D)$. All of the relevant data information on θ comes through the distribution $p(r|\theta)$. The strict implication is that no amount of data-massaging or data-processing can improve the data evidence on θ. Fig. 1 shows the normed likelihood functions $\bar{L}(\theta|r)=L(\theta|r)/\max_\theta L(\theta|r)$ for $r=0$, 1, 2, 5, and $n=100$. These figures illustrate the sorts of observed likelihood functions one might see in practice.

Fig. 1 illustrates that small changes in the realized number of defaults can have a substantial effect on the maximum likelihood estimator (MLE). Thus, for $n=100$, an increase by 1 in the number of defaults increases the MLE by .01. If the probability being estimated is large (e.g., 0.3), then a difference in the estimate of 0.01 is not perhaps as dramatic as when the realistic values are 0.01 or 0.02. Further, these small estimates are sharply determined, according to the shape of the likelihood functions.

A different point of view can be illustrated by the expected likelihood function for a given hypothetical value of θ. Fig. 2 plots $\sum_j \bar{L}(\theta|r_j)p(r_j|\theta_0)$ for $\theta_0=0.005$, 0.01, 0.02, and 0.05 and $n=100$. This function is rather more spread than the likelihood on given data (note that $\bar{L}(\theta|r)$ is concave in r for values near the most likely value $n\theta$).

3. Expert information

It is absolutely clear that there is some information available about θ in addition to the data information. For example, we expect that the portfolio in question is a low-default portfolio. Where does this expectation come from? We would be surprised if θ

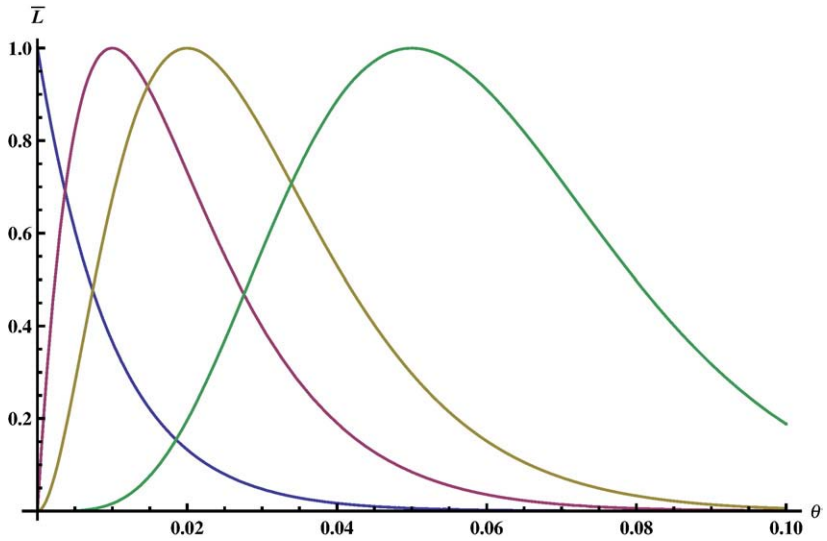

Fig. 1. Likelihood, $n=100$, $r=0,1,2,5$.

for such a portfolio turned out to be, say, 0.2. Further, there is a presumption that no portfolio has default probability 0. This information should be organized and incorporated in the analysis in a sensible way. This involves quantification of the information or, alternatively, quantification of the uncertainty about θ.

Quantification of uncertainty requires comparison with a standard. One standard for measuring uncertainty is a simple experiment, such as drawing balls from an urn at random as above, or sequences of coin flips. We might begin by defining events for consideration. Examples of events are $A=$"$\theta\leq 0.005$"; "$B=$"$\theta\leq 0.01$"; $C=$"$\theta\leq 0.015$, "etc. Assign probabilities by comparison. For example A is about as likely as seeing three heads in 50 throws of a fair coin. Sometimes it is easier to assign probabilities by considering the relative likelihoods of events and their complements. Thus, either A or "*not A*" must occur. Suppose A is considered twice as likely as "*not A*." Then the probability of A is 2/3, since we have fixed the ratio and the probabilities must add up to one. Some prefer to recast this assessment in terms of betting. Thus, the payout x is received if A occurs, $(1-x)$ if not. Again, the events are exhaustive and mutually exclusive. Adjust x until you are indifferent between betting on A and "*not A*." Then, it is reasonable to assume for small bets that $xP(A)=(1-x)(1-P(A))$ or $P(A)=(1-x)$. These possibilities and others are discussed in Berger (1980). It is clear that assessing probabilities requires some thought and some practice, but also that it can be done. It can be shown that beliefs that satisfy certain consistency requirements, for example that the believer is unwilling to make sure-loss bets, lead to measures of uncertainty that combine according to the laws of probability: convexity, additivity and multiplication. See for example DeGroot (1970).

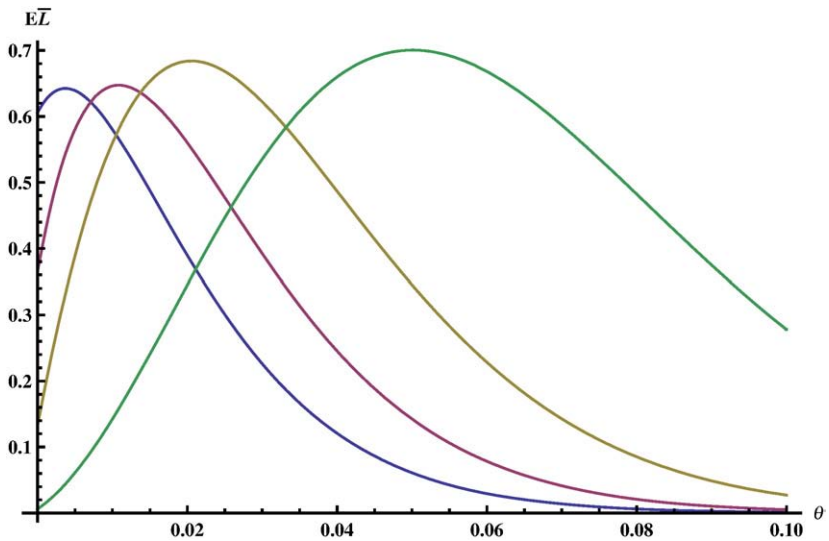

Fig. 2. Expected likelihood $\theta=0.005$, 0.01, 0.02. 0.05.

Once probabilities have been elicited, we turn to the practical matter of specifying a functional form for the prior distribution $p(\theta)$. A particularly easy specification is the uniform $p(\theta)=1$ for $\theta\in[0,1]$. This prior would sometimes be regarded as "uninformative," (with the implied additional property "unobjectionable") since it assigns equal probability to equal length subsets of [0,1]. The mean of this distribution is 1/2. Other moments also exist, and in that sense it is indeed informative (a prior expectation of default probability 1/2 might not be considered suitable for low-default portfolios). A generalization of the uniform in common use for a parameter that is constrained to lie in [0,1] is the beta distribution. The beta distribution for the random variable $\theta\in[0,1]$ with parameters (α,β) is

$$p(\theta|\alpha,\beta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}. \tag{3}$$

A somewhat richer specification is the beta distribution (Eq.(3)) modified to have support $[a, b]$. It is possible that some applications would require the support of θ to consist of the union of disjoint subsets of [0,1] but this seems fanciful in the current application. Let t have the beta distribution and upon change variables to $\theta(t)=a+(b-a)t$ with inverse function $t(\theta)=(\theta-a)/(b-a)$ and Jacobian $dt(\theta)/d\theta=1/(b-a)$. Then

$$p(\theta|\alpha,\beta,a,b)=\frac{\Gamma(\alpha+\beta)}{(b-a)\Gamma(\alpha)\Gamma(\beta)}((a-\theta)/(a-b))^{\alpha-1}((\theta-b)/(a-b))^{\beta-1} \tag{4}$$

over the range $\theta\in[a,b]$. This distribution has mean $E\theta=(b\alpha+a\beta)/(\alpha+\beta)$. The four-parameter beta distribution allows flexibility within the range $[a,b]$, but in some situations it may be too restrictive. A simple generalization is the seven-parameter mixture of two four-parameter betas with common support. The additional parameters are the two new $\{\alpha,\beta\}$ parameters and the mixing parameter λ.

$$\begin{aligned}p(\theta|\alpha_1,\beta_1,\alpha_2,\beta_2,a,b)=&\frac{\lambda\Gamma(\alpha_1+\beta_1)}{(b-a)\Gamma(\alpha_1)\Gamma(\beta_1)}((a-\theta)/(a-b))^{\alpha_1-1}((\theta-b)/(a-b))^{\beta_1-1}\\&+\frac{(1-\lambda)\Gamma(\alpha_2+\beta_2)}{(b-a)\Gamma(\alpha_2)\Gamma(\beta_2)}((a-\theta)/(a-b))^{\alpha_2-1}((\theta-b)/(a-b))^{\beta_2-1}.\end{aligned}$$

Computations with this mixture distribution are not substantially more complicated than computations with the four-parameter beta alone. If necessary, more mixture components with new parameters can be added, although it seems unlikely that expert information would be detailed and specific enough to require this complicated a representation. There is a theory on the approximation of general prior distributions by mixtures of conjugate distributions. By choosing enough beta-mixture terms the approximation of an arbitrary continuous prior $p(\theta)$ for a Bernoulli parameter can be made arbitrarily accurate. See Diaconis and Ylvisaker (1985). Useful references on the choice of prior distribution are Box and Tiao (1992) and Jaynes (2003).

4. Updating (learning)

With $p(\theta)$ describing expert opinion and the statistical model for the data information $p(r|\theta)$ at hand, we are in a position to combine the expert information with the data information to calculate $p(\theta|r)$, the posterior distribution describing the uncertainty about θ after observation of r defaults in n trials. The rules for combining probabilities imply $P(A|B)P(B)=P(A \text{ and } B)=P(B|A)P(A)$, or more usefully $P(B|A)=P(A|B)P(B)/P(A)$, assuming $P(A)>0$. Applying this rule gives Bayes' rule for updating beliefs

$$p(\theta|r)=p(r|\theta)p(\theta)/p(r). \tag{5}$$

The denominator $p(r)$, is the unconditional distribution of the number of defaults,

$$p(r)=\int p(r|\theta)p(\theta)d\theta. \tag{6}$$

$p(r)$ is also called the predictive distribution of the statistic r.

For the purpose of predicting the number of defaults in a portfolio of a given size, the predictive distribution(6)is relevant. For inference about the default probability θ, for example for input into the Basel capital formula, the posterior distribution (5) is relevant. Further discussion of the beta-binomial analysis sketched here and of applications to other models is given by Raiffa and Schlaifer (1961). On the Bayesian approach to econometrics see Zellner (1996), a reprint of the influential 1971 edition.

5. Prior distribution

I have asked an expert to specify a portfolio and give me some aspects of his beliefs about the unknown default probability. The portfolio consists of loans to highly-rated, large, internationally active and complex banks. The method included a specification of the problem and some specific questions over e-mail 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 by Kadane et al. (1980) and Kadane and Wolfson (1998). An example assessing a prior for a Bernoulli parameter is Chaloner and Duncan (1983).

Chaloner and Duncan follow Kadane et al. in suggesting that assessments be done not directly on the probabilities concerning the parameters, but on the predictive distribution. That is, questions should be asked about observables, to bring the expert's thoughts closer to familiar ground. Thus, in the case of defaults, a lack of prior knowledge might indicate that the predictive probability of the number of defaults in a sample of size n would be $1/(n+1)$. Departures from this predictive distribution indicate

prior knowledge. In the case of a Bernoulli parameter and a two-parameter beta prior, Chaloner and Duncan suggest first eliciting the mode of the predictive distribution for a given n (an integer), then assessing the relative probability of the adjacent values. Graphical feedback is provided for refinement of the specification. Examples used by Chaloner and Duncan consider $n=20$; perhaps the method would be less attractive for the large sample sizes and low probabilities we anticipate. The suggestion to interrogate experts on what they would expect to see in data, rather than what they would expect of parameter values, is appealing and I have to some extent pursued this with our expert.

It is necessary to specify a period over which to define the default probability. The "true" default probability has probably changed over time. Recent experience may be thought to be more relevant than the distant past, although the sample period should be representative of experience through a cycle. It could be argued that a recent period including the 2001–2002 period of mild downturn covers a modern cycle. A period that included the 1980's would yield higher default probabilities, but these are probably not currently relevant. The default probability of interest is the current and immediate future value, not a guess at what past estimates might be. There are 50 or fewer banks in this highly highly-rated category, and a sample period over the last seven years or so might include 300 observations as a high value. For our application, we considered a "small" sample of 100 observations and a "large" sample of 300 observations. Of course, the prior does not depend on the sample size.

We began by considering first the predictive distribution on 300 observations, the modal value was zero defaults. Upon being asked to consider the relative probabilities of zero or one default, conditional on one or fewer defaults occurring, the expert expressed some trepidation as it is difficult to think about such rare events. This was the method suggested by Chaloner and Duncan (1983) in an application involving larger probabilities and smaller datasets. Of course, the conditioning does not matter for the relative probabilities, but it may be easier for an expert to focus attention given the explicit conditioning. Our expert had difficulty thinking about the hypothetical default experiences and their relative likelihoods. The expert was quite happy in thinking about probabilities over probabilities, however. This may not be so uncommon in this technical area, as practitioners are accustomed to working with probabilities. 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 0.05. The median value was 0.0033. The expert remarked that the mean at 0.005 was larger than the median. Quartiles were assessed by asking the expert to consider the value at which larger or smaller values would be equiprobable given that the value was less than the median, then given that the value was more than the median. The former seemed easier to think about and was 0.00225 ("between 20 and 25 basis points"). The latter, the .75 quartile, was assessed at .025. Our expert found it much easier to think in terms of quantiles than in terms of moments.

This set of answers is more than enough information to determine a four-parameter beta distribution. I used a method of moments to fit parametric probability statements to the expert assessments. The moments I used were squared differences relative to the target values, for example $((a-0.0001)/0.0001)^2$. The support points were quite well-determined for a range of $\{\alpha, \beta\}$ pairs at the assessed values $\{a, b\}=[0.0001, 0.05]$. These were allowed to vary but the optimization routine did not change them beyond the 7th decimal place. The rather high value of b reflects the long tail apparently desired by the expert. The $\{\alpha, \beta\}$ parameters were rather less well-determined (the sum of squares function was fairly flat) and I settled on the values (1.9, 21.0) as best describing the expert's information. Changing the weights in the fitting routine did not substantially change any of the parameter values. The resulting prior distribution $p(\theta)$ is graphed in Fig. 3.

The median of this distribution is 0.0036, the mean is 0.0042. In practice, after the information is aggregated into an estimated probability distribution, then additional properties of the distribution would be calculated and the expert would be consulted

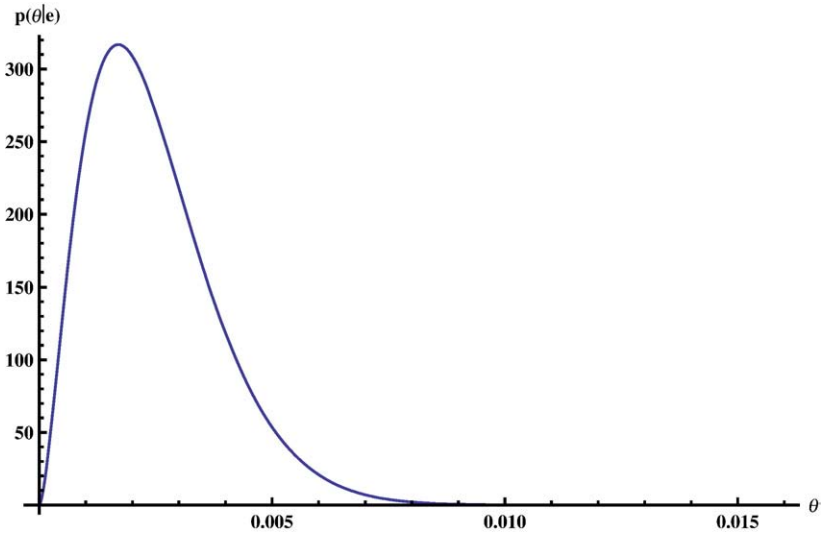

Fig. 3. Prior distribution representing expert information.

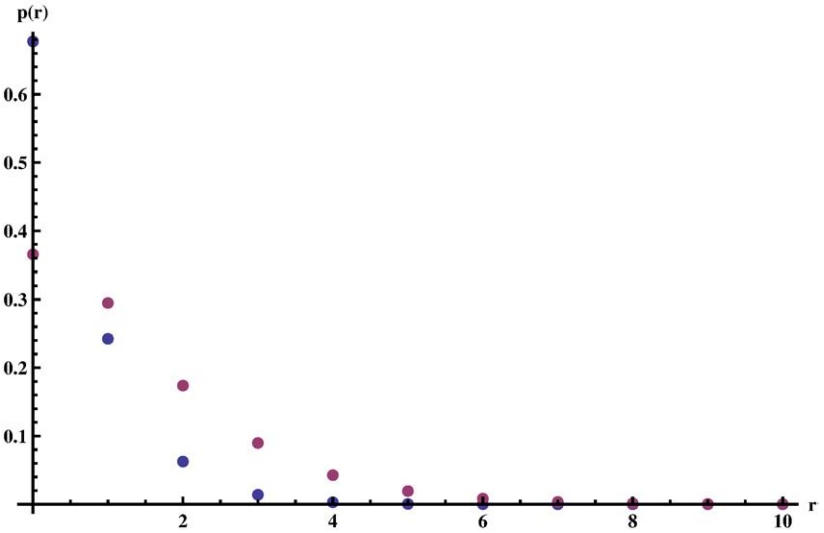

Fig. 4. Predictive distributions, $n=100$ (blue), 300. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

again to see if any changes were in order before proceeding to data analysis (Lindley, 1982). This process would be repeated as necessary. In the present application there was one round of feedback. This was valuable since the expert had had time to consider the probabilities involved. The characteristics reported are from the second round of elicitation.

The predictive distribution (6) corresponding to this prior is given in Fig. 4 for $n=100$ and $n=300$. With our specification, the expected value of $r, E(r)=\sum_{k=0}^{n} kp(k)$ is 0.424 for $n=100$ and 1.27 for $n=300$. The modal value is zero for both sample sizes. Defaults are expected to be rare events.

It is interesting to compute the unconditional expected likelihood

$$E\overline{L}(\theta)=\sum_j \overline{L}(\theta|r_j)p(r_j)$$

for comparison with Fig. 2. This is given in Fig. 5 for $n=\{100, 300, 500\}$.

6. Posterior analysis

Recall that the dataset information on the default probability is completely specified by the sample size n and the number of defaults r. Sample sizes between 50 and 100 (and less than 50) are not unrealistic and may be typical. The number of defaults $r=0$ is

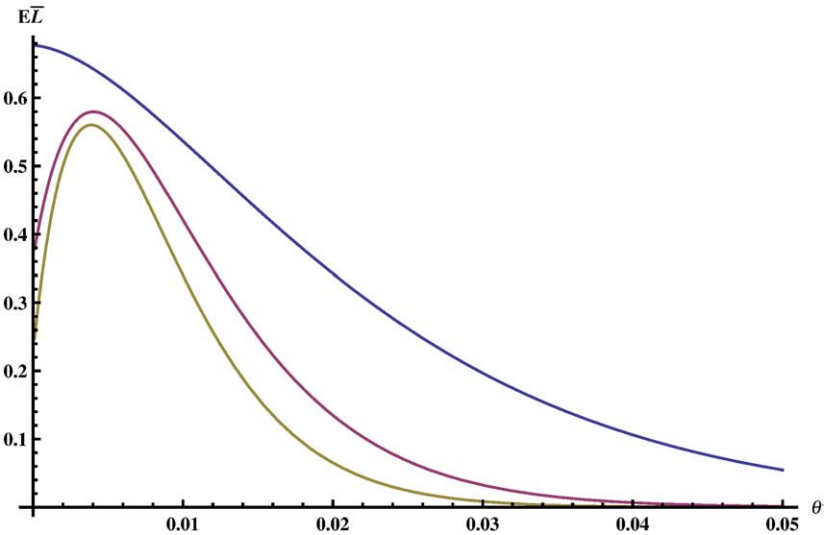

Fig. 5. Expected likelihoods, $n=100, 300, 500$.

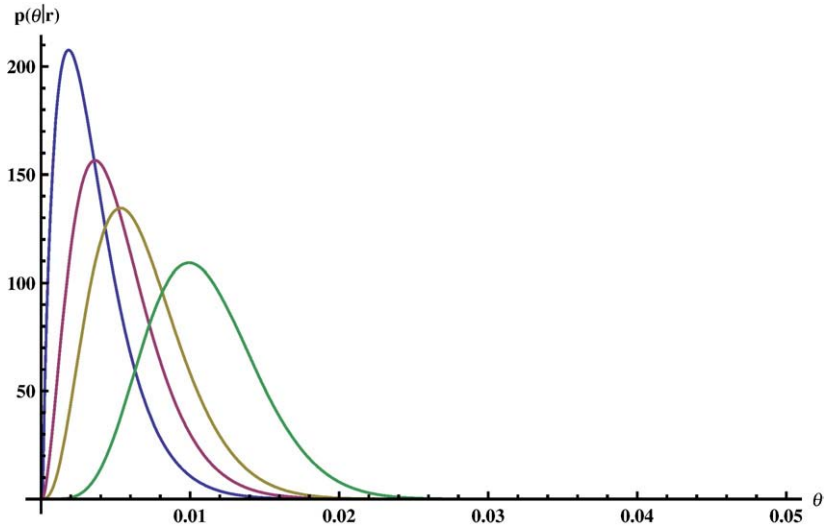

Fig. 6. Posterior distributions, $n=100$, $r=\{0,1,2,5\}$.

typical for these datasets and in fact this is the case that has prompted the literature on low-default portfolios. Thus, the results presented are applications in the sense that there are extremely relevant datasets with these summary statistics. The posterior distribution, $p(\theta|r)$, is graphed in Fig. 6 for $r=0$, 1, 2 and 5 and $n=100$ and in Fig. 7 for $r=0$, 1, 3 and 10 and $n=300$. The corresponding likelihood functions, for comparison, were given in Fig. 1. Note the substantial differences in location. Comparison with the prior distribution graphed in Fig. 3 reveals that the expert provides much more information to the analysis than do the data.

Given the distribution $p(\theta|r)$, 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 (2004). A natural value to use is the posterior expectation, $\bar{\theta}=E(\theta|r)$. The expectation is an optimal estimator under quadratic loss and is asymptotically an optimal estimator under a wide variety of loss functions. An alternative, by analogy with the maximum likelihood estimator $\hat{\theta}$, is the posterior mode $\dot{\theta}$. As a summary measure of our confidence we would use the posterior standard deviation $\sigma_\theta=\sqrt{E(\theta-\bar{\theta})^2}$. By comparison, the usual approximation to the standard deviation of the maximum likelihood estimator is $\alpha_{\hat{\theta}}=\sqrt{\hat{\theta}\left(1-\hat{\theta}\right)/n}$. These quantities are given in Table 1 for a variety of combinations of n and r.

Which procedure gives the most useful results for the hypothetical datasets? The maximum likelihood estimator $\hat{\theta}$ is very sensitive to small changes in the data. One might imagine that updating would be done periodically, leading to occasional substantial jumps in the estimator. For $n=100$, the MLE ranges from 0.00–0.05 as the number of defaults ranges from 0 to 5 (the last value is incredibly unlikely). The posterior mean ranges in the same case from 0.0036 to 0.011, and the posterior mode lies on a similar range slightly left shifted. The major differences between the posterior statistics ($\bar{\theta}$ and $\dot{\theta}$) and $\hat{\theta}$ occur at extremely unusual

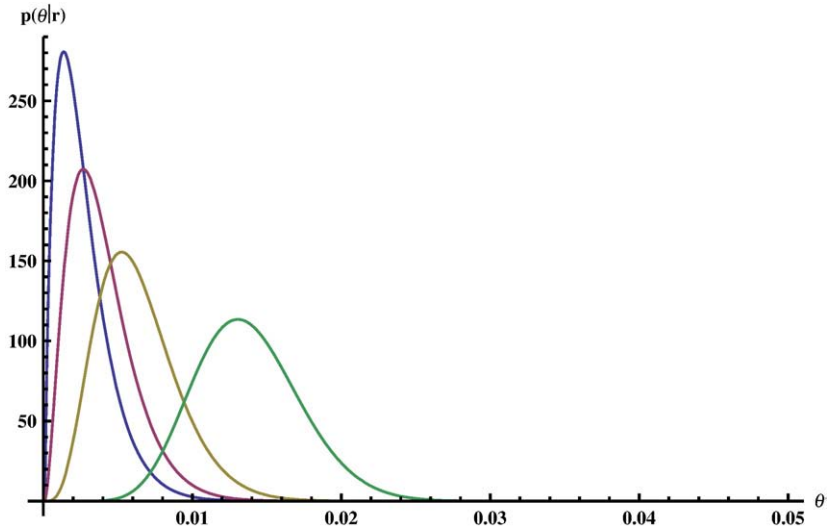

Fig. 7. Posterior distributions, $n=300$, $r=\{0,1,3,10\}$.

Table 1
Default probabilities: Location and precision

n	r	$\bar{\theta}$	$\dot{\theta}$	$\hat{\theta}$	σ_θ	$\sigma_{\hat{\theta}}$
100	0	0.0036	0.0018	0.000	0.0024	0 (!).
100	1	0.0052	0.0036	0.010	0.0028	0.0100
100	2	0.0067	0.0053	0.020	0.0031	0.0140
100	5	0.0109	0.0099	0.050	0.0037	0.0218
300	0	0.0027	0.0014	0.000	0.0018	0(!)
300	1	0.0039	0.0027	0.003	0.0022	0.0033
300	3	0.0064	0.0053	0.010	0.0027	0.0057
300	10	0.0137	0.0131	0.033	0.0035	0.0103
500	0	0.0021	0.0011	0.000	0.0015	0 (!)
500	2	0.0041	0.0032	0.004	0.0020	0.0028
500	10	0.0115	0.0108	0.020	0.0031	0.0063
500	20	0.0190	0.0185	0.040	0.0034	0.0088

Note: $\bar{\theta}$ is the posterior mean, $\dot{\theta}$ the mode, $\hat{\theta}$ the MLE, σ_θ the posterior S.D. and $\sigma_{\hat{\theta}}$ the S.D. of the MLE.

samples, for example the five-default sample in the 100-observation case. This case illustrates the importance of reporting the MLE in addition to the posterior statistics. The comparison gives an idea of the robustness of the inference. Note that a prior that mixes expert information with "diffuse" information, perhaps by mixing with a uniform on $[a, b]$ or a prior which is a mean-preserving spread of the real expert information would move the posterior statistics in the direction of the MLE. If there is a large movement when a small proportion of a diffuse prior is added that is an indication that the results are quite sensitive to the prior precision; that is, that the data and prior are in conflict. In this application, this information is revealed by a simple comparison of the MLE and the posterior statistics. This comparison takes us out of the realm of formal inference, and into the area of specification checking, or in Basel II terms, validation. But what would be the modeler's reaction to such a sample? Would it be that the default probability for this portfolio, thought to be an extremely safe portfolio, is indeed 0.05? A more appropriate reaction would be that there is something unusual happening, signaling a need for further investigation. Perhaps it is just a very unusual sample (in which case the estimate $\hat{\theta}$ is very unusual and it might be better to stick with $\bar{\theta}$ as an indication of the actual default probability). Or perhaps some assets have been misclassified or there are other errors in the data. Or perhaps economic conditions have become so dire that a portfolio with a 5% default is a low-default portfolio. If so, surely some other hints that things are not going well would be available.

7. Remarks on the Bayesian approach

The approach suggested here raises a number of issues worthy of further treatment.

7.1. Assessment and combination of expert information

There is a large literature on probability assessment. Much of this focuses on experts who are not necessarily familiar with formal probability concepts. The situation is somewhat simpler here, as the experts are used to dealing with probabilities and thinking about the ways probabilities combine (but not necessarily with assessing uncertainty about parameters in probabilistic terms).

Thinking about small probabilities is notoriously difficult; Kahneman and Tversky (1974) began a large literature. What are the easiest probability questions to assess when constructing a prior distribution? What are the most informative questions, in terms of tying down prior parameters tightly? How should information be fed back to the expert for revision? How should information from several experts be combined? This is addressed by Garthwaite et al. (2005), Lindley et al. (1979) and many others. Here there are essentially two reasonable possibilities. Answers to the same question from different experts can simply be entered into the GMM calculation as separate equations. Alternatively, they could be averaged as repeated measurements on the same equation (the difference here is only one of weighting). Or, the prior specification could be done for each expert m, and the results combined in a mixture, $p(\theta|e_1,\ldots, e_m)=\sum_m \alpha_m p(\theta|e_m)$, where α_m is the nonnegative weight assigned to the mth expert and $\sum_m \alpha_m=1$. This procedure should be combined with feedback to the experts and subsequent revision.

7.2. Robustness

The issue of robustness of the inference about the default probability arises at the validation stage. Modelers can expect to have to review their prior assessment mechanisms with validators and to provide justification for the methods used.

This is no different from the requirements for any other method of estimation of the default probability (and other required parameters). Prudent modelers will report not only the posterior distribution of θ as well as its mean $\bar{\theta}$ but summary statistics, including in this case the MLE, and any interesting or unusual features of the dataset. "Surprises" in the data will have to be explained. This is not specific to the Bayesian approach, but applicable to any method used. Bayesian robustness issues and procedures for assessing robustness of results are described by Berger and Berliner (1986). Some experimentation shows that inferences are not particularly sensitive to specification of the parameters a and b, as long as r/n is in the interval $[a, b]$, as expected. Note that the posterior distribution with a uniform prior on $[a, b]$ is simply proportional to the likelihood in that interval

(elsewhere, the posterior is zero). Thus, primary attention should be paid to the determination of α and β. Robustness is closely related to issues of supervision, as supervisors will review both the modeling efforts and the validation procedures of institutions.

7.3. Supervision

Subjectivity enters every statistical analysis. For many problems data information is substantial and the subjective elements are perhaps less important. In the present setting subjectivity enters explicitly in the specification of $p(\theta|r)$.

Subjectivity also enters in specification of $p(D|\theta)$, but we are used to that and the explicit dependence on judgement is usually dropped. Similarly, subjectivity enters in the classification of assets into "homogeneous" groups and many other places in settings involving supervision. Supervisors generally insist that the decisions made at the modeling level be logically based and validated. Thus, supervisors are willing to accept subjective decisions as long as they are well grounded. It is a small additional step to add subjective information about plausible parameter values. There should be evidence that due consideration was given to specification of $p(\theta|r)$ as well as the current requirement that $p(D|\theta)$ be justified. As in the case of validation, examples can be provided and standards set, while still relying on banks to perform their own analyses and validation. Newsletter No. 6 was written by the Basel Committee Accord Implementation Group's Validation Subgroup in response to banking industry questions and concerns regarding portfolios with limited loss data. Problem portfolios are those for which a "calculation based on historic losses … would not be sufficiently reliable to form the basis of a probability of default estimate…."(p.1). The newsletter notes that problem portfolios are also those which "may not have incurred recent losses, but historical experience or other analysis might suggest that there is a greater likelihood of losses than is captured in recent data."(p.1). The implication is that the actual probability of default is greater than the measured default rate. This case clearly points to disagreement between data information and a prior, where the prior is explicitly based on other data ("historical experience," not in the current sample) or expert opinion ("other analysis"). The newsletter does not suggest impossible mechanical solutions and instead sticks to sensible recommendations like getting more data. A section heading in the newsletter reads as follows: "A relative lack of loss data can at times be compensated for by other methods for assessing risk parameters." This is precisely what I am proposing. In reference to the Basel II document itself (Basel Committee on Banking Supervision, 2004), the newsletter quotes paragraph 449: "Estimates must be based on historical experience and empirical evidence, and not based purely on subjective or judgmental considerations." This seems to allow both data and nondata information, but not exclusively the latter, and thus to hold open the possibility of combining data evidence with nondata evidence in the formal system of conditional probability. Paragraph 448 notes that "estimates of PD, LGD and EAD must incorporate all relevant, material and available data, information, and methods." This seems to make a distinction between data and other sources of information, which is consistent with our analysis.

One danger is that an institution could claim about a bizarre assessment that it is the prior assessment of an expert who predicts no defaults. And indeed, it might be true. Some standards will be necessary, not just showing that the prior uncertainly was rigorously assessed, but that it meets some general standards of reliability. If asset groups were standardized across banks, then an agency could provide standardized descriptions of expert opinion. Supervisors do not currently seem to think such standardization appropriate or desirable. Could the agencies nevertheless provide some guidance? I think this would be feasible. Newsletter 6 states (p.4), "Supervisors expect to continue to share their experience in implementing the Framework in the case of LDPs in order to promote consistency." Could this mean that supervisors will share expert information to be incorporated into each bank's analysis? Clearly, the role of the supervisor, used to dealing with less formal subjectivity, will have to be defined when it comes to formal (probabilistically described) subjective information.

8. Conclusion

I have considered inference about the default probability for a low-default portfolio on the basis of data information and expert judgement. Examples consider sample sizes of 100 and 300 for hypothetical portfolios of loans to very safe, highly-rated large banks. The sample size of 100 is perhaps most realistic in this setting. I have also represented the judgement of an expert in the form of a probability distribution for combination with the likelihood function. This prior distribution seems to reflect expert opinion fairly well. Errors, which would be corrected through feedback and respecification in practice, are likely to introduce more certainty into the distribution rather than less. It is possible to study the posterior distributions for all of the most likely configurations of defaults in the samples. In each case the modal number of defaults is small. I have reported results for zero defaults through a number of defaults above any reasonable likelihood. In all of these the sample information contributes rather little relative to the expert information. Bounds for the likely value for the default probability (the most likely value and the expected value) are fairly tight within the relevant range of data possibilities. Thus, the data variability which is reasonably expected, and indeed data variability which is highly unlikely, will not affect sensible inference about the default probability beyond the second decimal place. These results raise issues about how banks should treat estimated default probabilities and how supervisors should evaluate both procedures and outcomes for particular portfolios.

Acknowledgements

I thank Jeffrey Brown, Mike Carhill, Hwansik Choi, Erik Larson, Mark Levonian, Anirban Mukherjee, Mitch Stengel, and seminar participants at Cornell University and the OCC. These thanks come with gratitude but without any implication of agreement with my views.

References

Balthazar, L., 2004. Pd estimates for basel ii. Risk 84–85 April.
Basel Committee on Banking Supervision, 2004. International Convergence of Capital Measurement and Capital Standards: a Revised Framework. Bank for International Settlements.
Basel Committee on Banking Supervision, 2005. Basel committee newsletter no. 6: validation of low-default portfolios in the basel ii framework. Technical Report. Bank for International Settlements.
Berger, James O., 1980. Statistical Decision Theory: Foundations, Concepts and Methods. Springer-Verlag.
Berger, James, Berliner, L. Mark, 1986. Robust bayes and empirical bayes analysis with contaminated priors. The Annals of Statistics 14 (2), 461–486 Jun.
Box, G.E.P., Tiao, G.C., 1992. Bayesian Inference in Statistical Analysis. John Wiley & Sons, New York.
Chaloner, Kathryn M., Duncan, George T., 1983. Assessment of a beta prior distribution: Pm elicitation. The Statistician, 32(1/2, Proceedings of the 1982 I.O.S. Annual Conference on Practical Bayesian Statistics), pp. 174–180. Mar.
DeGroot, Morris H., 1970. Optimal Statistical Decisions. McGraw-Hill.
Diaconis, Persi, Ylvisaker, Donald, 1985. In: Bernardo, J.M., DeGroot, M.H., Lindley, D.V., Smith, A.F.M. (Eds.), Quantifying Prior Opinion. Bayesian Statistics, vol. 2. Elsevier Science Publishers BV, North-Holland, pp. 133–156.
Dwyer, Douglas W., 2006. The distribution of defaults and bayesian model validation. Technical report, Moody's/KMV. November.
Garthwaite, Paul H., Kadane, Joseph B., O, Hagan, Anthony, 2005. Statistical methods for eliciting probability distributions. Journal of the American Statistical Association 100, 680–701.
Jaynes, E.T., 2003. Probability Theory: the Logic of Science. Cambridge University Press, New York.
Kadane, Joseph B., Wolfson, Lara J., 1998. Experiences in elicitation. The Statistician 47 (1), 3–19.
Kadane, Joseph B., Dickey, James M., Winkler, Robert L., Smith, Wayne S., Peters, Stephen C., 1980. Interactive elicitation of opinion for a normal linear model. Journal of the American Statistical Association 75 (372), 845–854 Dec.
Kahneman, D., Tversky, A., 1974. Judgement under uncertainty: heuristics and biases. Science 185, 1124–1131.
Lindley, D.V., Tversky, A., Brown, R.V., 1979. On the reconciliation of probability assessments. Journal of the Royal Statistical Society. Series A (General) 142 (2), 146–180.
Lindley, D.V., 1982. The improvement of probability judgements. Journal of the Royal Statistical Society. Series A (General) 145 (1), 117–126.
OCC, 2006. Validation of credit rating and scoring models: a workshop for managers and practitioners. Office of the Comptroller of the Currency.
Pluto, Katja, Tasche, Dirk, 2005. Thinking positively. Risk 72–78 August.
Raiffa, Howard, Schlaifer, Robert, 1961. Applied Statistical Decision Theory. Harvard Business School.
Wilde, Tom, Jackson, Lee, 2006. Low-default portfolios without simulation. Risk 60–63 August.
Zellner, Arnold, 1996. Introduction to Bayesian Inference in Econometrics. John Wiley & Sons, New York.