Default Estimation and Expert Information

Nicholas M. Kiefer

Cornell University

December 3, 2008

Abstract

Default is a rare event, even in segments in the midrange of a bank's portfolio. Inference about default rates is essential for risk management and for compliance with the requirements of Basel II. Most commercial loans are in the middle-risk categories and are to unrated companies. Expert information is crucial in inference about defaults. A Bayesian approach is proposed and illustrated using a prior distribution assessed from an industry expert. The binomial model, most common in applications, is extended to allow correlated defaults. A check of robustness is illustrated with an $\epsilon-$ mixture of priors.

Keywords: Bayesian inference, robustness, correlated defaults, Basel II, risk management, prior assessment

1 Introduction

Estimation of default probabilities (PD), loss given default (LGD, a fraction) and exposure at default (EAD) for portfolio segments containing reasonably homogeneous assets is essential to prudent risk management. It is also crucial for compliance with Basel II (B2) rules for banks using the IRB approach to determine capital requirements (Basel Committee on Banking Supervision (2004)). Estimation of small probabilities is tricky, and this paper will focus on estimating PD. The emphasis is on segments in the middle of the risk profile of the portfolio. Although the risk is in the middle of the asset mix, the probability of default is still "small." It is in fact likely to be about 0.01; defaults, though seen, are rare. The bulk of a typical bank's commercial loans are concentrated in these segments (segments differ across banks). Very low risk institutions are relatively few in number and they have access to capital through many avenues in addition to commercial loans. Very high risk loans are largely avoided and when present are often due to the reclassification of a safer loan as conditions change. To put this in perspective, the middle-quality loans are approximately S&P Baa or Moody's BBB. In practice the bulk of these loans are to unrated companies and the bank has done its own rating to assign the loans to risk "buckets." The focus of this paper is on estimation of the default probability for such a risk bucket on the basis of historical information and expert knowledge.

Throughout the paper we take a probability approach to the quantitative description of uncertainty. There are many arguments that uncertainty is best described in terms of probabilities. The classic axiomatic treatment is Savage (1954). In the case of default modeling, where measuring and controlling risk is the aim, it is natural to focus on anticipating defaults, or at least anticipating the aggregate number of defaults. Suppose there are a number of default configurations, and we wish to assign numbers to these events and to use these numbers to describe the likelihood of the events. Simple arguments based on scoring rules (for example minimizing squared prediction error) or odds posting (avoiding certain losses) imply that these numbers must combine like probabilities. For constructions see De Finetti (1974). Lindley (1982b) elaborates on the development using scoring rules, Heath and Sudderth (1978) and Berger (1980) on betting. Reasoning about probabilities is not easy. There is a long literature beginning with Kahneman and Tversky (1974) demonstrating that people in practice find it difficult to think about probabilities consistently. Theoretical alternatives to probabilistic reasoning include possibility measures, plausibility measures, etc. These are reviewed and evaluated by Halpern (2003). Although these practical and theoretical objections to probability are often used to criticize the Bayesian approach, they apply equally to the likelihood specification and the modeling approach to risk management. While recognizing these objections, this paper will use the probability approach, noting that alternatives invariably lead to incoherence (De Finetti (1974)).

2 A Statistical Model for Defaults

The simplest and most common probability model for defaults of assets in a homogeneous segment of a portfolio is the Binomial, in which the defaults are independent across assets and over time, and defaults occur with common probability θ. This is the most widely used specification in practice and may be consistent with B2 requirements calling for a long-run average default probability. Note that specification of this model requires expert judgement, that is, information. Denote the expert information by e. The role of expert judgement is not usually explicitly indicated at this stage, so it is worthwhile to point to its contribution. First, consider the statistical model. The independent Bernoulli model is not the only possibility.

Certainly independence is a strong assumption and would have to be considered carefully. External factors not explicitly modeled, for example general economic conditions, could induce correlation. There is evidence that default probabilities vary over the business cycle (for example Nickell, Perraudin, and Varotto (2000)); we return to this topic below. The Basel prescription is for a marginal annual default probability, however, and correlation among defaults is accommodated separately in the formula for the capital requirement. Thus, many discussions of the inference issue have focussed on the binomial model and the associated frequency estimator. Second, are the observations really identically distributed? Perhaps the default probabilities differ across assets, even within the group. Can this be modeled, perhaps on the basis of asset characteristics? The 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. For specificity we will continue with the Binomial specification, but will turn attention below to an extension allowing correlated defaults (also possibly consistent with B2) below. Let d_i indicate whether the ith observation was a default ($d_i = 1$) or not ($d_i = 0$). The Bernoulli model (a single Binomial trial) for the distribution of $d_i$is $p(d_i|\theta, e) = \theta^{d_i}(1-\theta)^{1-d_i}$. Let $\mathbf{D} = \{d_i, i = 1, ..., n\}$ denote the whole data set and $r = r(\mathbf{D}) = \sum_i d_i$ the count of defaults. Then the joint distribution of the data is

$$\begin{aligned} p(\mathbf{D}|\theta, e) &= \prod_{i=1}^{n} \theta^{d_i}(1-\theta)^{1-d_i} \qquad (1) \\ &= \theta^r(1-\theta)^{n-r} \end{aligned}$$

As a function of θ for given data $\mathbf{D}$ this is the likelihood function $L(\theta|\mathbf{D}, e)$. Since this distribution depends on the data $\mathbf{D}$ only through r (n is regarded as fixed), the sufficiency principle implies that we can concentrate attention on the distribution of r

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

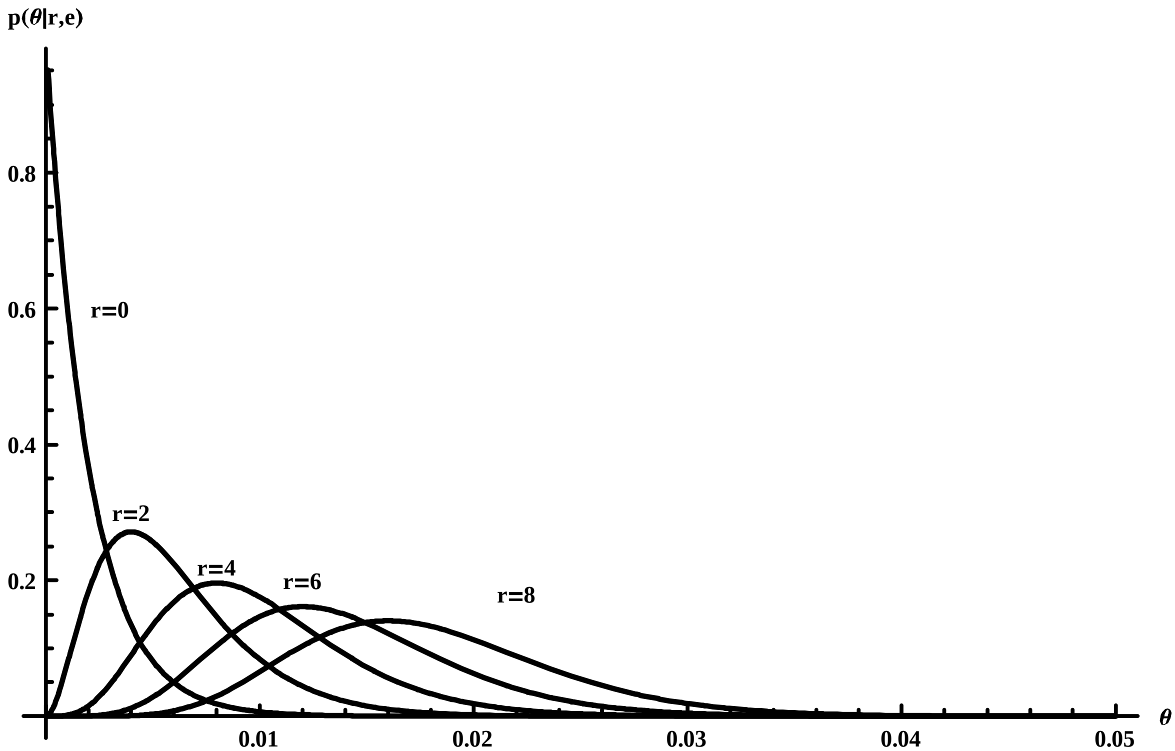
a Binomial(n,θ) distribution. This is so well known that it is easy to underappreciate the simplification obtained by passing from 1 to 2. Instead of separate treatment for each of the 2^n datasets possible, it is sufficient to restrict attention to n+1 data set types, characterized by the value of r. This theory of types can be made the basis of a theory of asymptotic inference. See Cover and Thomas (1991). In our application, the set of likely values of r is small, and an analysis can be done for each of these values of r, rather than for the $\binom{n}{r}$ distinct datasets corresponding to each value of r. Thus, by analyzing a few likely data set types, we analyze in effect all of the most likely data realizations.

Regarded as a function of θ for fixed r, 2 is the likelihood function. Figure 1 shows the likelihood functions for n=500, a reference data set size, and r={0,2,4,6,8}.

3 Uncertain Default Probabilities

Equation 2 is a statistical model. It generates probabilities for all default configurations as a function of a single parameter θ which remains unspecified. The default probability θ is an unknown, but that doesn't mean that nothing is known about its value. In fact, defaults are widely studied and risk managers, modelers, validators, and supervisors have detailed knowledge on values of θ for particular portfolio segments. The point is that θ is unknown in the same sense that the future default status of a particular asset is unknown. The fact that default is in the future is not important; the key is that it is unknown and the uncertainty can be described and quantified. We have seen how uncertain defaults can be modeled. The same

Figure 1: Likelihood Functions n=500



methods can be used to model the uncertainty about θ. Continuing with the logic used to model default uncertainty, we see that uncertainty about values of θ are coherently described by probabilities. We assemble these probability assessments into a distribution describing the uncertainty about θ given the expert information e, $p(\theta|e)$.

The distribution $p(\theta|e)$ can be a quite general specification, reflecting in general the assessments of uncertainty in an infinity of possible events. This is in contrast with the case of default configurations, in which there are only a finite (though usually large) number of possible default configurations. However, this should not present an insurmountable problem. Note that we are quite willing to model the large number of probabilities associated with the possible different default configurations with a simple statistical model - the binomial. The same approach is taken for the prior specification. That is, we can fit a few probability assessments by an expert to a suitable functional form and use that distribution to model prior uncertainty. There is some approximation involved, and care is necessary. In this regard, the situation is no different from that present in likelihood specification.

A convenient functional form is the beta distribution

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

which has mean $\alpha/(\alpha+\beta)$ and variance $\alpha\beta/((\alpha+\beta)^2(1+\alpha+\beta))$. A particularly easy generalization is to specify the support $\theta \in [a,b] \subset [0,1]$. 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 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. For example it may not be flexible enough to allow combination of information from many experts. A simple generalization is the 7-parameter mixture of two 4-parameter Betas with common support. Computations with this mixture distribution are not substantially more complicated than computations with the 4-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. A useful further generalization is given by the 9-parameter mixture allowing different supports for the two mixture components. By choosing enough

Beta-mixture terms the approximation of an arbitrary continuous prior $p(\theta|e)$ for a Bernoulli parameter can be made arbitrarily accurate, in the sense that the sequence of approximations can be chosen to converge uniformly to $p(\theta|e)$. Note that there is nothing stochastic in this argument. The proof follows the proof of the Stone-Weierstrass approximation theorem for approximation of continuous functions by polynomials. See Diaconis and Ylvisaker (1985). The use of mixtures in simulating posterior densities is advocated by Hoogerheide, Kaashoek, and van Dijk (2007).

4 Inference

With the likelihood and prior at hand inference is a straightforward application of Bayes rule. Given the distribution $p(\theta|e)$, we obtain the joint distribution of r, the number of defaults, and θ :

$$p(r,\theta|e) = p(r|\theta,e)p(\theta|e)$$

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

$$p(r|e) = \int p(r,\theta|e)d\theta \tag{5}$$

If the value of the parameter θ is of main interest we divide to obtain the conditional (posterior) distribution of θ :

$$p(\theta|r,e) = p(r|\theta,e)p(\theta|e)/p(r|e) \tag{6}$$

which is Bayes rule. Since Basel II places more emphasis on the default probability than on the number of defaults in a given portfolio segment, we focus our discussion on $p(\theta|r,e)$. Note as an aside that if defaults are a sample from an infinite exchangeable sequence then the marginal distribution of the number of defaults can always be written as a binomial mixture, so the parametric specification may not be as restrictive as it seems. See De Finetti (1974).

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 that might be in the middle of a bank's portfolio. These are typically commercial loans, mostly to unrated companies. If rated, these might be about S&P Baa or Moody's BBB. The method included a specification of the problem and some specific questions followed by a discussion. General discussions of the elicitation of prior distributions are given by Kadane, Dickey, Winkler, Smith, and Peters (1980), Garthwaite, Kadane, and O'Hagan (2005) 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. In the case of a Bernoulli parameter and a 2-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 ("dropoffs"). Graphical feedback is provided for refinement of the specification. Examples consider n=20. Gavasakar (1988) suggests an alternative method, based on assessing modes of predictive distributions but not on dropoffs. Instead, changes in the mode in response to hypothetical samples are elicited and an explicit model of elicitation errors is proposed. The method is evaluated in the n=20 case and appears competitive. 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. This approach may be less attractive in the case of large sample sizes and small probabilities, and in our particular application, where the expert was sophisticated about probabilities. Our expert found it easier to think in terms of the probabilities directly than in terms of defaults in a hypothetical sample.

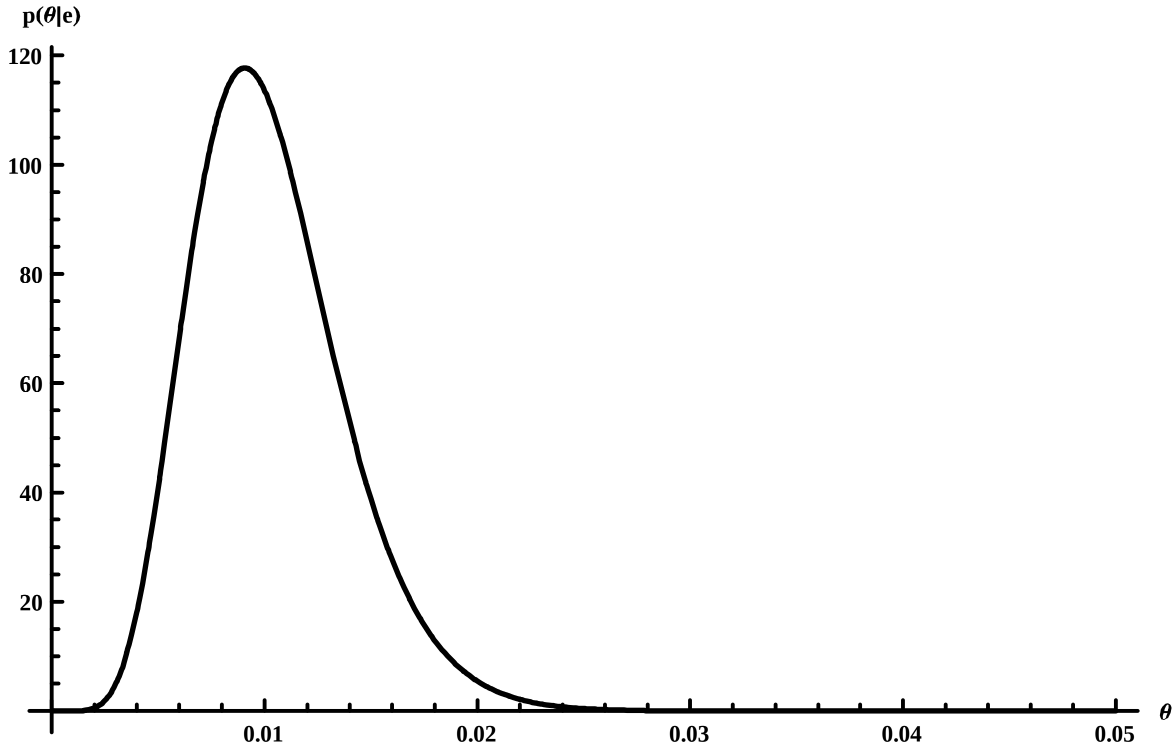
Thinking about uncertainty in terms of probabilities requires effort and practice (possibly explaining why it is so rarely done). Nevertheless it can be done once experts are convinced it is worthwhile. Indeed, there is experimental evidence in game settings that elicited beliefs about oponents' future actions are better explanations of responses than empirical beliefs - Cournot or fictitious play - based on weighted averages of previous actions. For details see Nyarko and Schotter (2002).

The precise definition of default is at issue. In the economic theory of the firm, default occurs when debt payments are missed and ownership and control of the firm passes from existing owners (shareholders in the case of a corporation) to debtholders. As a lesser criterion, loans that are assigned to "nonaccrual" may be considered defaulted. We simply note the importance of using consistent definitions in the assessment of expert information and in data definition.

We did the elicitation first assuming a sample of 500 asset-years. For our application, we also considered a "small" sample of 100 observations and a "large" sample of 1000 observations, and occasionally an enormous sample of 10000 observations. Considering first the predictive distribution on 500 observations, the modal value was five defaults. Upon being asked to consider the relative probabilities of five or four defaults, conditional on four or five defaults occurring (the conditioning does not matter here, for the probability ratio, but it is thought to be easier to think about when posed in this fashion), the expert expressed some trepidation as it is difficult to think about such rare events. Ultimately, the expert gave probability ratios not achievable by the binomial model even with known probability. This experience supports the implication of Gavasakar (1988) that dropoff probabilities are problematic. 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 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 he did not want to rule out the possibility that the rates were much higher than anticipated (prudence?). Quartiles were assessed by asking the expert to consider the value at which larger or smaller values would be equiprobable given the value was less than the median, then given the value was more than the median. The median value was 0.01. The former was 0.0075. The latter, the .75 quartile, was assessed at .0125. The expert seemed to be thinking in terms of a normal distribution, perhaps using informally a central limit theorem combined with long experience with this category of assets.

This set of answers is more than enough information to determine a 4-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 $\{a, b\} = [0.0001, 0.3]$. These were allowed to vary (the parameter set is overdetermined) but the optimization routine did not change them beyond the 7th decimal

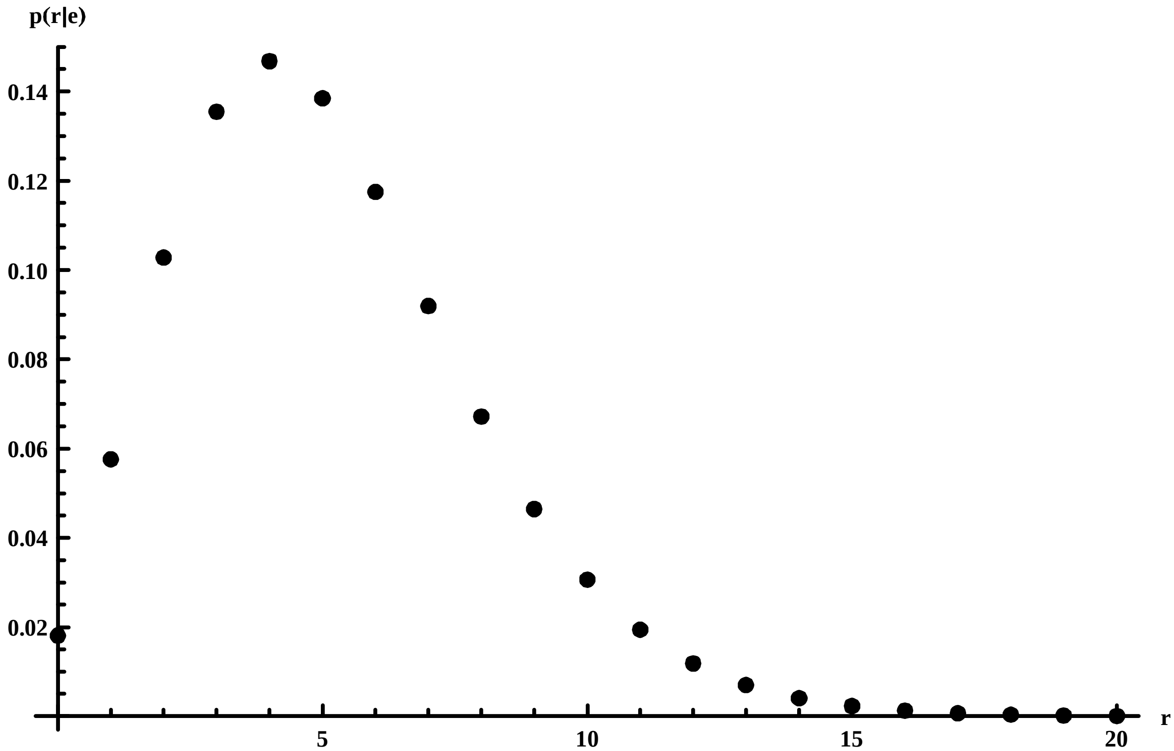
Figure 2: Expert Information : Closeup



place. Thus, the expert was able to determine these parameter values consistently with his probability assessments. Further, changing the weights did not matter much either. Probably this is due to the fact that there is almost no probability in the upper tail, so changing the upper bound made almost no difference in the assessed probabilities. Thus 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 (7.9, 224.8) as best describing the expert's information. The resulting prior distribution $p(\theta|e)$ has virtually no probability on the long right tail. A closup view of the relevant part of the prior is graphed in Figure 2.

The median of this distribution is 0.00988, the mean is 0.0103 and the standard deviation is 0.00355. In practice, after the information is aggregated into an esti-

Figure 3: Predictive Distribution p(r|e)



mated probability distribution, then additional properties of the distribution would be calculated and the expert would be consulted again to see if any changes were in order before proceeding to data analysis Lindley (1982a). This process would be repeated as necessary. In the present application there was one round of feedback, valuable since the expert had had time to consider the probabilities involved. The characteristics reported are from the second round of elicitation. An application within a bank might do additional rounds with the expert, or consider alternative experts and a combined prior.

The predictive distribution 5 corresponding to this prior is given in Figure 3 for n=500.

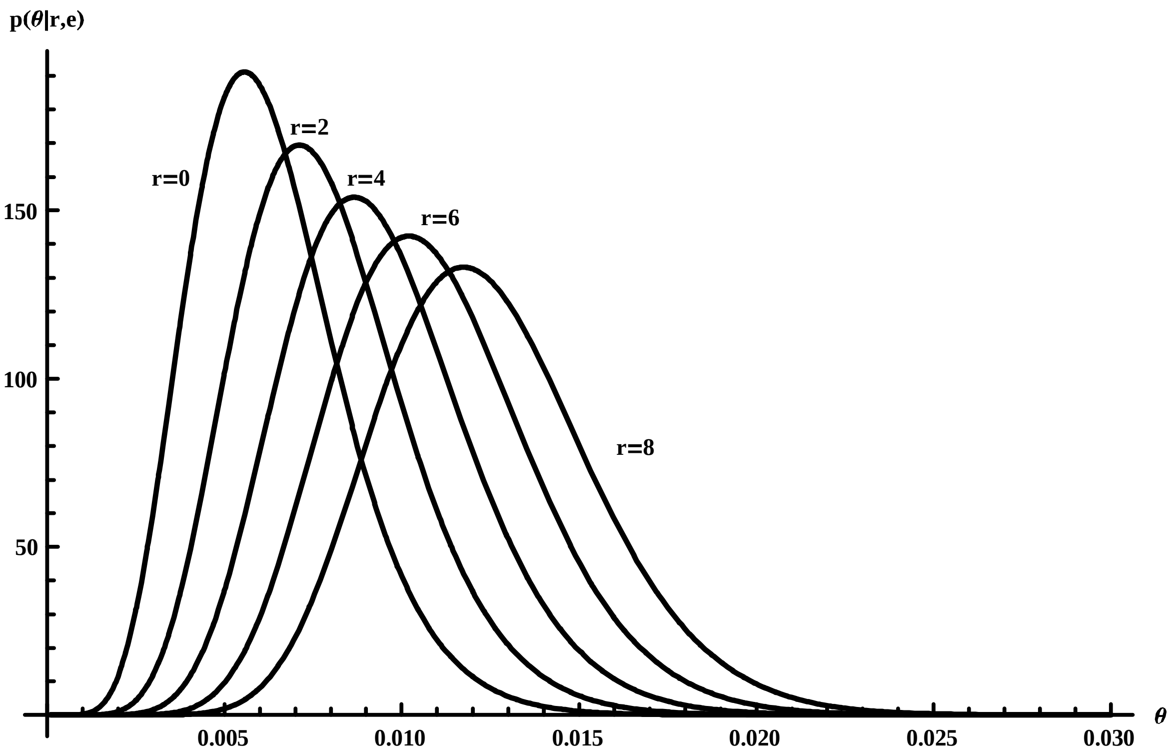
With our specification, the expected value of $r, E(r|e) =_{k=0}^{n} kp(k|e)$ is 5.1 for n=500. Total defaults numbering 0-9 characterize 92% of expected data sets. Thus, carrying out our analysis for these 10 data types, comprising about 2^{62} distinct datasets, a trivial fraction of the 2^{500} possible datasets, actually covers 92% of the expected realizations. Defaults are expected to be rare events. Thus we are not analyzing one particular dataset, rather we provide results applicable to 92% of the likely datasets.

6 Posterior Analysis

The posterior distribution, $p(\theta|r,e)$, is graphed in Figure 4 for r = 0, 2, 4, 6, and 8 and n=500. The corresponding likelihood functions, for comparison, were given in Figure 1. Note the substantial differences in location. Comparison with the likelihood functions graphed in Figure 1 and the prior distribution graphed in Figure 2 reveals that the expert provides much more information to the analysis than do the data.

Given the distribution $p(\theta|r,e)$, we might ask for a summary statistic, a suitable estimator for plugging into the required capital formulas as envisioned by Basel Committee on Banking Supervision (2004). A natural value to use is the posterior expectation, $\overline{\theta} = E(\theta|r,e)$.The expectation is an optimal estimator under quadratic loss and is asymptotically an optimal estimator under bowl-shaped loss functions (use the Bernstein - von Mises theorem on convergence of the posterior to a normal distribution and Anderson's lemma on minimum loss, van der Vaart (1998), Theorem 10.1 and Lemma 8.5). An alternative, by analogy with the maximum likelihood estimator $\widehat{\theta}$, is the posterior mode θ_m As a summary measure of our confidence we would use the posterior standard deviation $\sigma_\theta = \sqrt{E(\theta - \overline{\theta})^2}$. By comparison, the usual approximation to the standard deviation of the maximum likelihood estimator is $\sigma_{\widehat{\theta}} = \sqrt{\widehat{\theta}(1-\widehat{\theta})/n}$.These quantities are given in Table 1 for r=0-9 and r=20, 50,

Figure 4: Posterior Densities $p(\theta|r,e)$



100, 200. As noted, the r=0-9 case covers the 2^{62} most likely datasets out of the possible 2^{500}. Together, these comprise analyses of 92% of likely datasets. The r=20 case is an extremely low probability outcome - less than 0.0001 - and is included to show the results in this case. There are approximately 2^{118} datasets corresponding to r=20. The rows for r=50, 100, and 200 are included as a further "stress test" and will be discussed below. Their combined prior probability of occurrence is less than 10^{-14}.

Table 1: Default Probabilities - Location and Precision, n=500

r	$\overline{\theta}$	θ_m	$\widehat{\theta}$	σ_θ	$\sigma_{\widehat{\theta}}$
0	0.0063	0.0081	0.000	0.0022	0 (!).
1	0.0071	0.0092	0.002	0.0023	0.0020
2	0.0079	0.0103	0.004	0.0025	0.0028
3	0.0086	0.0114	0.006	0.0026	0.0035
4	0.0094	0.0125	0.008	0.0027	0.0040
5	0.0102	0.0136	0.010	0.0028	0.0044
6	0.0109	0.0147	0.012	0.0029	0.0049
7	0.0117	0.0158	0.014	0.0030	0.0053
8	0.0125	0.0169	0.016	0.0031	0.0056
9	0.0132	0.0180	0.018	0.0032	0.0060
20	0.0215	0.0296	0.040	0.0040	0.0088
50	0.0431	0.0425	0.100	0.0053	0.0134
100	0.0753	0.0749	0.200	0.0065	0.0179
200	0.1267	0.1266	0.400	0.0069	0.0219

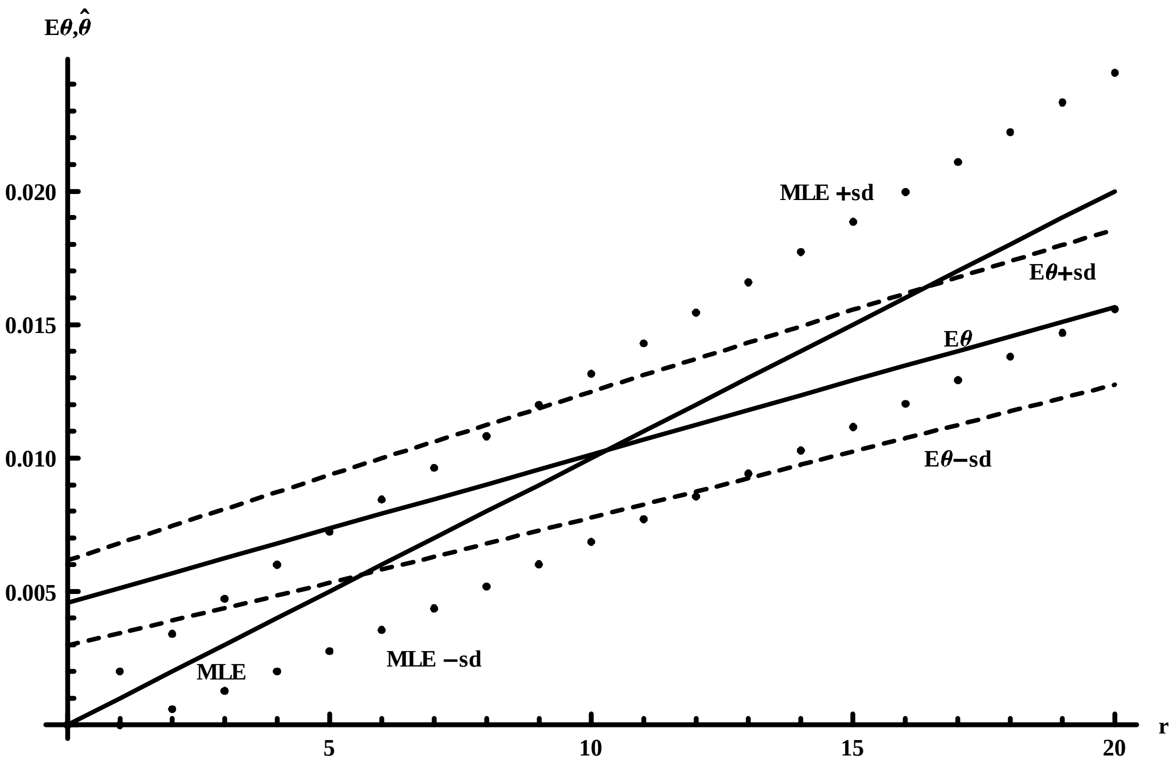
For values of r below its expected value the posterior mean is greater than the MLE, for values above the posterior is less than the MLE, as expected. As is well-known and widely discussed, the MLE is unsatisfactory when there are no observed defaults (Basel Committee on Banking Supervision (2005), Pluto and Tasche (2005), BBA, LIBA, and ISDA (2005), Kiefer (2008)) The Bayesian approach provides a coherent resolution of the inference problem without resort to desperation (sudden reclassification of defaulted assets, technical gimmicks).

Expert information will have larger weight in smaller sample sizes, and smaller relative weight for larger sample sizes. For n=1000, for example, r=5-15 reflects 76% of the most likely datasets; r=0-20 represents 97%. To put this in perspective, the cases r=0-20 correspond to approximately 2^{138} datasets out of a possible 2^{1000}. Thus, 97% of the likely observations are contained in the small fraction 2^{-862} of the possible datasets, or 0.0021 of the possible types. A substantial simplification results from concentrating on the distribution of the sufficient statistic and use of expert judgement to characterize possible samples. Naturally, this simplification depends critically on the use of expert judgement in specification of the likelihood function (our choice admits a sufficient statistic) and in specification of the prior distribution. Rather than resorting to extensive tabulation, we report results for 97% of likely samples in Figure 5. The error bands, dotted for the MLE and dashed for the prior mean, are plus/minus one standard deviation.

7 Robustness - the cautious Bayesian

Suppose we are rather less sure of our expert than he is of the default probability. Or, more politely, how can we assess just how important the tightly-held views of the expert are in determining our estimates? Table 1 gives one answer by comparing the MLE and the posterior location measures. Another answer was proposed by Kiefer (2008) , who considered a less-certain expert with a prior with the same location but substantially higher variance than the actual expert. An alternative

Figure 5: $E\theta$ and MLE $\hat{\theta}$ for n=1000



approach, more formal and based on the literature on Bayesian robustness (Berger and Berliner (1986)) is to mix the actual expert's prior with an alternative prior, and see exactly how seriously the inferences are affected by changes in the mixing parameter. Berger and Berliner (1986) in fact suggested mixing in a class of distributions, corresponding to different amounts or directions of uncertainty in the prior elicitation. In this spirit, we will mix the expert's 4-parameter beta distribution with a uniform distribution. Here, there are two clear possibilities. One is to mix with the uniform on [a,b], accepting the expert's bounds but examining robustness to alpha and beta. The second is to mix with the uniform on [0,1], allowing all theoretically feasible values of θ. We choose the latter approach. This is not a completely comfortable approach. Although the uniform is commonly interpreted as an uninformative prior, it in fact has a mean of 1/2, not a likely value for our default probability by any reasonable prior. An alternative might be to mix with a prior with the same mean as our expert's distribution, but maximum variance. We do not pursue this here. Our results suggest that it would not make much difference; the key is to mix in a distribution with full support, so that likelihood surprises can appear. We choose to mix the expert's prior with a uniform on all of [0,1]. This allows input from the likelihood if the likelihood happens to be concentrated above b (or below a). The mixture distribution is

$$p(\theta|e,\epsilon) = (1-\epsilon)p(\theta|\alpha,\beta,a,b)I(\theta \in [a,b]) + \epsilon \tag{7}$$

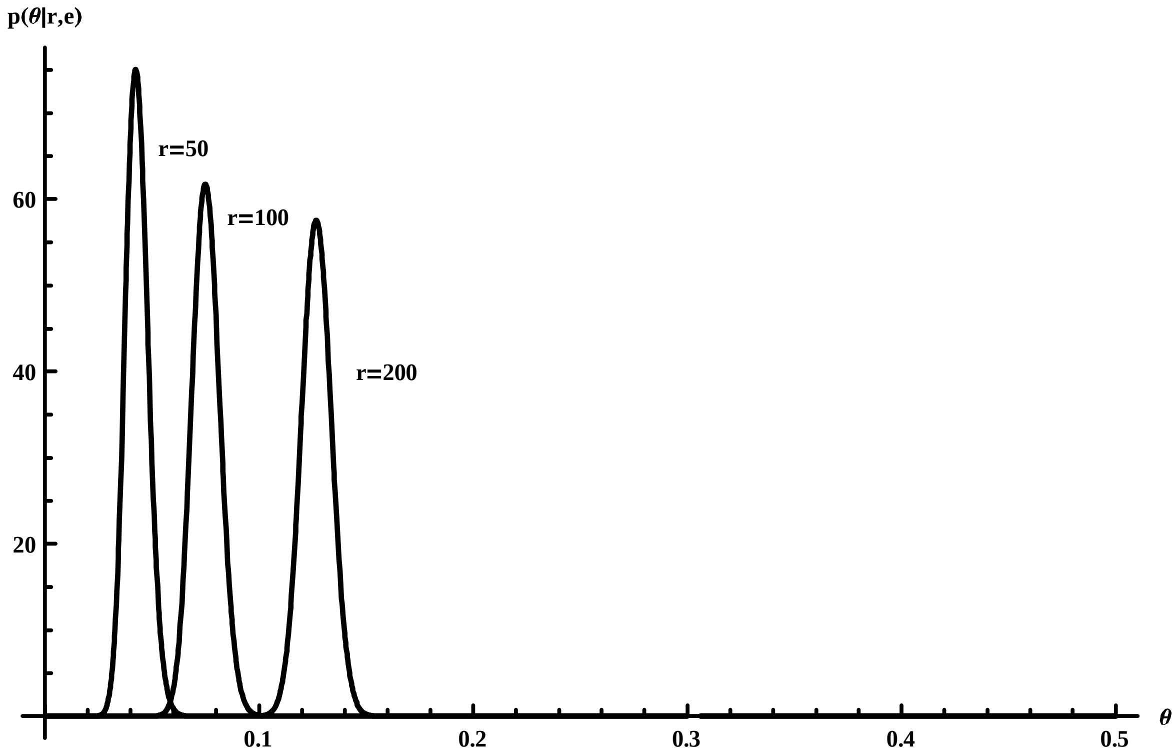
for $\theta \in [0,1]$. The approach can be used whatever prior is specified, not just the 4-parameter beta. Our robust prior is in the 9-parameter mixture family consisting of our expert's 4-parameter beta mixed with the 4-parameter beta with parameters $\{\alpha,\beta,a,b\} = \{1,1,0,1\}$and mixing parameter ϵ. Table 2 shows the posterior means for the mixture priors for $\epsilon = \{0.01, 0.1, \ 0.2, 0.3, 0.4\}$.

Table 2: Robustness - Posterior means for mixture priors, n=500

r	$\overline{\theta}; \epsilon = .01$	$\overline{\theta}; \epsilon = .1$	$\overline{\theta}; \epsilon = .2$	$\overline{\theta}; \epsilon = .3$	$\overline{\theta}; \epsilon = .4$
0	0.0063	0.0063	0.0062	0.0061	0.0061
1	0.0071	0.0071	0.0071	0.0071	0.0070
2	0.0079	0.079	0.0079	0.0079	0.0078
3	0.0086	0.0086	0.0086	0.0086	0.0086
4	0.0094	0.0094	0.0094	0.0094	0.0094
5	0.0102	0.0102	0.0102	0.0102	0.0102
6	0.0109	0.0109	0.0110	0.0110	0.0110
7	0.0117	0.0117	0.0117	0.0118	0.0118
8	0.0125	0.0125	0.0125	0.0125	0.0126
9	0.0132	0.0133	0.0133	0.0134	0.0134
20	0.0358	0.0358	0.0386	0.0398	0.0405
50	0.1016	0.1016	0.1016	0.1016	0.1016
100	0.2012	0.2012	0.2012	0.2012	0.2012
200	0.4004	0.4004	0.4004	0.4004	0.4004

Mixing the expert's prior with the uniform prior makes essentially no difference to the posterior mean for data in the likely part of the set of potential samples. For r=20, unlikely but not outrageous, using the robust prior makes a substantial difference. For the extremely unlikely values, 50, 100, 200, the differences are dramatic. The actual value of ϵ makes almost no difference. The numbers for $\epsilon = 0.001$, not shown in the table, give virtually the same mean for all r. For comparison, we recall the values of $\overline{\theta}$ for r={20,50,100,200} from Table 1. These are {0.0215,0.0431,0.0753,0.1267}. Figure 6 shows the posterior distributions for our expert's prior, $p(\theta|r, e)$ for r=50, 100, and 200. It is clear that the prior plays a huge role here, as the likelihood mass is concentrated near .1, .2 and .4, while the

Figure 6: Posteriors with ϵ=0, n=500

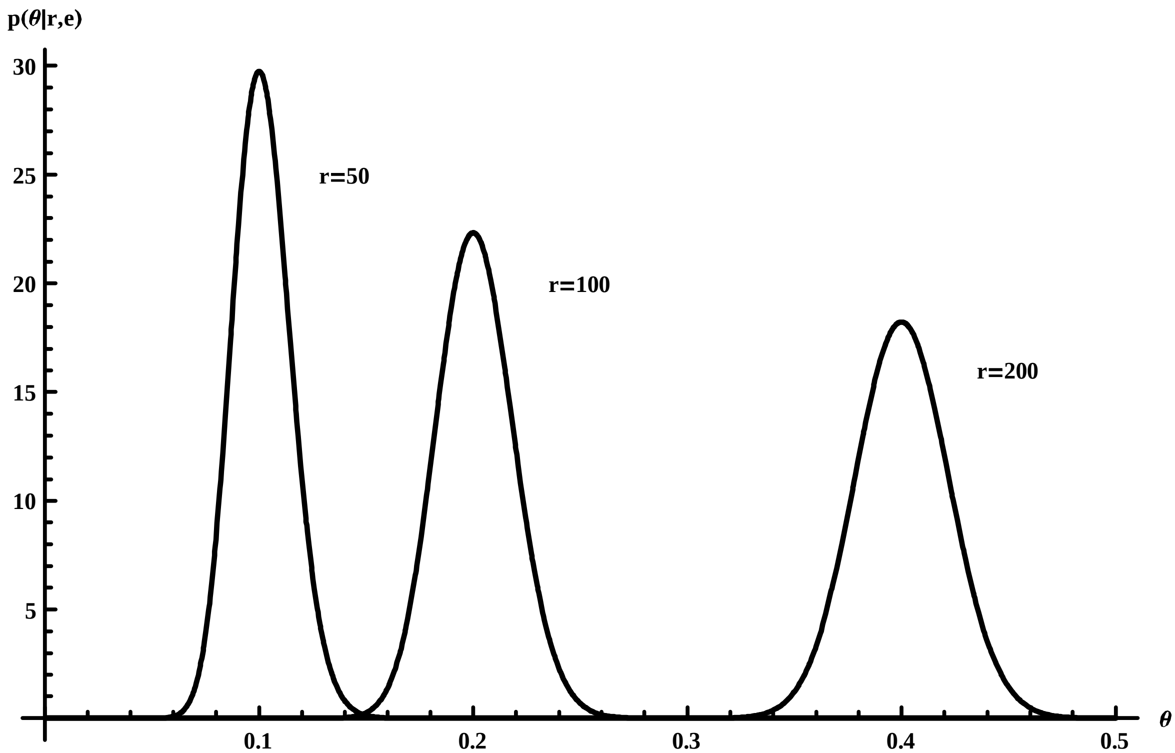


prior gives only trivial weight to values greater than about .03, see Figures 1 and 2. On the other hand, Figure 7 shows the posterior corresponding to 7 with 1% mixing ($\epsilon = 0.01$). Here, the likelihood dominates, as the likelihood value near the expert's prior is vanishingly small relative to the likelihood in the tail area of the mixing prior.

Thus, the robust analysis with even a very small nonzero mixing fraction can reveal disagreements between the data and the expert opinion which are perhaps masked by the formal analysis. This robust analysis may have a role to play in the validation phase.

In what sense is the robust analysis useful? We are really bringing something outside the model, namely the uniform distribution representing no one's beliefs, into the analysis as a formal tool for diagnostic analysis. The spirit is the same as

Figure 7: Posteriors with ϵ=0.01, n=500



usual procedures associated with good statistical practice - residual analysis, out of sample fits, forecast monitoring, or comparison with alternative models. All of these procedures involve stepping away from the specified model and its analysis, and asking, post estimation, does the specification make sense? Post-estimation model evaluation techniques are often informal, sometimes problem specific, and require sound statistical judgement OCC (2006). The analysis of robustness via an artificial prior is an attempt to merge the formal analysis with the informal post-estimation model checking. A related method, checking for irrelevant data using a mixture distribution, is proposed by Ritov (1985) and this might have a role as well.

8 Heterogeneity

It is clearly contemplated in the Basel II guidance that heterogeneity is mitigated by the classification of assets into homogeneous groups before estimation of the group-specific default probability. However, there may be remaining heterogeneity, due to systematic temporal changes in asset characteristics or to changing macroeconomic conditions. For fairly low-default portfolios there is unlikely to be enough data to sort out differences between years. However, there is evidence from other markets that default probabilities vary over the cycle, see for example Nickell, Perraudin, and Varotto (2000). The B2 capital requirements are based on a one-factor model due to Gordy (2003) that acommodates systematic temporal variation in asset values and hence in default probabilities. This model can be used as the basis of a model that allows temporal variation in the default probabilities, and hence correlated defaults within years. The value of the ith asset in time t is

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

where ϵ_{it} is the time and asset specific shock and x_t is the common time shock, inducing correlation ρ across asset values within a period. The random variables are assumed to be standard normal and independent. A mean of zero is attainable through translation without loss of generality since we are only interested in default probabilities. Suppose default occurs if $v_{it} < d$, a default threshold value. The overall or marginal default rate we are interested in is $\theta = \Phi(d)$. However, in each period the default rate depends on the realization of the systematic factor x_t; denote this θ_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 is

$$\theta_t = \Phi[(d-\rho^{1/2}x_t)/(1-\rho)^{1/2}]$$

Thus for $\rho \neq 0$ there is random variation in the default probability over time and this induces correlation in defaults across assets within a period. The distribution is given by

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

using the standard normal distribution of x_t and $\theta = \Phi(d)$. Differentiating gives the density $p(\theta_t|\theta, \rho, e)$, where the role of expert judgement e is quite clear. The parameters are θ, the marginal or mean default probability, for which we have already assessed a prior, and the asset correlation ρ. Values for ρ are in fact prescribed in Basel Committee on Banking Supervision (2006) for different asset classes. First, we defer to the B2 formulas and set ρ to a prescribed value (0.20) and drop ρ from the notation. Then, we consider the effect of generalizing this specification. Moving toward the posterior for θ in this new specification we note that the conditional distribution of the number of defaults in each period is (from 2)

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

from which we obtain the distribution conditional on the parameter of interest

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

and the distribution for $\mathbf{R} = (r_1, ...r_T)$ is

$$p(\mathbf{R}|\theta, e) = \prod_{t=1}^{T} p(r_t|\theta, e) \tag{8}$$

where we condition on $(n_1, ..., n_T)$. Regarded as a function of θ for fixed $\mathbf{R}$, 8 is the likelihood function. With this in hand we can calculate the posterior distribution using 5 and 6

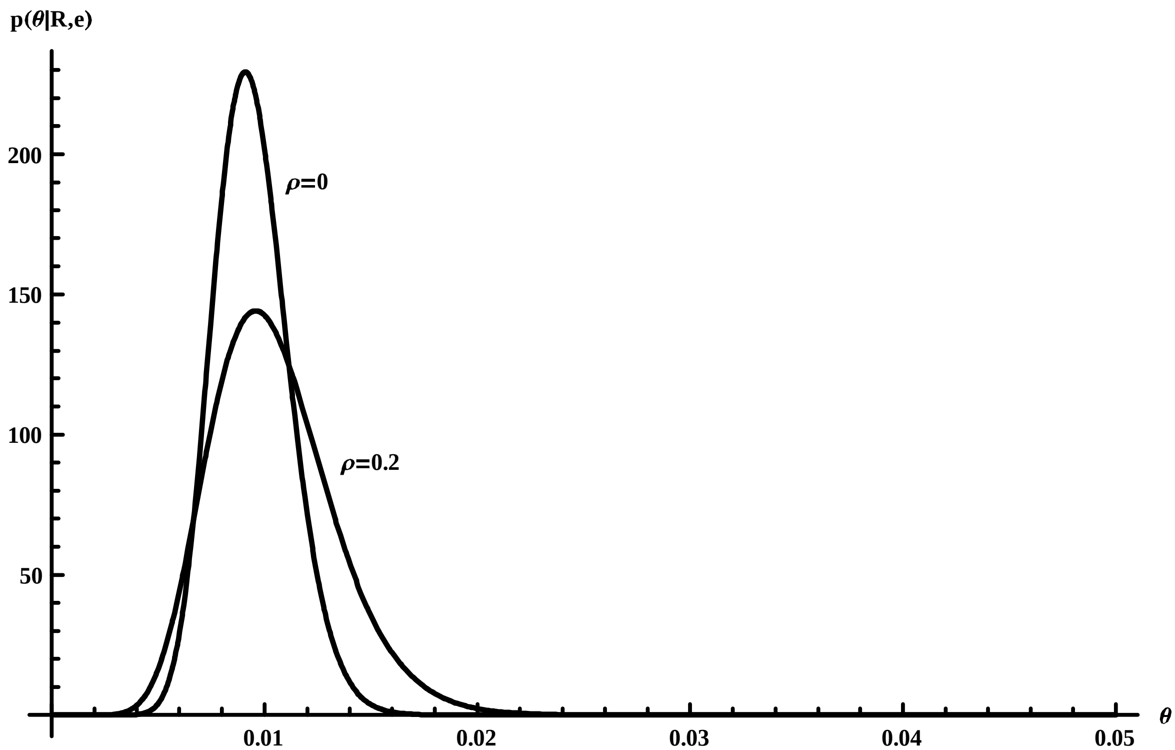
$$p(\theta|\mathbf{R}, e) = p(\mathbf{R}|\theta, e)p(\theta|e)/p(\mathbf{R}|e)$$

where $p(\mathbf{R}|e) = \int p(\mathbf{R}|\theta, e)p(\theta|e)$ is the predictive distribution, the marginal distribution of the data.

In order to illustrate these calculations we need data which have defaults separately by year. Thus, there is no longer a single integer sufficient statistic - although a vector of integers is not substantially more complicated. To fix ideas, we use a hypothetical bucket of mid-portfolio corporate bonds of S&P-rated firms in the KMV North American Non-Financial Dataset. Default rates were computed for cohorts of firms starting in September 1993 and running through September 2004. In total there are 2197 asset/years of data and 20 defaults, for an overall empirical rate of 0.00913. The posterior expectation is $E\theta = 0.0104$ and the posterior standard deviation $\sigma_\theta = 0.0029$. The corresponding statistics for the Binomial model without heterogeneity are $E\theta = 0.0094$ and $\sigma_\theta = 0.0018$. Figure 8 gives the posterior densities for θ with and without heterogeneity. Allowing for heterogeneity clearly decreases the data information and therefore increases the posterior uncertainty about the long-run default probability θ.

Our specification with the correlation at the value specified in the international guidance shows that the data are considerably less informative about the default probability in the generalized model, even though no new parameters are introduced

Figure 8: Posterior with Heterogeneity



- i.e., the reduction in data information is not simply a degrees of freedom effect. We next consider the model with ρ unknown but with a distribution centered on the B2 value. This will give an indication of the sensivity of the results to the regulatory specification, and will indicate to some extent the data implications for ρ. We consider two priors for ρ, a Beta distribution with standard deviation 0.05 and a Beta with standard deviation 0.32. These, together with the result for ρ fixed at its mean value provide coverage of a range of precisions all centered on the B2 value (0.20). In both cases we specify this prior independently of the prior on the default probability. Of course, this does not imply that the posterior has these quantities independent and indeed we find positive posterior correlation. The two Beta distributions have parameters (12.6, 50.4) and (0.12, 0.48) respectively.

For the intermediate specification with the standard deviation of ρ equal to 0.05, the posterior moments are $E\theta = 0.0103$, $E\rho = 0.187$, $\sigma_\theta = 0.0028$, $\sigma_p = 0.0455$, and $\rho_{\theta\rho}$ (the posterior correlation between the default probability and the asset value correlation) = 0.119. Thus allowing uncertainty about the value of the asset correlation does not substantially change the result for the parameter of interest, θ.

For the less informative specification with the standard deviation of ρ equal to 0.32, the posterior moments are $E\theta = 0.0098$, $E\rho = 0.120$, $\sigma_\theta = 0.0025$, $\sigma_p = 0.0804$, and $\rho_{\theta\rho} = 0.213$. Here too, with very weak prior information about the asset correlation, the inference on the interest parameter θ is essentially unchanged. The posterior expectation of ρ is reduced, though there is not much posterior information about this parameter, as might be expected with a fairly rich model and a dataset with few defaults. Note that the posterior expectation is within one standard deviation of the posterior expectation from the intermediate model, and in fact is within one standard deviation of the value prescribed by B2.

We estimated this model using direct numerical integration (with Mathematica$^{\text{TM}}$

6.0). Estimation of much more complicated models is now in principle straightforward using Markov Chain Monte Carlo (MCMC) and related procedures (see Robert and Casella (2004) and Geweke (2005)). These techniques may be useful in the context of validation (essentially specification analysis) procedures that banks are expected to employ. See OCC (2006).

9 Conclusion

I have considered inference about the default probability for a midrange portfolio segment on the basis of data information and expert judgement. Examples focus on the sample size of 500; results are also presented for the 1000 observations, some portfolios in this risk range are much larger. These analyses are relevant to hypothetical portfolios of middle-risk commercial loans. These are predominantly to unrated companies; if rated these would be approximately S&P Baa or Moody's BBB. I have also represented the judgement of an expert in the form of a probability distribution, for combination with the likelihood function. The expert is a practitioner experienced in risk management in well-run banks. The 4-parameter Beta distribution seems to reflect expert opinion fairly well. Errors, which would be corrected through additional feedback and respecification in practice, are likely to introduce more certainty into the distribution rather than less. We consider the possible likely realizations of the sufficient statistic for the specified statistical model. In the default case, the number of realizations is linear in the sample size (while the number of potential distinct samples is exponential). Using the expert information, it is possible to isolate the most likely realizations. In the sample of 500, five defaults are expected. In this case, our analysis of 0 through 9 defaults covers 92% of expected datasets. Our analyses of samples of 1000 covered 97% of the likely realizations. The Binomial model is not the only possible model. The one-factor model underlying the B2 capital calculations allows temporal variation in default rates, inducing correlation among defaults within periods. This highly structural model leads naturally to a generalization of the Binomial model. The Bayesian analysis remains straightforward and appropriate. An application to a bucket of mid-portfolio corporate bonds is used to illustrate the effects of generalizing the Binomial model to allow heterogeneity and correlated defaults.

At the validation stage, modelers can be expected to have to justify the likelihood specification and the representation of expert information. Analysis of the sensitivity of the results to the prior should be a part of this validation procedure. We propose using a mixture of the expert's prior and an alternative, less informative prior. In our case, we mix the prior with a uniform distribution on the unit interval. While it is not likely that the uniform describes any expert's opinion on the default probability, mixing in the uniform allows unexpected disagreement between the prior and the data to appear vividly. An example shows that even a trivially small weight on the alternative will do. Of course, within the context of the model, the decision based on the expert's posterior is correct. A broader view might suggest something wrong with the specification - of either the likelihood or the prior. Perhaps these do not refer to the same risk class, or perhaps the default definitions are inconsistent. The situation is not unlike that arising in ordinary validation exercises in which the model is evaluated in terms of residual analysis or out-of-sample fits. These involve considerations which are relevant but which are outside the formal model. As a result there are a number of different methods in use, corresponding to different ways in which models can fail, and expert judgement remains crucial in this less formal context as well as in the formal specification of the likelihood and the prior. For further discussion, see OCC (2006). Finally, the use of the posterior distribution solely to obtain estimators to insert into the B2 capital formulas is somewhat crude. Risk managers at the bank level might use the full posterior distributions in risk management decisions, calculating for example implied distributions of losses and probabilities of different levels of losses taking into account not only uncertainty about defaults but uncertainty about default probabilities. Supervisors and risk management officials might consider the distribution of minimum required capital and choose standards accordingly using a loss function. These possibilities seem not yet to be at hand. The use of a loss function in determining capital is suggested in Kiefer and Larson (2004).

10 Acknowledgments

Affiliations: Cornell University, Departments of Economics and Statistical Science and Center for Analytic Economics, 490 Uris Hall, Ithaca, NY 14853-7601, US. email:nicholas.kiefer@cornell.edu; US Department of the Treasury, Office of the Comptroller of the Currency, Risk Analysis Division, and CREATES, www.creates.au.dk, funded by the Danish National Research Foundation. I thank Jeffrey Brown, Mike Carhill, Hwansik Choi, Erik Larson, Mark Levonian, 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. 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.

References

Basel Committee on Banking Supervision (2004): "International Convergence of Capital Measurement and Capital Standards: A Revised Framework," Bank for International Settlements.

——— (2005): "Basel Committee Newsletter No. 6: Validation of Low-Default Portfolios in the Basel II Framework," Discussion paper, Bank for International Settlements.

——— (2006): "International Convergence of Capital Measurement and Capital Standards: A Revised Framework, Comprehensive Version," Bank for International Settlements.

BBA, LIBA, and ISDA (2005): "Low Default Portfolios," Discussion paper, British Banking Association, London Investment Banking Association and International Swaps and Derivatives Association, Joint Industry Working Group.

Berger, J., and L. M. Berliner (1986): "Robust Bayes and Empirical Bayes Analysis with Contaminated Priors," *The Annals of Statistics*, 14(2), 461–486.

Berger, J. O. (1980): *Statistical Decision Theory: Foundations, Concepts and Methods*. Springer-Verlag.

Chaloner, K. M., and G. T. Duncan (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), 174–180.

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

De Finetti, B. (1974): *Theory of Probability*, vol. 1. New York: Wiley.

Diaconis, P., and D. Ylvisaker (1985): "Quantifying Prior Opinion," in *Bayesian Statistics 2*, ed. by J. M. Bernardo, M. H. DeGroot, D. Lindley, and A. Smith, pp. 133–156. Elsevier Science Publishers BV (North-Holland).

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

Gavasakar, U. (1988): "A Comparison of Two Elicitation Methods for a Prior Distribution for a Binomial Parameter," *Management Science*, 34:6, 784–790.

Geweke, J. (2005): *Contemporary Bayesian Exonometrics and Statistics*. New York: John Wiley & Sons.

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

Halpern, J. Y. (2003): *Reasoning about Uncertainty*. Cambridge: MIT Press.

Heath, D., and W. Sudderth (1978): "On Finitely Additive Priors, Coherence, and Extended Admissibility," *Annals of Statistics*, 6, 335–345.

Hoogerheide, L. F., J. F. Kaashoek, and H. K. van Dijk (2007): "On the shape of posterior densities and credible sets in instrumental variable regression models with reduced rank: an application of flexible sampling methods using neural networks," *Journal of Econometrics*, 139, 154–180.

Kadane, J. B., J. M. Dickey, R. L. Winkler, W. S. Smith, and S. C. Peters (1980): "Interactive Elicitation of Opinion for a Normal Linear Model," *Journal of the American Statistical Association*, 75(372), 845–854.

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

Kahneman, D., and A. Tversky (1974): "Judgement Under Uncertainty: Heuristics and Biases," *Science*, 185, 1124–1131.

Kiefer, N. M. (2008): "Default Estimation for Low Default Portfolios," *Journal of Empirical Finance*, 16.

Kiefer, N. M., and C. E. Larson (2004): "Evaluating Design Choices in Economic Capital Modeling: A Loss Function Approach," in *Economic Capital: A Practitioner Guide*, ed. by A. Dev, chap. 15. Risk Books, London.

Lindley, D. V. (1982a): "The Improvement of Probability Judgements," *Journal of the Royal Statistical Society. Series A (General)*, 145(1), 117–126.

Lindley, D. V. (1982b): "Scoring Rules and the Inevitability of Probability," *International Statistical Review / Revue Internationale de Statistique*, 50(1), 1–11.

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

Nyarko, Y., and A. Schotter (2002): "An Experimental Study of Belief Learning Using Elicited Beliefs," *Econometrica*, 70(3), 971–1005.

OCC (2006): "Validation of Credit Rating and Scoring Models: A workshop for Managers and Practitioners," Office of the Comptroller of the Currency.

Pluto, K., and D. Tasche (2005): "Thinking Positively," *Risk*, August, 72–78.

Ritov, Y. (1985): "Robust Bayes Decision Procedures: Gross Error in the Data Distribution," *The Annals of Statistics*, 13(2), 626–637.

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

Savage, L. J. (1954): *Foundations of Statistics*. New York: Wiley.

van der Vaart, A. W. (1998): *Asymptotic Statistics*. Cambridge University Press.