The Probability Approach to Default Probabilities

Nicholas M. Kiefer_
Cornell University,
Departments of Economics and Statistical Sciences,
490 Uris Hall, Ithaca, NY 14853-7601, US.
email:nmk1@cornell.edu;
US Department of the Treasury
O¢ ce of the Comptroller of the Currency
Risk Analysis Division, 250 E. St. SW, DC 20219

OCC Economics Working Paper 2007-1

March 15, 2007

Abstract

The probability approach to uncertainty and modeling is applied to default probability estimation. Default estimation for low-default portfolios has attracted attention as banks contemplate the requirements of Basel II.s IRB rules. Nicholas M. Kiefer proposes the formal introduction of expert information into quantitative analysis. An application treating the incorporation of expert information on the default probability is considered in detail.
Keywords: Bayesian inference, Bayesian estimation, expert information, Basel II, risk management
JEL Classi.cations: C11, C13, C44, G18, G32
_Disclaimer: The views herein are those of the author and do not necessarily represent the views of the O¢ ce of the Comptroller of the Currency. Acknowledgement: I thank Je¤rey Brown, Mike Carhill, Erik Larson, Mark Levonian, Katja Pluto, Mitch Stengel, Dirk Tasche, and seminar participants at Cornell University and the OCC. These thanks come with gratitude but without any implication of agreement with my views.

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 as well as for compliance with Basel II rules for banks using the IRB approach to determine capital requirements (Basel Committee on Banking Supervision (2004)). Estimation of small probabilities has attracted considerable recent attention; see Basel Committee on Banking Supervision (2005), Balthazar (2004), BBA, LIBA, and ISDA (2005),and Pluto and Tasche (2005). The focus of this paper is on estimation of the default probability for a risk bucket on the basis of historical information and expert knowledge. Section 2 argues for the probability approach to uncertainty measurement. The probability approach to default modeling is uncontroversial, although perhaps the extent of the constraints imposed by the simple independent Bernoulli model are underappreciated. This model is briefly described in Section 3. In section 4 we argue that exactly the same considerations that lead to the probability approach for defaults should lead to the probability approach to default probabilities. As an example, we consider describing expert information in the form of a Beta distribution on the default probability. The probability approach allows coherent combination of expert and data information through Bayes Rule, taken up in Section 5. Section 6 considers estimators of PD based on the probability approach and compares them with alternatives, including the maximum likelihood estimator (the empirical default fraction) and a recent suggestion based on the upper endpoint of a confidence interval of prespecified coverage.

2 Uncertainty

Uncertainty is best described in terms of probabilities. This thesis can be based on prediction scoring, avoidance of sure losses in betting, Pareto optimality, axiomatic development, etc. Important references are De Finetti (1974) and Savage (1954). These arguments lead to a requirement of *coherence*. This weak requirement is just that systems of numbers describing uncertainty will not be such that another system can beat them in prediction, or that, if used for betting, they will not admit sure losses. This simple requirement is enough to insure that the predictions must combine like probabilities. Thus, let E, F, denote events (e.g., "asset 2 and only asset 2, defaults"). Let x_i be numbers used to quantify the uncertainty about events. The three properties implied by coherence are: P1: Convexity: $0 \leq x_i \leq 1$. P2: Additivity: Let x_1 refer to the event E and x_2 the event $\sim E$. Then $x_1 + x_2 = 1$. P3: Multiplication: Let x_1 correspond to E, x_3 to F given E, and x_4 to E and F. Then $x_4 = x_1 x_3$.

These three properties define a system of probabilities. The probabilities underlie statistical models. The probability approach to describing and modeling default uncertainty using statistical models is central to risk management and to the requirements of Basel II (under the IRB rules). In the case of default modeling, where measuring and controlling risk is the aim, it is widely accepted that the probability approach is the correct approach to default uncertainty. The fact that probabilities combine in accordance with convexity, additivity and multiplication is central for moving from probabilities of default on an asset, to default rates in a segment, to rates in a portfolio, and to a default probability for the bank. It is less well accepted that uncertainty about the unknown default probability can be usefully modeled in exactly the same way, using a statistical model, for exactly the same reasons.

3 Uncertain Defaults

The requirement of coherence prescribes relations among the probabilities of related events, but does not specify what these probabilities are. The usual approach in statistical modeling is to choose a statistical model that generates all the relevant probabilities as a function of a small number of parameters. The simplest and most common probability model for defaults of assets in a homogeneous segment of a portfolio is the Bernoulli, in which the defaults are independent across assets and over time, and defaults occur with common probability θ. Note that specification of this model requires expert judgement, that is, information. We will denote the expert information by e. Let d_i indicate whether the ith observation was a default ($d_i = 1$) or not ($d_i = 0$). The Bernoulli model for the distribution of d_i is $p(d_i|\theta, e) = \theta^{d_i}(1-\theta)^{1-d_i}$. Let $D = \{d_i, i = 1, ..., n\}$ denote the whole data set and $r = r(D) = \sum_i d_i$ the count of defaults. Then the joint distribution of the data is

$$\begin{aligned} p(D|\theta, e) &= \prod \theta^{d_i}(1-\theta)^{1-d_i} \\ &= \theta^r (1-\theta)^{n-r} \end{aligned} \tag{3.1}$$

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

$$p(r|\theta, e) = \binom{n}{r}\theta^r(1-\theta)^{n-r} \tag{3.2}$$

a Binomial (n, θ) distribution. This is a tremendous simplification, since, for fixed n, there are only a small number of likely defaults r, so analysis can be done for all likely datasets. The Binomial model is by no means always appropriate and its use requires judgement. For practical purposes, perhaps the most important shortcoming is the independence assumption. If there is heterogeneity in the default probabilities, perhaps due to changing macroeconomic conditions, the defaults will cluster. When defaults are infrequent, this may be difficult to capture in a statistical model. Thus, the Binomial specification is best when the sample is homogeneous and therefore may not be reliable over long periods or when the "bucket" includes dissimilar assets. The Basel II formula relies on an estimator for the marginal probability of default and then allows for correlation at a second stage in the model used to develop the formula for required capital. Direct treatment of correlation is possible and is proposed by Gossl (2005) and McNeil and Wendin (2006). These papers use a Bayesian approach to formulate an hierarchical model for dependence, but do not focus on incorporation of expert information. Because our emphasis is on the role of expert information, rather than on refinement of the likelihood specification, we will focus on the Binomial model.

4 Uncertain Default Probabilities

Equation 3.2 is a model for describing default probabilities (probabilities for different default configurations in a portfolio segment), but it is an incomplete model in that the parameter θ 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. We have seen how uncertain defaults can be modeled. The same methods can be used to model the uncertainty about θ. Define events E_i relevant to describing the uncertainty about θ, for example $E_1 =$ "$\theta < .0001$"; E_2 : "$\theta < .0005$," etc. Uncertainty about values of θ are coherently described by probabilities on these events. We assemble these probability assessments into a distribution describing the uncertainty about $\theta, p(\theta|e)$. Our approach is a classical Bayesian approach as described in Raiffa and Schlaifer (1961).

Now, $p(\theta|e)$ 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 We are quite willing to model the large number of probabilities associated with the possible different default configurations with a simple statistical model; in fact, a 1-parameter model. The same can be done with the prior specification. We can fit a few probability assessments by an expert to a suitable functional form and use that distribution to model prior uncertainty. Of course, as with the likelihood approach, there is some approximation involved, and care is necessary.

A convenient functional form, is given by the beta distribution

$$p(\theta|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \tag{4.1}$$

which has mean $\alpha/(\alpha+\beta)$, variance $\alpha\beta/((\alpha+\beta)^2(1+\alpha+\beta))$ and mode $(\alpha-1)/(\alpha+\beta-2)$. The special case of $\alpha = \beta = 1$ is the uniform distribution on the unit interval. Beta-binomial analysis is described in Raiffa and Schlaifer (1961).

As an illustration, consider a segment of loans which might be in the middle of a bank's portfolio in terms of risk. These loans might be roughly equivalent to S&P BBB or Moody's Baa. The bulk of these loans are to unrated companies and the bank has done its own rating to assign the loans to risk "buckets." We have consulted an experienced industry expert on these assets. The expert puts the default probability for assets in this portfolio at 0.01 (an expressed median value). When asked to condition on the probability being less than 0.01, and then considering the conditional median the expert returned the 25% quantile 0.0075. The corresponding question returned the 75% point at 0.0125. Thus this expert reports a rather tight distribution centered on 0.01 and nearly symmetric, probably reflecting extensive experience with portfolios active in this risk segment. We can fit these assessments to a beta distribution, choosing parameters to reflect our expert's information, resulting in $\alpha = 6.8$ and $\beta = 647$. This will be a sufficient representation of expert opinion for the point being made in this paper. In practice, the elicitation process is extensive. More information would be extracted from the expert, perhaps wider families of distributions than the beta would be considered, and there would be some iteration back and forth between the statistician and the expert.

We use the prior 4.1 to illustrate the approach without introducing conceptually unnecessary complications. In fact, essentially any distribution could be used. The important thing is that the distribution accurately reflects the expert information. A simple parametric refinement is to restrict the range, say from [0,1] to [a,b]. A more general option is to use mixtures of Beta distributions. It can be shown that any continuous prior on [0,1] can be arbitrarily well approximated by a mixture of Betas. The convenient functional forms used below may not be available (depending on the prior), but calculations through direct numerical integration or through simulation using Markov Chain Monte Carlo (Robert and Casella (2004)) are routine.

5 Inference

Given the distribution $p(\theta|e)$, we can multiply the probabilities in accord with the multiplication rule to 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,

$$p(r|e) = \int p(r, \theta|e) d\theta \tag{5.1}$$

If the value of the parameter θ is of main interest (rather than the number of defaults) we can divide to obtain the conditional (posterior) distribution of θ :

$$p(\theta|r, e) = p(r|\theta, e)p(\theta|e)/p(r|e) \tag{5.2}$$

which is Bayes rule.
Using specifications 3.2 in which expert opinion appears in the likelihood specification and 4.1 in which expert opinion is reflected in the values of α and β we find for the predictive distribution 5.1

$$p(r|e) = \frac{\Gamma(r+\alpha)\Gamma(n-r+\beta)\Gamma(\alpha+\beta)\Gamma(n+1)}{\Gamma(r+1)\Gamma(n-r+1)\Gamma(\alpha)\Gamma(\beta)\Gamma(n+\alpha+\beta)} \tag{5.3}$$

and for the posterior 5.2

$$p(\theta|r, e) = \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+r)\Gamma(\beta+n-r)} \theta^{\alpha+r-1}(1-\theta)^{\beta+n-r-1} \tag{5.4}$$

With our example prior distribution for expert opinion the predictive distributions 5.3 of the number of defaults in a portfolio segment of size 100 is plotted in Figure 1.

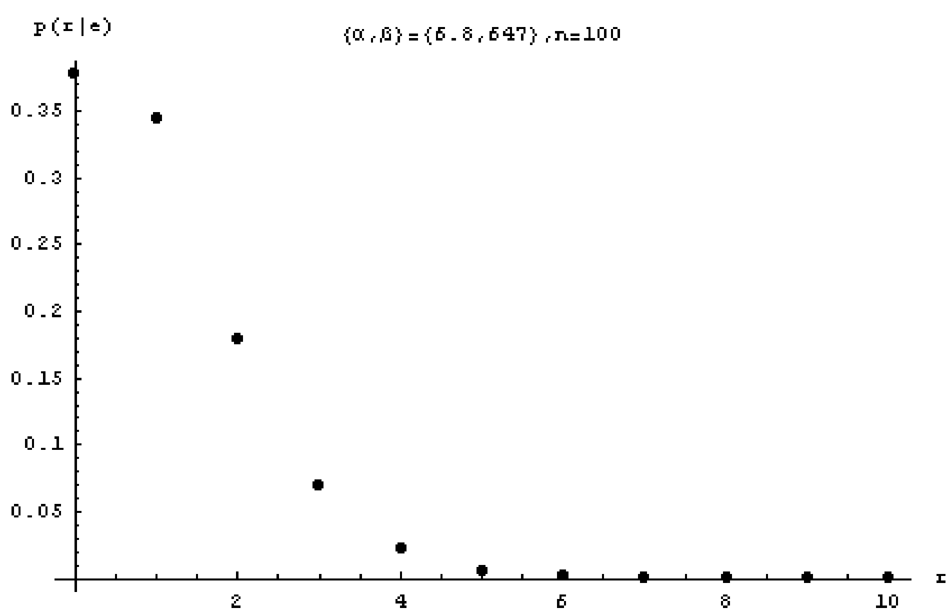

Figure 1: Predictive Distribution $p(r|e)$ for n=100

A candidate estimator for PD, suitable for plugging into the formulas given by the Basel committee's capital model, is

$$\overline{\theta} = E(\theta|r, e) = (\alpha + r)/(\alpha + \beta + n) \tag{5.5}$$

We will not spend much effort justifying the use of the mean; except to point out that the mean is optimal with respect to squared error loss. For any summary statistic, it is appropriate to report an indicator of its reliability, for example the standard deviation $\sigma_\theta = \sqrt{E(\theta - E(\theta|r, e))^2}$.

The posterior distributions for a sample size of 100 with 0,1, and 5 defaults are plotted in Figure 2. These distributions represent in full the post-data uncertainty about θ given the expert information and likely (and some unlikely) samples.

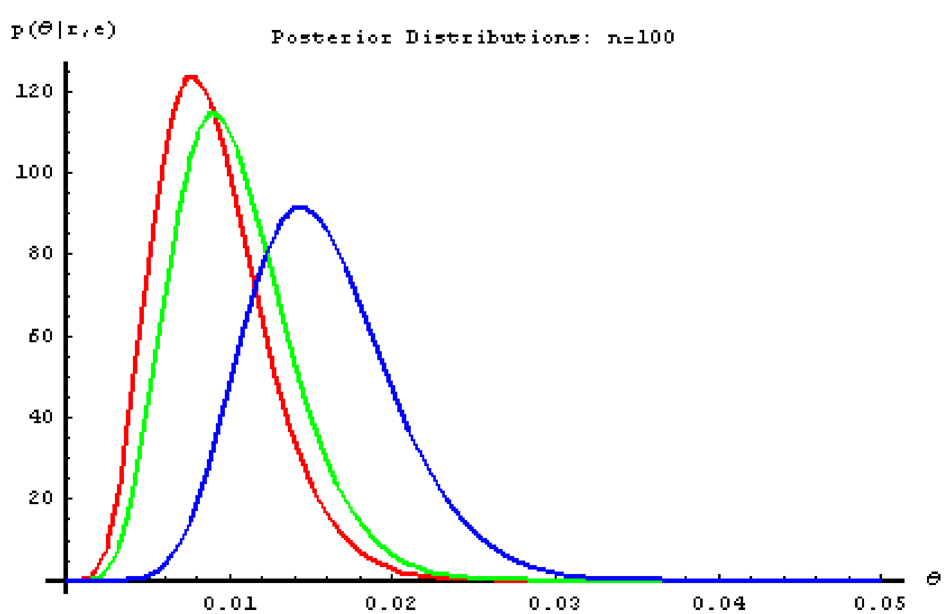

Figure 2: Posterior Distributions for r=0 (red), r=1 (green) and r=5 (blue).

6 Estimators

We advocate point estimators based on the probability approach. In the examples tabulated below we focus on the posterior mean 5.5. An alternative estimator in wide use is the maximum likelihood estimator (MLE) $\widehat{\theta} = r/n$. For large samples and θ well away from the extremes of 0 and 1, the MLE is a very good estimator in many applications, since the data evidence can be expected to dominate prior evidence provided by the expert. That is, the data evidence is accumulated as the data set increases in size, while the amount of expert evidence is fixed. When that occurs, the likelihood approach and the probability approach coincide as the sample becomes large. Thus, the likelihood estimator can be given a slightly strained probability interpretation. However, for θ near zero (the case in applications we consider) this domination does not occur for any practical sample size (Kiefer (2006)).

An estimator recently proposed is the confidence estimator (Pluto and Tasche (2005)). We consider a simple case of this estimator. The authors give refinements including extensions to simultaneous estimation of default probabilities for different buckets. Our comments apply equally to these cases. The principle attractive feature of the confidence estimator is that it gives a nonzero estimator in the case of zero observed defaults, a case which is extremely likely to occur in low-default (probability) portfolio segments. Define F(u|r,n) as the probability of obtaining r or fewer successes in a sequence of n independent Bernoulli trials each with success probability u. Thus

$$F(u|r,n) = \sum_{i=0}^{r} \binom{n}{i} u^i (1-u)^{n-i}$$

Then define the estimator for a prespecified number $\delta \in [0,1]$ chosen by the modeler,

$$\theta_\delta = F^{-1}(\delta|r,n). \tag{6.1}$$

For the case r=0, the estimator takes a simple form, $\theta_\delta = 1 - \delta^{(1/n)}$. A few comments on this procedure are in order. Although we criticize this estimator on both fundamentals and performance, it does represent a major advance in thinking about default probabilities. Specifically, it suggests that the unbiased estimator can be abandoned and that a practical estimator might adjust the unbiased estimator toward more likely (in this case nonzero) values. One wonders why it is considered easier to think about realistic values for δ than it is to think about θ directly. Pluto and Tasche interpret the choice of δ as a specification of "conservatism." In any case, there does not seem to be a direct probability interpretation of the estimator θ_δ. Rather, θ_δ is interpreted as that value of θ at which the probability of seeing r or fewer defaults is equal to δ. It is difficult to see a justification for this approach, other than the appeal of bounding the estimate away from zero (provided δ, a choice parameter for the modeler, is bounded away from zero). Conservatism, if desired, could be accommodated by using a percentile estimator based on the posterior distribution - perhaps using a 60% or 70% point. This calculation is straightforward given the posterior. The resulting estimator could be greater or lesser than the MLE, so the confidence estimator cannot be given this interpretation, although the spirit may be the same. Of course, the posterior mean has a direct probability interpretation. The maximum likelihood estimator has an approximate probability interpretation as well as an interpretation as that value of θ which maximizes the probability of seeing the sample actually observed.

The performance of the alternative estimators is given in Table 1 for n=100 and all plausible values of the number of realized defaults. Results are given based on the posterior distribution using our expert's information; hypothetical information from an alternative, less confident expert with $\alpha = 1.5$, and $\beta = 150$,(this expert has approximately twice the prior standard deviation as our actual expert); from the likelihood alone, and the confidence estimator $\theta_{0.1}$.

Table 1: Point Estimates

n	r	$\overline{\theta}$	$\overline{\theta}_{lce}$	$\widehat{\theta}$	$\theta_{0.1}$
100	0	0.0090	0.0060	0	0.0228
100	1	0.0103	0.0099	0.01	0.0383
100	2	0.0117	0.0139	0.02	0.0523
100	3	0.0130	0.0179	0.03	0.0656
100	4	0.0143	0.0219	0.04	0.0783
100	5	0.0157	0.0258	0.05	0.0908

Notes: The posterior mean, the same based on a less confident expert, the MLE and the confidence estimator.

As expected, the posterior mean for our expert is tightly clustered around the prior expectation (about 0.0104) for the sample size 100, even with the unlikely value of 5 defaults. Our hypothetical less-confident expert also supplies estimators which do not seem unreasonable, though of course they are much more sensitive to sample variation. The MLE reflects the well-appreciated problem that the estimator is zero when zero defaults are observed. The confidence estimator depends on the user-specified δ; the value chosen 0.10, is suggested in Pluto and

Tasche (2005); the values 0.5, 0.25, 0.05, 0.01 and 0.001 are also considered in that paper. For comparison, the θ_δ for these values or δ and for r=0 are 0.007, 0.014, 0.030, 0.045, and 0.067. The confidence estimator is often greater than the alternative estimators and always greater than the MLE, reflecting what the proposers call converatism. It is conservative in that it certainly overstates risk relative to the MLE, and this might be an area of application in which upside errors are less problematic than downside errors.

Let us confront these estimators with a stress test, in the spirit of the validation exercises expected of financial institutions. See OCC (2000) and OCC (2006). Table 2 reports the same complement of estimates, now for a sample of size 10 and a sample of size 1000.

Table 2: Point Estimates, Stress Test

n	r	$\overline{\theta}$	$\overline{\theta}_{lce}$	$\widehat{\theta}$	$\theta_{0.1}$
10	0	0.0102	0.0093	0	0.2057
10	1	0.0118	0.0155	0.1	0.3368
10	2	0.0133	0.0217	0.2	0.4496
1000	0	0.0041	0.0013	0	0.0023
1000	10	0.0102	0.0100	0.01	0.0154
1000	50	0.0343	0.0447	0.05	0.0600

Notes: The posterior mean, the same based on a less confident expert, the MLE and the confidence estimator.

For the very small sample size the probability estimators both appear reasonable. The preferred estimator is the one that actually reflects expert judgement, given in columns 3 and 4. The MLE is very sensitive to the number of defaults and gives the estimator zero for zero-default samples. The confidence estimator is very conservative for the small sample. For very large samples all estimators work better, as expected, though the experts pull estimators toward the prior means while the confidence estimator is always above the likelihood estimator. The MLE for r=0 is again a potential problem, but this is a very unlikely sample for n=1000.

Two final comments on the confidence estimator are appropriate. First, consider an example in which a sample of 100 observations is drawn from a Binomial(100,0.01) distribution. A zero will be observed with probability 0.366; the MLE will be zero and $\theta_{0.1}$ will be positive. This might be considered desirable, though our expert might regard the estimate 0.023 (see Table 1) quite surprising.

It is, after all, more than twice the true value in our example. With probability 0.634, the number of defaults will be positive and the confidence estimator will substantially overestimate the true value of the default probability. Second, consider a model for nondefaults - survival. In our samples with r defaults, there are $n-r$ survivals. One might consider estimating a survival probability ρ rather than the default probability θ. Of course, $\rho = 1-\theta$, so the associated prior is obtained by a simple linear change of variables. In our example, if the prior on θ has parameters (α, β), then the prior on ρ has parameters (β, α).There is no new prior information imposed by looking at the singular joint distribution, and there is no new data information in the distribution of $n-r$ that is not in the distribution of r. Thus, we would not in practice analyze these probabilities separately. Clearly, in the probability approach, $E(\theta|e) = 1 - E(\rho|e)$ and $E(\theta|r,e) = 1 - E(\rho|r,e)$. Since $\widehat{\rho} = (n-r)/n$, we have $\widehat{\theta} = 1 - \widehat{\rho}$ in the likelihood approach as well. Consider, however, the estimators $\theta_{0.1}$ and $\rho_{0.1}$(defined by interchanging r and n-r in 6.1). For a sample of size 100 with 5 defaults, we find $\rho_{0.1}$ =0.975 and $\theta_{0.1}$ =0.091. Referring to our discussion above (Section 2), we see that as descriptions of the uncertainties regarding defaults, the confidence estimators violate property P2, additivity, and hence are incoherent.

7 Conclusion

Expert information is crucial in risk management and specifically default modeling. This information is typically based on a mix of subjective judgement, related information not specifically modeled, and long experience with related data sets. The expert information appears in the assignment of assets to segments on the basis of their risk, in the definitions of the segments, in the choice of sample period and in the chosen statistical model. This paper argues that the expert information on the likely values of the parameters of the risk model should also be formally incorporated in the analysis.

Any measure of uncertainty should satisfy a reasonable system of properties known as coherence. The probability approach to modeling defaults is coherent, widely accepted and uncontroversial. The same justifying arguments imply that uncertainty about the unknown default rates should be modeled by probabilities. In the case of default modeling, a parametric model is customary. We use the same approach to modeling expert information about the unknown parameters -

using answers to questions about the unknown default rate to fit a parametric model to the expert's beliefs. In practice, this procedure is intensive and requires more effort and expert involvement than that given in our example. With this probability distribution in hand, updating beliefs with data information entering through the likelihood function is straightforward using Bayes Rule.

Our examples illustrate the application of the probability approach. We have used the opinions of an expert on the likely default probabilities for a risk segment in the middle of a bank's portfolio. The expert was quite certain about the likely ranges of the default probability. As a check, we also constructed a hypothetical, less-confident expert and calculated the posterior statistics. The probability approach is feasible as well as logical. We considered the likelihood approach, perhaps appropriate for very large samples and risky portfolio segments. As often noted, that approach gives an estimator of zero for default probabilities in segments with no defaults in the sample. That value is considered unacceptable. A recent proposal for estimating positive default probabilities using the upper endpoint of an approximate confidence interval for $\widehat{\theta}$ with specified coverage is also examined and is shown to be incoherent.

References

Balthazar, L. (2004): "PD Estimates for Basel II," *Risk*, April, 84–85.

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.

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.

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

Gossl, C. (2005): "Predictions based on certain uncertainties - a Bayesian credit portfolio approach," Discussion paper, HypoVereinsbank AG.

Kiefer, N. M. (2006): "Default Estimation for Low Default Portfolios," OCC Working Paper.

McNeil, A. J., and J. P. Wendin (2006): "Bayesian Inference for Generalized Linear Mixed Models of Portfolio Credit Risk," Discussion paper, Department of Mathematics, ETH Zurich.

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

——— (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.

Raiffa, H., and R. Schlaifer (1961): *Applied Statistical Decision Theory.* Harvard Business School.

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.