Likelihood Approaches to Low Default Portfolios

Alan Forrest
Dunfermline Building Society

Version 1.1 22/6/05
Version 1.2 14/9/05

1. Abstract

This paper proposes a framework for computing conservative Probabilities of Default (PDs) for Low Default Portfolios (LDPs) that is

- statistically valid and robust,
- flexible and extendable to a wide variety of conditions and business areas,
- involves all observed data,
- involves expert opinion explicitly and quantitatively, and
- graduates from the LDP to normal regimes smoothly, without the need for artificial cut-overs.

The framework is justified by statistical theory and illustrated by examples which also extend the range of examples explored in recent studies. The example outputs agree well with the Confidence Interval approach and support the "most prudent estimate" principle. However, in general, the most prudent estimates will sometimes more conservative than this framework.

Finally, in applying the Vasicek approach to interloan correlation, this paper details a numerical approximation of the relevant expectations that is deterministic and more efficient than approximations using simulation proposed in recent studies.

2. Introduction

2.1 The context and recent developments of Low Default Portfolios are covered in the June 2005 LDP Expert Working Group paper to the CRSG [CRSG]. That paper discusses a wide range of topics, but this paper concentrates on LDPs for which few default data can ever be made available: [CRSG] point 7(c). Here the issues are

- How to cope accurately and conservatively with few defaults.
- How to involve expert opinion.
- At what point and how to cut over from Low Default to the normal regime.

The Appendix to [CRSG], by Alan Cathcart [C], proposes a method for treating LDPs. Other methods are also noted in [C], among them the method of Pluto and Tashe [PT] which is of special interest in this paper. Both [C] and [PT] use a binomial model for the observed default data, using PDs among the model parameters. From the observed default levels they build the class of data outcomes at least as poor as the observed data in regard of defaults. A conservative PD is found as the point at which the probability of this class falls below an agreed level. Further, [PT] develops a useful "most prudent estimate" principl which allows the intuitive extension of the analysis to many grades. There is still debate about which such confidence level to use.

2.2 This paper takes the same model based approach to PD estimation but differs in the final step. Instead of the probability of a class of data outcomes, Likelihood and Likelihood Ratio are used in this paper.

This small change is crucial. It makes a direct connection with the classical theory of statistical inference, with its well-known approximations valid for high default cases – see section 3.

It is also easily generalized to many gradings and more complex situations and this is reflected in the breadth of the examples in section 4. Note that it is not clear how to generalize the confidence interval methods of [C] to this situation: what combinations of default are worse than others? Is the combination - 4 low risk defaults, 0 high risk defaults, say - better or worse than 3 low risk and 2 high risk?

Likelihood is also fundamental to Bayesian methodologies which this paper touches but does not explore far. Furthermore, the method here could be mixed with other approaches, such as non parametric bootstrapping put forward in [SH], but this is not explored.

2.3 In [C] and [PT], additional complexity and realism are built into the model via a correlation between the separate accounts in the portfolio. Once again the best value of correlation is not settled, although 12% is used consistently. This paper also makes this development, proposing on the way an improvement to the numerical approximation methodology used in [C] (See example 4.3).

2.4 This paper aims to draw out the advantages of the Likelihood Framework

- Based on Classical Statistical Theory – well-established theory, accepted in the profession to be a robust and accurate way of making statistical inference. Applied in practice by professional Statisticians in all contexts of data analysis: Medicine, Epidemiology, Government, Environment and Business Operations. Standard statistical programming packages perform the required calculations.
- Works for Low and High default portfolios, moving seamlessly between them – indeed standard statistical model building processes, automatic or hand made, use Likelihood as the basis of their choice of best model. The method outlined here simply makes explicit, for Low Defaults, what happens in high default portfolio model building.
- Accommodates the refinements used in [C], eg inter-loan correlation.
- Gives the required output – PDs and their conservative limits are found directly as solutions of convex optimisation problems, or by reading off graphs in simple situations.
- A General Open Framework – allows business to set explicitly the way in which expert opinion and data interact. Separates the expert opinion stage from the data stage cleanly so that each can be examined and checked.

2.5 The scope of this paper is as follows

- The statistical background and examples of the Likelihood framework.
- Expert opinion expressed as portfolio gradings.
- Involvement of correlation via the Vasicek formulation to the same extent as in [C, PT].

Out of scope

- The process of converting expert opinion into gradings – assume that such “elicitation” has been performed to the satisfaction of the business and regulator beforehand.
- Development of Bayesian approaches beyond the basic Likelihood, such as classes of Priors, interpretation of Posteriors, optimisation of response etc.

- The correct confidence level to choose – from statistical tradition, this note assumes 95% confidence throughout and other choices are covered in [CRSG].
- The correct level of correlation in the Vasicek model.
- Refined analysis of Likelihood Ratio cut levels – this is the subject of on-going research.
- Specific issues raised by particular business types or industry regulations.

2.6 Terms and definitions used in this paper

Likelihood: the probability that precisely the data observed could have occurred. It depends on the model used and is a function of the parameters of the model. In this note the parameters include the PDs .

Maximum Likelihood: the largest value of likelihood among all relevant combinations of the model parameters.

Likelihood Ratio: the ratio of the Likelihood to the Maximum Likelihood, a number always less than 1. For theoretical reasons, often rescaled as the positive quantity $-2\log(LR)$, LR being the Likelihood Ratio (see example 4.2 ahead).

Cut: the value at which to cut the Likelihood Ratio in order to find the confidence region for the model parameters (including PDs). If x is the cut value, then the confidence region is the set of parameters **p** where $-2\log LR(\mathbf{p}) < x$.

Confidence Region: see *Cut*.

Grading: a division of the portfolio into groups according to risk level, so that we know or assume that all loans in one grade are definitely more or less risky than all loans in another grade.

Prior Odds: a subjective quantification of how the parameters of the model lie, made before the data are introduced to the analysis. It is a function of the parameters in the model and is higher for parameter combinations considered more probable. The Prior is usually gathered indirectly from experts in a formal “elicitation” exercise.

An example of a Prior is the grading of a portfolio, where the expert’s opinion about ranking of risk is reflected in inequalities between the PDs of loans in each grade (see examples D,E and F ahead). In the examples of this note we do not use any other kind of Prior, but more general Priors are clearly possible. This would be the subject of a separate study into Bayesian techniques.

3. Background and overview of the Likelihood framework

3.1 The statistical theory of inference which underlies the Likelihood approach in this paper is well covered in standard texts, eg [S]. This section gives an overview of the most relevant parts.

The aim of inference is to deduce facts about statistical models that describe observed data. From an observed dataset, a statistical model is proposed, for example the binomial model that describes the LDPs. This model has parameters, eg the PDs. The model and the data together determine the likelihood function of the parameters. The "best" estimate of the parameters is found at the maximum likelihood.

A parameter combination **p** will be within the 95% confidence region of the maximum likelihood estimate if, by the observed data, the hypothesis,

H0: parameters = **p**,

cannot be rejected in favour of

H1: parameters not = **p**

with 95% confidence. Likelihood ratio gives the best test of this hypothesis: if LR is above a certain value then the hypothesis rejected. Thus the confidence region is determined by levels of LR.

3.1.a 3.1 skates over the issue of whether the same cut value applies for each **p**. It does not; but to a good approximation it does, and this uniformity gets rapidly better as the size of the portfolio increases. For portfolios of over 10 loans per grade, the principle of uniform cut level applies in practice. Note that Basel II requires sufficiently many obligors (not defaults) before modelling can proceed, and 10 per grade is a reasonable lower bound that theory can suggest to the regulator.

Thus the confidence region is determined by a cut value as described in 2.6 above.

3.2 For large portfolios with many defaults, the cut level can be determined quickly. The value –2logLR is expected to be distributed as Chi-squared with an appropriate number of degrees of freedom. This approximation is understood to be adequate for practical use down to about 5 defaults per grade.

In the examples in this paper, the number of grades equals the number of degrees of freedom. Therefore the 95% confidence cut value is simple the 95th percentile of the appropriate Chi-squared distribution, which can be found in standard statistical tables. For example, if there is only one grade then the cut value is 3.85, approximately.

3.3 For fewer defaults, the choice of cut value may not be well approximated by Chi-squared considerations. Indeed for 0 defaults, the cut value for 95% confidence will be 5.99 (=-2log(0.05)) whatever the degree of freedom.

Therefore cut values have to be determined specially for default values between 1 and 5. Note that these cut values depend only on the number of defaults and degrees of freedom, not on the size of portfolio.

The following table gives approximate cut values for the case of a single grade:

Number of Defaults	Level of cut (95% confidence)
0	6.0
1	4.7
2	4.25
3	4.15
4	4.05
5	4.0
6,7	3.95
8,9,10	3.9
>10	3.85 = Chi-squared value

3.4 Point 3.3 notwithstanding, this paper takes a simplified approach to determining cut values: Simply insist that the large default regime covers the low default portfolio regime as well, with the exception of the 0 default case as noted above.

Number of Defaults	Level of cut (95% confidence)
0	6.0
>=1	Chi-squared value

This is an adequate approximation for the examples that follow, but a more careful analysis along the lines of 3.3 will be needed when there are many grades in a portfolio.

3.5 The following items summarise the main steps of the framework method. The next section will elaborate how these steps work in practice.

1. Examine the Expert opinion, ECAI grades etc available, interpreting these as a grading of the portfolio and as Prior assumptions about the PDs.
2. Calculate the Likelihood function for the observed data, taking into account any imposed parameters (eg correlations) as well as the restrictions on the PDs imposed by point 1.
3. Find the maximum Likelihood and compute the Likelihood Ratio function, LR. The point of maximum likelihood gives the "best estimate" of the PDs.
4. Decide the cut level (See 3.3,3.4). This level depends on the number of grades and the required confidence level.
5. The confidence region for the PDs is where –2logLR is below the cut.
6. The confidence region of PDs as a whole then gives the conservative values for the PD of each of the accounts in the portfolio.

4. How it works in practice

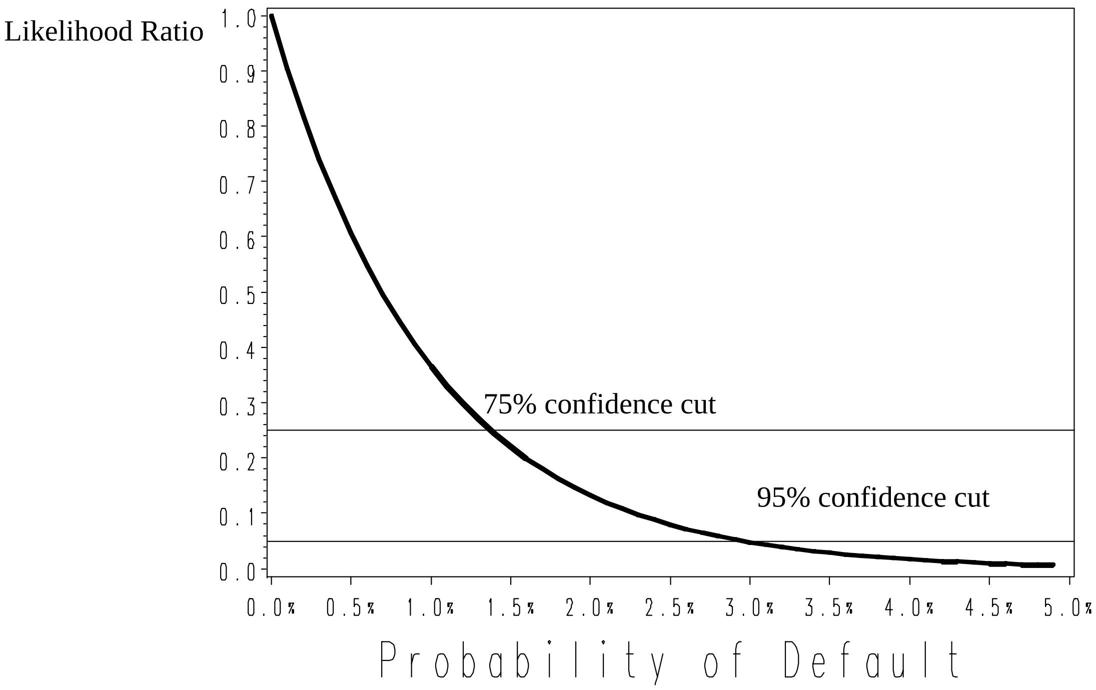
The following six examples examine the framework's performance on a portfolio of 100 accounts over 1 year, under variation of three conditions: Number of grades (1 or more), Number of defaults (0 or more), Correlation (0% or 12%).

The section ends with a summary of the PD values obtained and a comparison with the outcomes of the confidence interval approach in [C].

4.1 Single Grade, no defaults, no correlation. This is a portfolio of 100 loans, uniform for the purposes of risk, with no default in the course of a year. A single parameter defines the portfolio's statistical behaviour: probability of individual default, p. In this case the likelihood is L(p) = (1-p)^100.

The Likelihood is maximised at p=0 – this is the "most likely" value of p, which is intuitive. A conservative limit for p can be found by looking where the Likelihood Ratio (= Likelihood in this case) is above a certain value. Classically a cut of 0.05, corresponding to 95% confidence, is chosen, but other confidence levels are discussed in other LDP papers.

The graph below gives the Likelihood Ratio. The horizontal cuts indicate conservative estimates for p: 3% approximately at the 95% confidence level, and 1.3% at the 75% confidence level.

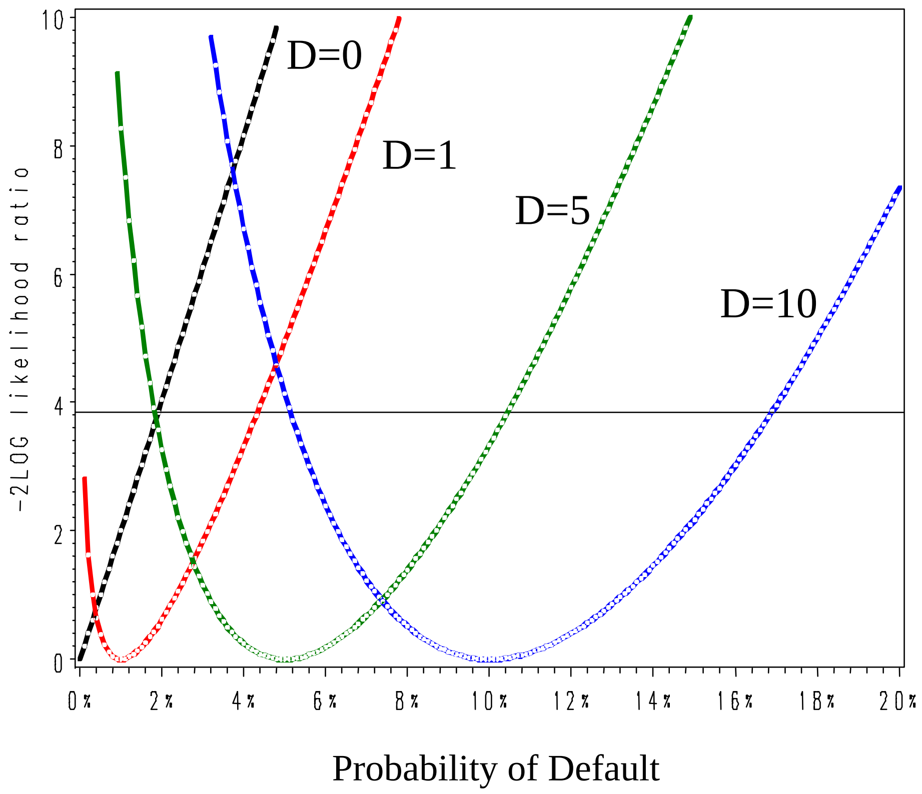


4.2 Single Grade, several defaults, no correlation. A uniform portfolio of 100 loans with D defaults. Once again individual PD, p, defines the system. The Likelihood is the probability that precisely this outcome of defaults arises: L(p) = p^D (1-p)^(100-D).

The maximum likelihood is found at p = D/100 , once again intuitively correct. To get a conservative limit for p, it is correct to consider the Likelihood Ratio, i.e. the Likelihood as a proportion of the maximum likelihood: LR(p) = L(p) / L(D/100) . For theoretical reasons, it's convenient to present –2 * the logarithm of this quantity:

$$-2\log LR(p) = 2(\log(D/100) - \log L(p)).$$

This Log Likelihood Ratio is plotted in the graph below for various default levels.



Following the methodology of 3.4, we can use the 3.85 cut line to read from the graph that, for example, when there are 10 defaults out of 100, PD lies between 5.1% and 16.9% with 95% confidence.

4.3 Single Grade, several Defaults, correlation. This introduces the effect of correlation, as noted and referenced in [C]. Here the Likelihood in 4.2 is modified by replacing p with the expression

$$\Phi\left(\frac{\Phi^{-1}(p) + Y\sqrt{\rho}}{\sqrt{1-\rho}} \right)$$

then taking an expectation with respect to Y, a standard normal variate N(0,1). This becomes an expected likelihood which we treat just like likelihood in the analysis before:

$$\mathbf{E}\left[\left(\Phi\left(\frac{\Phi^{-1}(p) + Y\sqrt{\rho}}{\sqrt{1-\rho}} \right) \right)^{D} \left(1 - \Phi\left(\frac{\Phi^{-1}(p) + Y\sqrt{\rho}}{\sqrt{1-\rho}} \right) \right)^{N-D} \right]$$

As before, N=100 and D is the number of defaults. This expectation can be approximated effectively by an average of M terms, as Y runs over regular quantiles of the standard normal distribution:

$$\frac{1}{M}\sum_{i=1}^{M}\left[P\left(p, \frac{i-1/2}{M} \right)^{D}\left(1-P\left(p, \frac{i-1/2}{M} \right) \right)^{N-D}\right]$$

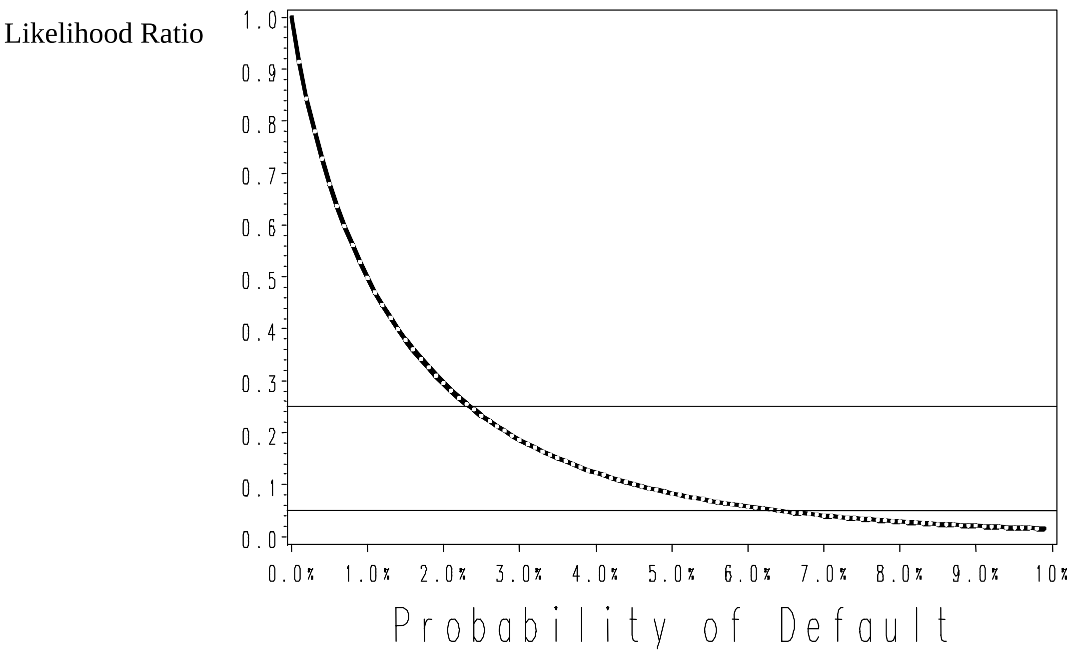
where

$$P(p,x)=\Phi\left(\frac{\Phi^{-1}(p)+\Phi^{-1}(x)\sqrt{\rho}}{\sqrt{1-\rho}} \right)$$

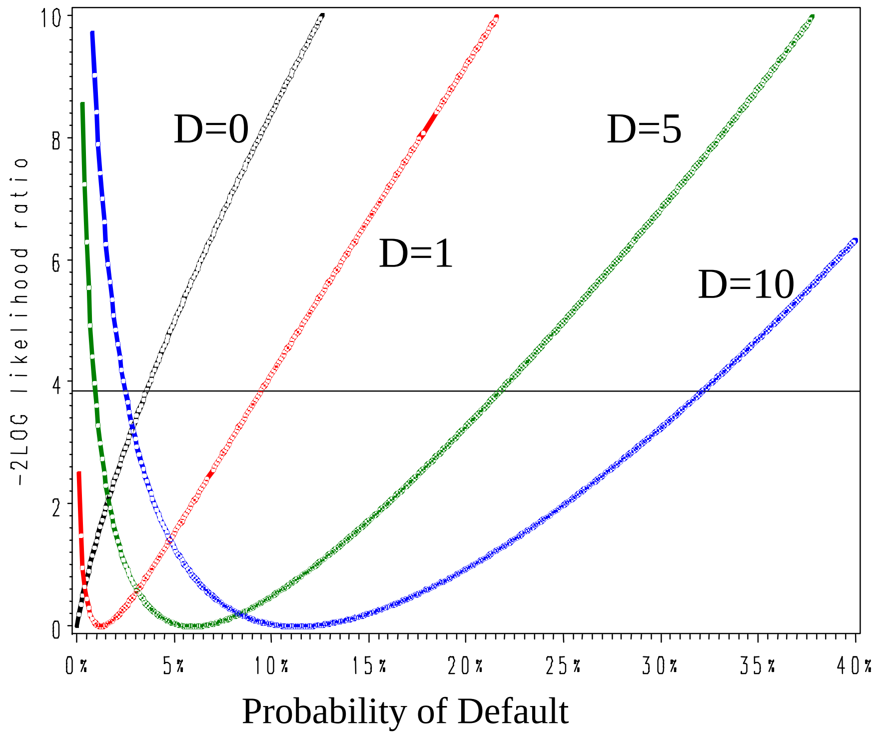
Note that this differs from the simulation approach in [C]. There are two advantages of the approach here:

- It's deterministic – the approach in [C] is likely to be accurate, but there remains a chance that the approximation is unluckily wide.
- It's more accurate - the accuracy of the approach in [C] is order square root of 1/M, whereas the accuracy of the expression above is order 1/M. In practice, this means that M can be chosen smaller in the method used here, reducing computation time and resource.

The following graph gives the case of no defaults and correlation = 12% , indicating that at 95% confidence p < 6.5% , and at 75% confidence p< 2.3% approximately.



For several defaults the log likelihood ratio can be computed as before, but using the expected likelihood instead. The following graph shows –2logLR for various values of default as before, but with correlation = 12%.



For example, looking along the 3.85 cut, for 10 defaults out of 100 cases, with correlation set at 12%, p is found between 2.5% and 32% with 95% confidence.

4.4 Multiple Grades, no Defaults, no correlation. As in [PT], the portfolio is now split into grades each of which has possibly a different PD. The grades are ranked from least to most risky and this imposes inequalities on the PDs.

In this example, there are two grades: the least risky with 70 accounts, the most risky with 30 accounts. The PDs are p and q respectively, with $p \le q$ imposed.

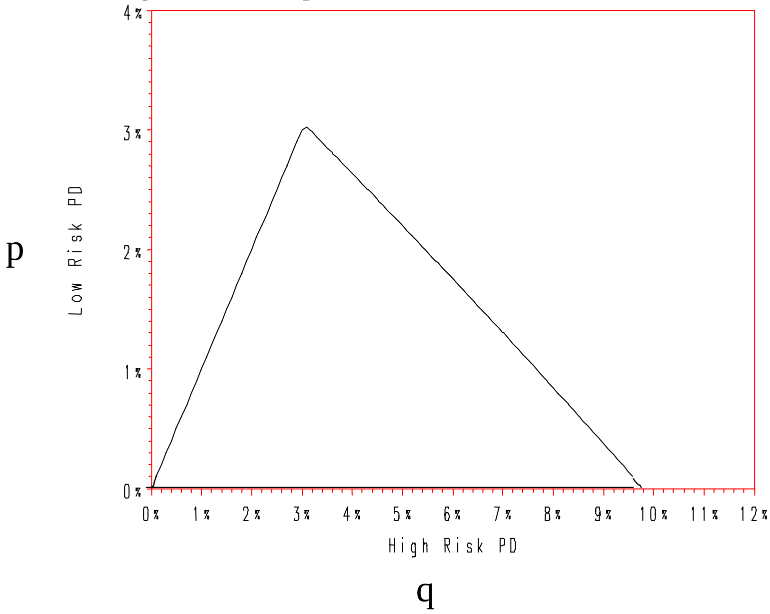
In the case of no defaults, the Likelihood is:

$$L(p,q) = (1-p)^{70} (1-q)^{30} .$$

Maximum likelihood is found at p=q=0 , as expected , but the conservative region for p and q is found where Likelihood exceeds 0.05 (95% confidence). The following graph shows the 95% confidence region for p and q together (within the triangle), with the imposed restriction $p \le q$.

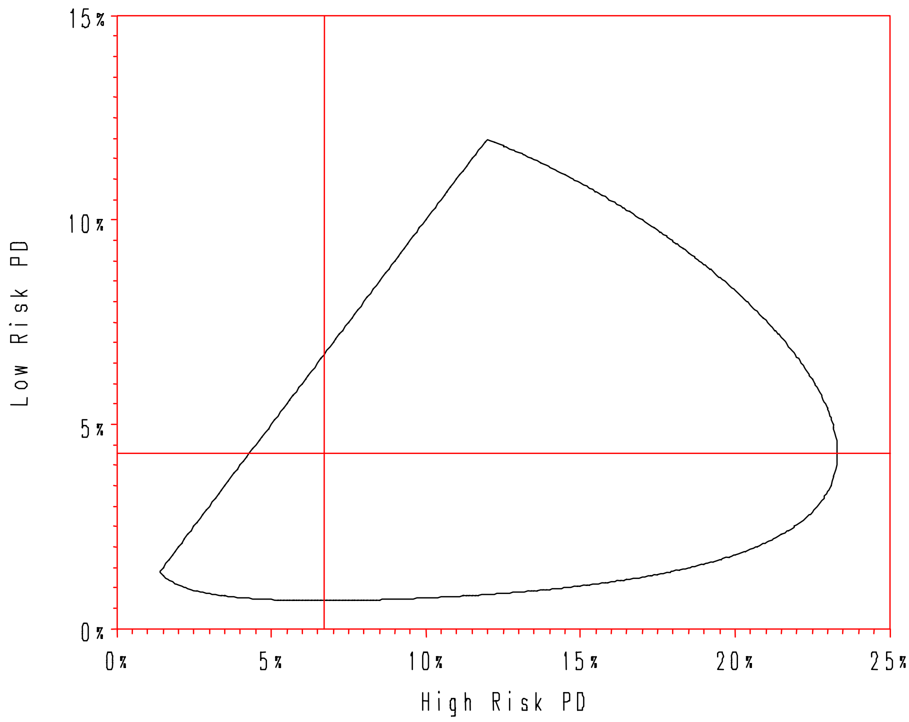
The extremes of the region give conservative estimates for each parameter separately: $q \le 9.8\%$ and $p \le 3\%$. These values agree with the "most prudent estimate" approach in [PT] nicely: the conservative value for p is found on the line p = q (i.e. 100 equally graded accounts), while the conservative value of q is found on the line p=0 (i.e.30 equally graded accounts).

95% confidence region for example 4.4.



4.5 Multiple Grades, some defaults, no correlation. For example, assume 3 defaults out of 70 low risk accounts and 2 defaults out of 30 high risk accounts.

The likelihood function is given by $L(p,q) = p^3(1-p)^{67} q^2(1-q)^{28}$ which achieves its maximum at the intuitive $p=3/70 = 4.3\%$, $q = 2/30 = 6.7\%$. As before, the likelihood ratio is $LR(p,q) = L(p,q)/L(3/70,2/30)$. The conservative region for p and q is where $-2\log LR(p,q) < 5.99$, this being the 95% confidence value for Chi-squared with 2 degrees of freedom. This region is illustrated in the following graph, noting the restriction $p \le q$ as in example 4.4 above.



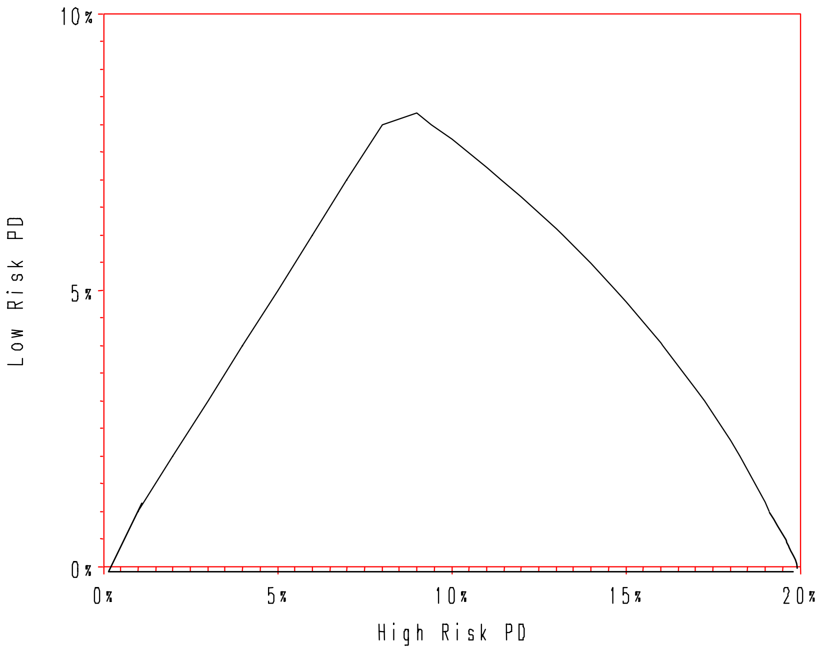
Reading off the graph, we find that q <= 23% and p <= 12% are conservative estimates at the 95% confidence level.

Note that again the value of p is found on the line p=q, so that the "most prudent" [PT] approach is appropriate. Unfortunately, this is not a general principle. For example, if there are 0 low risk defaults and many high risk defaults then the confidence region will miss the p=q line completely and the conservative estimates of p will not agree with the most prudent principle. In general, the most prudent principle will produce over-conservative estimates.

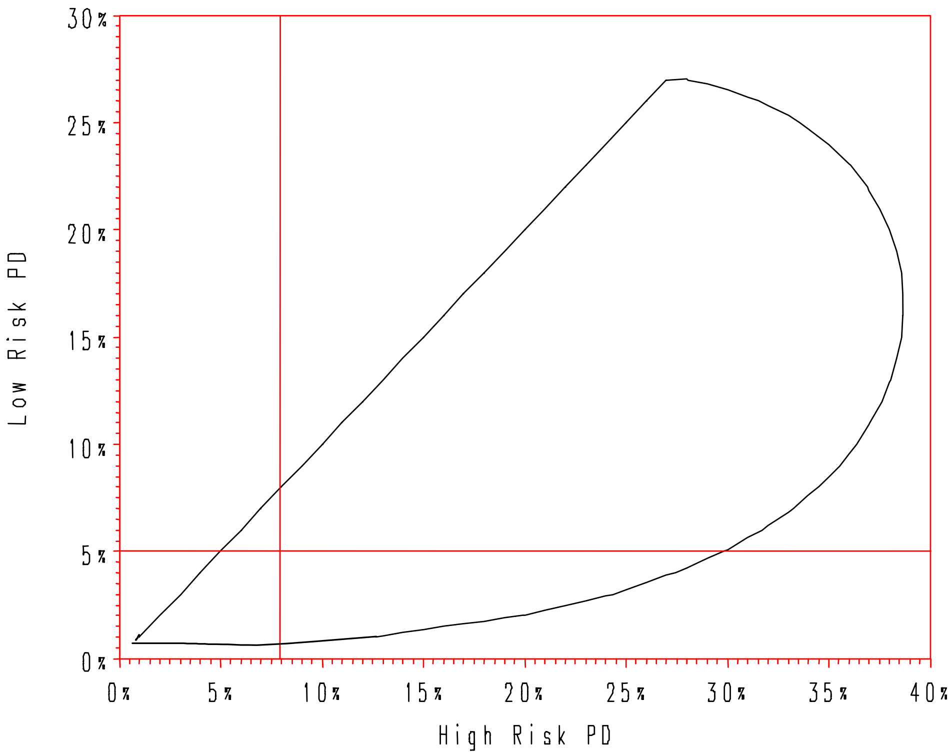
4.6 Multiple Grades, Some Defaults, Some Correlation This combines the features of all the previous examples. Here we have N1=70 low risk accounts, with D1 defaults, and N2=30 high risk accounts with D2 defaults, giving an expected likelihood function

$$\mathbf{E}\Big[\Big(\Phi\Big(\frac{\Phi^{-1}(p)+Y\sqrt{\rho}}{\sqrt{1-\rho}} \Big)\Big)^{D_1} \Big(1-\Phi\Big(\frac{\Phi^{-1}(p)+Y\sqrt{\rho}}{\sqrt{1-\rho}} \Big)\Big)^{N_1-D_1} \Big(\Phi\Big(\frac{\Phi^{-1}(q)+Y\sqrt{\rho}}{\sqrt{1-\rho}} \Big)\Big)^{D_2} \Big(1-\Phi\Big(\frac{\Phi^{-1}(q)+Y\sqrt{\rho}}{\sqrt{1-\rho}} \Big)\Big)^{N_2-D_2} \Big]$$

For no defaults at all and correlation = 12%, the 95% confidence region for p and q (cut value for –2logLR = 5.99) is charted below. The conservative PDs are p=8% and q=20% approximately.



For 3 defaults in the low risk group and 2 defaults in the high risk group, the maximal likelihood estimates are p = 5.0% and q = 7.9% approximately. From the 95% confidence region plotted below, we read the conservative limits p <= 27% and q <= 39% approximately.



4.7 Table summary of example conservative PDs. With comparison with output that would result from [C] or [PT] methodology.

4.7.1 Single Grade

No correlation

Defaults/100	Likelihood approach (cut –2LogLR at 3.85 for D>0)	95% "Confidence Interval" Limit from [C]
0	3.0%	3.0%
1	4.4%	4.6%
5	10.4%	10.2%
10	16.9%	16.4%

Correlation 12%

Defaults/100 R=12%	Likelihood approach (cut –2LogLR at 3.85 for D>0)	95% "Confidence Interval" Limit from [C]
0	6.3%	6.3%
1	9.5%	10%
5	21.5%	20%
10	32%	29%

Thus for all levels of default the two methods agree quite closely.

4.7.2 Two Grades (70 low risk, 30 high risk)

No Correlation

Defaults Low risk, high risk	Likelihood approach (cut –2LogLR at 5.99) low risk PD , high risk PD	95% "Confidence Interval" Limit from [PT]
0, 0	3.0% , 9.8%	3.0% , 9.8%
3, 2	12% , 23%	NA

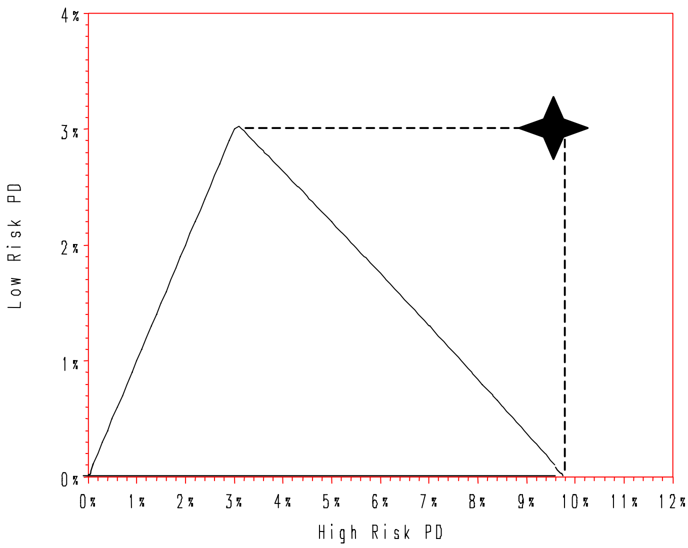
Note that, in this case, the two methods are identical where there are 0 defaults.

Correlation 12%

Defaults Low risk, high risk	Likelihood approach (cut –2LogLR at 5.99) low risk PD , high risk PD	95% "Confidence Interval" Limit from [PT]
0, 0	8% , 20%	NA
3, 2	27% , 39%	NA

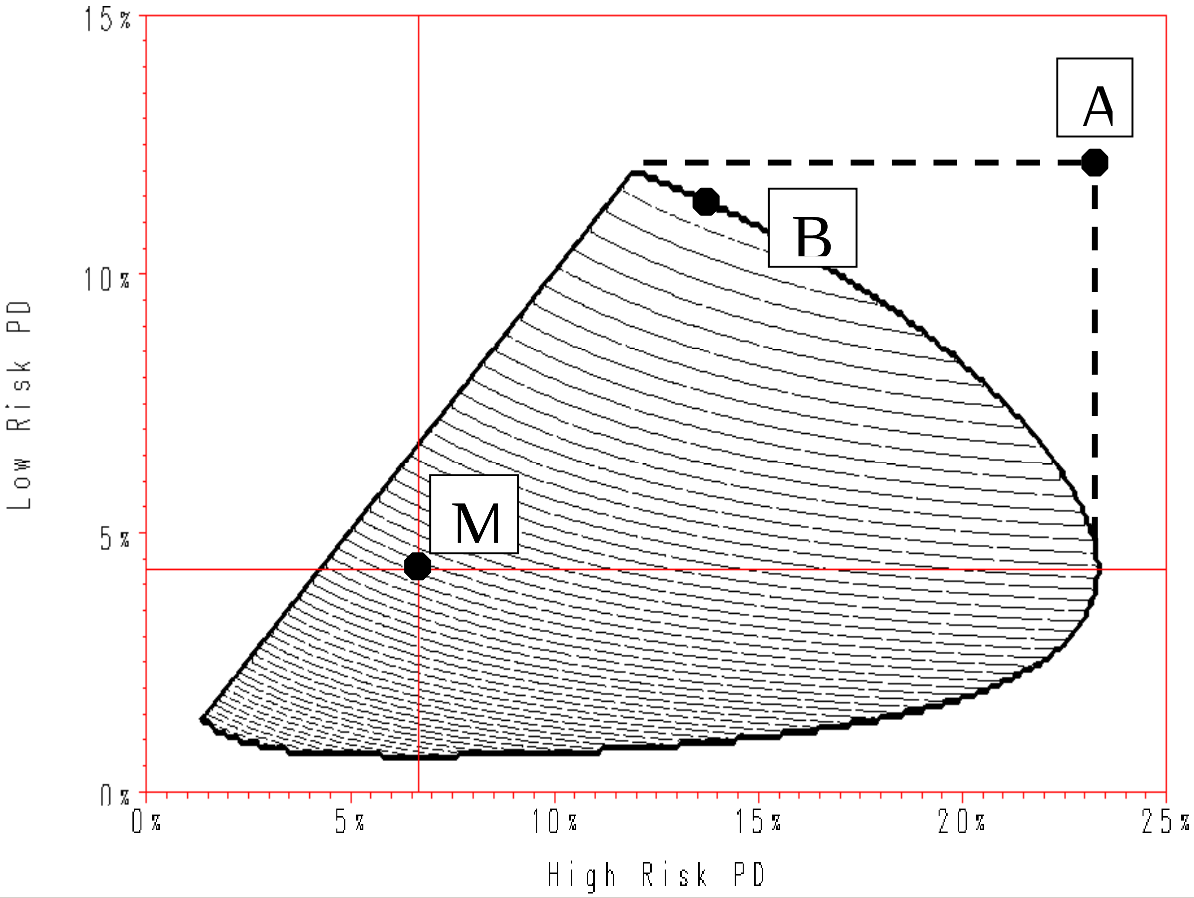
5. Better coordinated PDs.

5.1 Note that for examples 4.4-4.6, with two or more gradings, the conservative PDs are decided separately. The combined choice of conservative PDs lies well outside the confidence region. For example, in the following chart, from 4.4, the star indicates the combined conservative PDs. This is over conservative.



This combined choice is also the cause of inconsistencies, as can be seen in the tables of section 4.7. For example, compare the PDs arising from two grades with 3 and 2 defaults, with the PD arising from one grade with 5 defaults. Adding information about grades should reduce the PD of one of the grades at least. This doesn't happen with the choices of PD in the tables in 4.7.

5.2 Rather, we need a rule to pick combinations of PDs within the confidence region. Here we propose to choose the point in the confidence region which maximises the Basel II K factor. This is illustrated in the following diagram from example 4.5.



The diagram shows the 95% confidence region, with contours indicating the levels of Basel II Capital risk weight (assuming in this example that LGDs are all 10% and EADs are equal between the 100 relevant exposures). The diagram also marks the points of maximum likelihood (M) and two choices of conservative PDs. The first choice (A) is the point chosen in example 4.5 and which has the weaknesses pointed out in 5.1. The second choice (B) is the point where Risk Weight is maximised.

	High Risk PD	Low Risk PD	Risk Weight
Maximum Likelihood (M)	6.7%	4.3%	32.7%
Ultra-Conservative PD (A)	23.4%	12.1%	53.2%
RW maximal PD (B)	13.7%	11.4%	48.7%

It's clear that there are significant differences between the two choices of conservative PD.

5.3 Overall, the RW-maximal choice appears to be more acceptable and consistent with its expected use in Basel II calculations and in validation of risk gradings. In the letter use, this example, which shows little difference between choice (B) and the choice (12.1%) for equal PDs; so the grading is probably not validated by the data. This can be made more precise in terms of a hypothesis test.

6. Areas for development

6.1 **Cut Levels.** In this paper, accurate Likelihood Ratio cut levels from hypothesis testing are not attempted except in the 0 default case, although they are estimated in the one grade examples (see 3.3). These cut levels, their significance level and principle, will need to be justified by the business and theoretical context. The next step will be to issue clear guidance or definitive examples on these points, without restricting the modelling choices open to the professional.

6.2 **More sophisticated Priors.** Expert opinion can be gathered in forms other than gradings and this will be reflected in more subtle inequalities on PDs or in different kinds of Prior Odds functions.

7. References

[C] Alan Cathcart , Appendix to [CRSG]

[CRSG] Expert Group paper on Low Default Portfolios, FSA internal publication, June 2005

[PT] Katja Pluto, Dirk Tasche. Estimating Probabilities of Default on Low Default Portfolios, Deutsche Bundesbank publication, December 11, 2004.

[SH] Til Schuermann, Samuel Hanson. Estimating Probabilities of Default. Federal Reserve Bank of New York Staff Report no 190, July 2004.

[S] S.D Silvey. Statistical inference. Chapman and Hall, London, 1995.