Basel Compliant Modelling with Little or No Data

Synopsis

Traditionally, two types of credit risk models have been predominant in the industry: (a) statistical models constructed on fairly large data samples, generally used for for consumer credit and (b) expert models, generally used for corporate credit. With the advent of the Basel Accord the boundary has blurred because lenders have been forced to construct statistical models where numbers of defaulting accounts are low and because it is a regulatory requirement that expert models take account of actual risk experience even if the number of defaults is very small.

This talk describes how models can be constructed and validated for situations where there are few defaulting accounts and few write-offs. It highlights and compares the various approaches available and examines creative examples that have been implemented by lenders. We discuss some of the underlying principles of such modelling, we examine what the Basel regulators have to say on the matter and we suggest areas where more academic research would be useful to assist the credit practitioner.

This talk should be of interest to (a) mortgage lenders who, even if they have a reasonable number of defaulting accounts, may be short of information on property foreclosures, (b) corporate credit lenders, (d) lenders who wish to introduce a new product and (e) any other lender who has to model with small samples of accounts or with a low number of defaulting accounts.

In this paper we do not present a complete survey of all the approaches that are possible. Rather, we examine a sprinkling of model types, discussing what we have seen in practice and/or have adopted as best practice ourselves. This does not imply that we are not aware of the weaknesses in our approaches. In fact we welcome input from academia in all areas so that we can improve our methods the next time around.

Finally, we mention one area in particular that we believe may have been missed by the regulators leading to a potentially serious under-capitalisation for lenders who have a small number of obligors each with a large credit facility.

Contents

1. Purpose, Background and Summary

The purpose of this paper is simply to inform on methodologies to show the reader how small sample modelling for Basel has been tackled in practice.

There are many situations in credit when one is modelling a binary outcome and there are insufficient observations for one of the outcomes. Even if a lender has sufficient derogatory accounts for modelling it may have insufficient Defaults, Write-offs/ Possessions etc. further down the line.

What, then, does “insufficient” mean in this context? It is possible to define it more formally. For example one could set bounds on the sample size by insisting that “the chance of a particular realisation of a Gini Coefficient at below X% of its true value” is to be <5%. In practice the only way to determine this bound is to perform simulations. We recognise that some guidance from academia would be beneficial here. Also, the value of X depends on the cost of making incorrect decisions within the business. Again, guidance would be beneficial.

Generally, if the characteristics are continuous the the number of bad accounts required is fewer because the continuous variables, or functional forms of them, can be used directly in the model with fewer degrees of freedom than for discrete variables. That said, there are usually a fair sprinkling of discrete variables, particularly for application scorecards. It is now well known within the credit industry that the approach towards scorecard development should change as the number of bad accounts diminishes [Ref5]. In this paper we discuss modelling when there are fewer than 150 bad accounts.

We do not recommend a particular method, only give examples of approaches that we have see. The underpinning considerations though are:

1. Keeping the models as transparent as possible mathematically, whilst being sufficiently flexible to solve the problem in hand
2. Using Bayesian statistics to solve the small sample problems
3. Being inventive: every problem has at least one good solution

2. Principles

Ask a modern statistician for methods to deal with situations where

- there are small samples
- there needs to be an admixture of data and expertise, possibly because of
  - the small sample size,
  - poor data quality
  - some data variables not available

and he/she should suggest that Bayesian statistical techniques are used for modelling as these present the only consistent way of combining expertise and data.

We believe that this is indeed true and we present examples below where we have used such techniques. However, it does not display the whole picture as Bayesian methods are only combination approaches and do not specify the shape of the statistical model that is to be used (although there are Bayesian model comparison methods), the approach to be used to collect the data nor indeed do they fully define the the approach to be used to combine the data and expertise. Nonetheless, any technique that does not combine probabilities logically and which does not treat degrees of belief as probabilities (which is what the Bayesian approach does) seems to us to require much more justification.

3. Regulators Views

The financial regulators' contribution to the problems caused by small numbers of defaults (Low Default Portfolios – LDPs) has been to try to define exactly what 'small' means and then to provide guidance on how to calculate conservative confidence intervals for such. The conservatism is necessary within the EU's Capital Requirements Directive. There have been two key papers published on low default portfolios [Ref 1 and Ref 2]. The FSA's paper is non-binding, as it is there to be of assistance only. Nevertheless, we believe that the regulator requires this or an equivalent approach to be adopted.

It seems to us that both the UK regulator (the FSA) and the German regulator (the Bundesbank/Bafin) think through the technical issues very carefully and any paper that either of them produces is generally of a high technical quality. Having said this though, our views on LDPs are two-fold:

- The papers are too complex for lenders to understand and duplicate.
- A key element seems to have been been missed by the regulators.

We present a simpler calculation in Appendix A. This might seem mathematically complex but is, in fact, much simpler than the alternatives and produces a straightforward formula that can be used by anyone with a pocket calculator or a spreadsheet to calculate a conservative confidence level. We discuss the use of this calculation and compare it with the enlightening, but complex, Benjamin, Cathcart and Ryan [BCR] alternative offered by the FSA [ref 1].

The key reason for the use of our approach (which was derived before the BCR approach was formulated) is that the Basel Accord capital formulae are, believe it or not, deliberate simplifications to make it easy for the banks to calculate capital. Returning to, what is essentially a confidence interval method, that requires advanced mathematics to assimilate it does not seem to us to be in the spirit of the thrust of the Accord. We thus utilise the BCR paper to validate our own approach, reasoning that our approach is simpler for lenders to implement (even though the BCR approach may be more technically correct) , rather than use the BCR approach directly.

The various regulators are obviously concerned with what is a low-default portfolio and how it should be assessed. However, there is no logical reason why a concrete limit on the number of defaults should be set – the methods should be applicable regardless of this number – and the regulators do not provide guidance on statistical modelling when the number of defaults is low (too low for standard modelling) but above the minimum.

4. Some Statistical Issues

(a) Multiple Observations

One issue that occurs for any lender who is producing a behaviour scorecard using historic data is that it is possible to increase the sample size by taking multiple snapshots of the same account at different time periods. Indeed, we have adopted this approach when constructing such scorecards and when constructing Probability of Possession models for mortgages. The problem with this type of approach is that all regressions (linear or logistic) assume that the observations are independent. If this is not the case then one should ideally use some adjusted regression process to undertake the modelling.

We have recognised this problem but have not used special techniques. We discuss our reasoning. As a simplified case we assume that each observation has been sampled 3 times and that the 3 observations are 100% correlated. In this situation the result from any regression run would be identical to that from a run where the observations had been sampled once, if the regressions were allowed to maximise likelihood in an unconstrained fashion. There are only two issues:

- Most regression programs select variables based on statistical tests and the results of these tests would differ under the two situations
- The statistics on the parameters would also differ

Thus to use standard logistic regression in this situation the analyst would have to replace the automatic variable inclusion mechanism with his/her own experience and also make a personal judgment as to whether the final model is over-fitted or not. As experienced scorecard builders, the

Rhino Risk consultants feel confident in this situation but realise that those with less experience might wish to have a more formal statistical method and one that is easier to justify to the regulator (although the FSA does not seem to have seen this a particularly relevant issue, probably because they, like us, value experience above perfect accuracy).

(b) Changing the Outcome Definition

All modellers, including ourselves, adopt, where possible, a staged approach towards scorecard construction, building first a model using a longer outcome period and a less stringent outcome definition (e.g. 2+ in arrears for mortgages rather than the Default definition) and then calibrating the result to the required outcome period/definition. This is a very sensible approach. However, if the outcome period is very long it can generate difficulties as bias can creep in. For example, if one had a 3 year outcome period but had only lent to a particular class of customers (e.g. those with adverse credit history) in the last year then these customers would look better than average simply because they had not had the full 3 years in which to go bad. This issue can be resolved by one of three methods:

1. Only include observations with the the full outcome period
2. Weight the observations
3. Use Survival Analysis

The first option is possible but, when one is using longer outcome periods in order to cater for insufficient bad accounts, it does not then seem sensible to exclude the recent customers. We have used option 2 and option 3. We initially tried the Weighting option (2) but now prefer to use Survival Analysis as statisticians who have validated our models prefer it and record-weighting can cause various difficulties in the interpretation of confidence intervals.

It should be noted that various other options are available that have been insufficiently explored in the literature. One of these is to augment the outcome with credit bureau scores. Also, the use of a bureau score in the model itself can provide much confidence in the final result as such scores are known to be predictive under most situations (the only exception possibly being for sub-prime business)

5. Bayesian Models

We have assisted a large bank in the construction of their Basel PD mortgage models using Bayesian techniques of the type outlined in [ref 4]. The approach we have used is one that has been shown to work before but which could definitely be improved by expert statisticians. Ref4 describes the approach more fully but we briefly highlight it here. The method works where there are characteristics divided into attributes (for continuous variables a similar method based on functional forms is more appropriate):

1. Characteristics are selected that are reasonably uncorrelated [by examining a correlation matrix between the variables]
2. For each characteristic
    1. The goods and bads for each attribute are recorded as “g” and “b”
    2. An expert is asked for his/her views on the bad rate for each attribute
    3. These expert bad rates are converted to odds, factored and then converted back to bad rates again, the factor being chosen so that the overall bad rate is that from the sample being examined. The results for an attribute are “g_e” and “b_e”
    4. If the expert's views are in total disagreement with the data then a review is performed to ensure that he/she has been informed correctly of the underlying problem/ data and that the expert chosen is the most appropriate person in this instance.
    5. For each attribute the expert's bad rates are converted into Good and Bad accounts “G” and “B” where $G = \lambda g_e$, $B = \lambda b_e$ and λ is a factor chosen to weight the expert's contribution.
    6. A new bad rate for the attribute is calculated from the pooled bad accounts b+B and the pooled good accounts g+G
    7. Weights of evidence are calculated
3. For each original characteristic a new characteristics is formed where the value of the characteristic is the weight of evidence from the above calculation
4. To calculate the points for a particular customer the weights of evidence are simply added as

we know that, from step 1, the characteristics are uncorrelated.

For this approach to be successful a number of assumptions have to be justified

1. The selection of the expert has to be justified
2. The parameter λ has to be selected as a balance to ensure that
   1. the expert's views are not over-valued
   2. random fluctuations in the data are not given prominence

In the above approach the choice of λ can be made to depend on the sample size. We have typically assumed that if the sample defaults are very small for an attribute then the expert will have had less experience in proportion. Thus the value of λ should decrease.

Although we have had some success with the above approach we would much prefer it if

- There was more statistical support in the literature for how to weight the expert's knowledge
- We could integrate functional forms into the approach rather than having to coarse-class continuous characteristics.

We would hope that our next venture into this type of modelling will have a more secure foundation in these two areas. Nonetheless, the approach is practical and can easily be understood and implemented. The expert views can also be replaced by information from relevant external data sources.

6. 'Distributional Assumption' Models

For modelling the "probability of possession (property foreclosure) given default" for mortgages (a key component of LGD) we have come across situations when the lender has no possessions, or very few possessions, through a combination of the following circumstances:

- The recent benign economic climate with low interest rates and rapidly increasing house prices has meant that there have been very few property possessions
- Lenders have often not recorded information historically

Even in the no-possession situation we have been able to devise a sensible approach towards model construction. This approach is based on some assumptions about the causal mechanism for a possession, on disaggregating the underlying distributions and on making some assumptions about the form of these distributions. We explain this below:

One key variable at least – Loan to Value ratio - must be used for the modelling of Probability of Possession given Default. There is a logical reason why this might be expected to be a very relevant feature of the Probability of Possession assessment. It is not to the financial advantage of customers who have minimal equity in their property to concern themselves with avoiding repossession. In fact, the use of LTV as a key variable is consistent with the Option-Theoretic approach used for valuing debt, whereby equity holders (the customers, in this case) are considered to have a "PUT-option" on the value of the collateral where the "strike price" is the debt. The rationale is that if the asset value falls below the value of the debt the equity holders are better off by giving the assets to the lenders rather than repaying the debt. *This Option-theoretic approach, developed by Merton, is used by MoodysKMV in their credit default risk models.*

If there is insufficient data to obtain the relationship with LTV by adopting a forward forecasting approach then the following approach is an alternative, based on a variation of Baye's formula. The formula assumes that an initial primary sample is selected from Defaults. If insufficient defaults exist it will have to be selected from the accounts most severely in arrears:

$$\textbf{Odds(Possession|LTV)} = \frac{\textbf{p(LTV|Possession)}}{\textbf{p(LTV|non-Possession)}} \times \textbf{Odds(Possession)}$$

The Odds of Possession (given default) can thus be calculated for each LTV if the following are known:

(1) The LTV distribution for Possessions

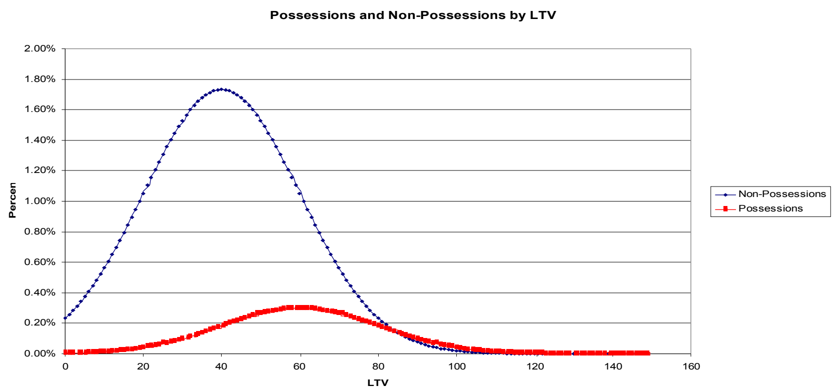
(2) The LTV distribution for the non-Possessions [that are defaults]
(3) The Odds of Possession

The key aspect of this approach is that the above three pieces of information do not have to emanate from the same source. The first of these can be obtained by examining historic Possessions from external sources. The second can be obtained from extant defaults. The last can be a recessionary-based forecast [e.g. from the Council of Mortgage Lenders] or a current figure as appropriate for the scenario being tested. The graphs below illustrate the calculation. If full data is not available then distributional assumptions can be made (e.g. the distribution of (1) can be assumed to be normal, with mean calculated from external data and standard deviation the same as that from (2)).

The first graph shows the two LTV distributions [for Possessions(1) and non-Possessions(2)] with the former suitably scaled for the Probability of Possession(3). This scaling is done so that the area under the Possessions curve divided by the sum of the areas under the two curves is equal to the possession rate.

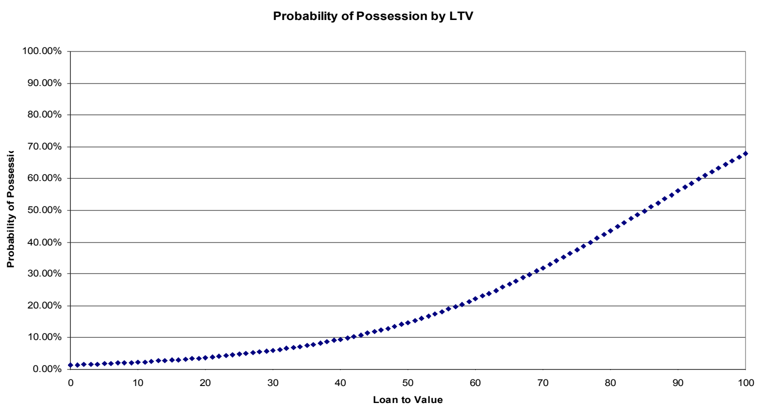
The second graph shows the Possession Rate by LTV calculated by examining graph 1 and then dividing the Possessions by the sum of the Possessions and non-Possessions for each LTV level.

Probability of Possession Calculation by LTV



Rhino Risk

Probability of Possession by LTV



36

We have thus been able to construct a logical model based on disparate data sources and some distributional assumptions. Such a model can be installed and then replaced with a more empirical version as data becomes available. The above approach can be used to determine the effect of stressed house prices by recalculating what the LTV would be at the time of Possession, rather than using a current LTV.

7. Causal Models

For calculating shortfall for mortgages we have used models that have more structure in them. These models can be developed from data sources or, if no data is available, can be devised to be logically sensible and consistent. We describe a typical approach below. All that is required for the calculation is a model (e.g. a scorecard or a segmentation) to predict the mean haircut (price reduction at sale over what has been estimated) for the property, a haircut distribution and various parameters (e.g. future estimated house price movements).

The Shortfall model for a mortgage lender measures the expected loss given the property has been repossessed. This calculation is sensitive to the house price that is being used in the stressed or unstressed scenario that the lender is considering. The model is explained without discounting to NPV to make the explanation simpler to understand.

We assume in the following example that:

The Exposure at Sale (this includes costs and interest) is £65,000
The Property Sale Price is £75,000
The Indexed Sale Price is £100,000 from available house price indices

(Once the property has been sold the Indexed Sale Price can be calculated from the Nationwide or Halifax regional house price indices. For properties that have not yet reached the Sale point, or even the Default or Possession point, the Indexed Sale Price has to be replaced by an Estimated Index Sale Price which consists of the indexed valuation at the current date plus a forecast up to the point of sale)

Because all costs are included in the Exposure at Sale, the bank will make a loss for a customer if the

Exposure at Sale exceeds the Property Sale Price
(e.g. If £65,000 > £75,000 – *which it is not and so there is no loss*).

If we divide both sides of the above inequality by the Indexed Sale Price then we get the inequality:

$$\frac{\text{Exposure at Sale}}{\text{Indexed Sale Price}} > \frac{\text{Property Sale Price}}{\text{Indexed Sale Price}}$$

This can be expressed as

LTV_at_Sale > mean_haircut (e.g. 65% > 75% - *which is in fact false*) (1)

The haircut will generally be less than 100% as repossessed properties do not usually achieve their market value at sale. *Note: our use of the term 'haircut' is different from the standard usage, in that, in the example above, the haircut would normally be regarded as 25%.*

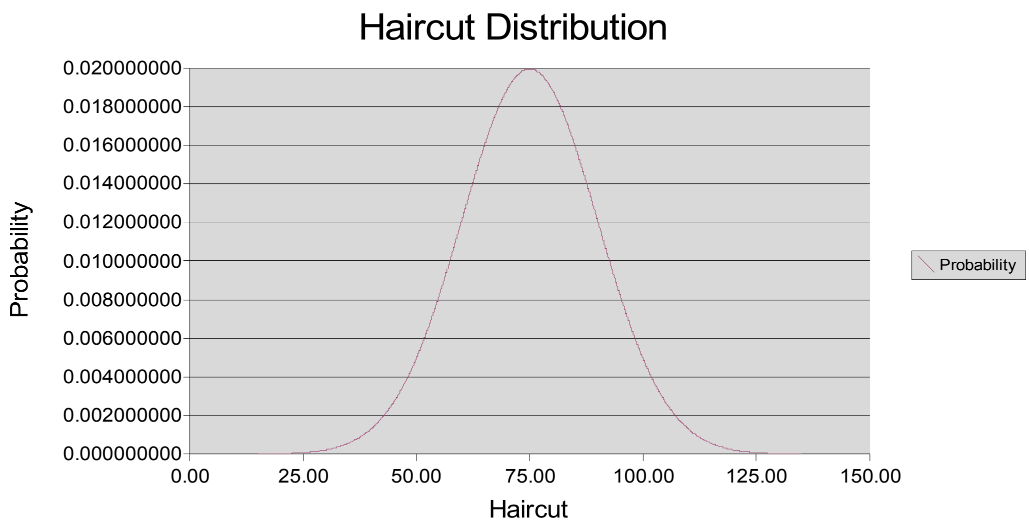
For the Expected Shortfall model the mean haircut can be calculated separately for each property using a specially developed scorecard. Thus, when the shortfall model is applied to the bank's retail mortgage book, each property will have its own separate mean haircut as well as its own separate LTV (i.e. LTV_at_Sale).

However, generally when we apply the shortfall calculation we are looking into the future up to the point of distressed sale of the property (should it occur) . We thus do not know two key features for the shortfall calculation:

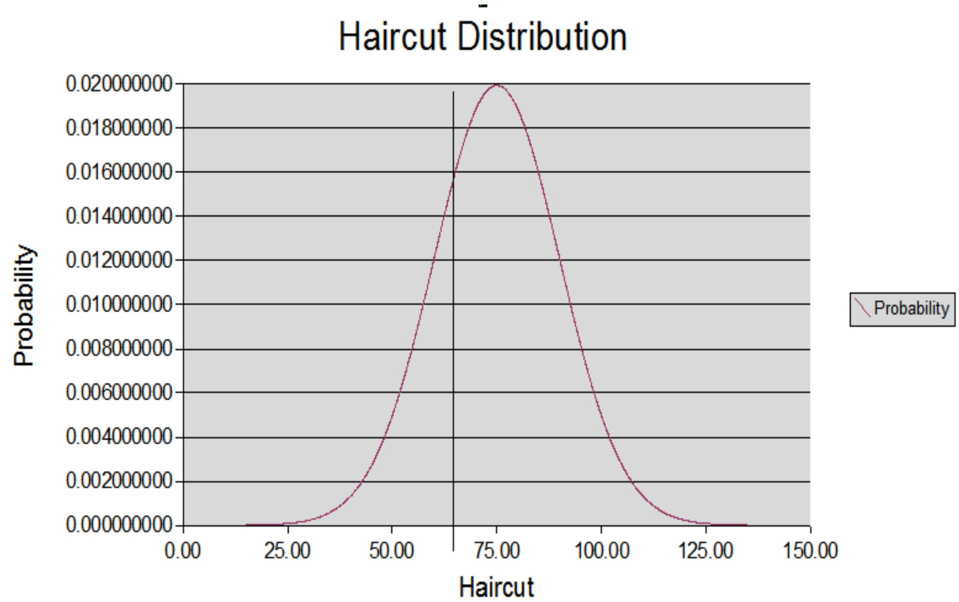
1. We do not know the indexed sale price which we require to calculate the LTV_at_Sale
2. We do not know the haircut.

The first of these we address by using the indexed property value up to the current date and then using a forecast up to the point of Sale. The severity of the forecast will be scenario-dependant and will depend on the stressed scenario that we are using.

For the haircut we can immediately see that we would not know the haircut even if the property were to be sold today as, without sight of the property and without a recent valuation, its state would be an unknown. We represent this lack of knowledge by a probability distribution. Each haircut has a probability of occurring. This probability, for each possible haircut, is called the "haircut probability distribution" [strictly speaking statistically this is the *haircut density function*] . The curve below shows the haircut distribution for a particular property. The mean haircut is 75% in this example:



We place our LTV line of 65% on this graph to get:



We now calculate the probability of the haircut being below the fixed LTV line. This probability is calculated as the area under the haircut distribution to the left of the LTV line. Thus, in distinction from equation (1), when we assume a probabilistic evaluation of the haircut then there is a positive probability of there being a loss (because this area is > 0). We show how this loss is calculated in the Excel table in Appendix B. We first define the columns in the table:

	Excel Column	*Definition*
1	Number of Standard Deviations from the Mean	A value of -4 means that the haircut is 4 standard deviations below the mean
2	Probability	This is the probability of the haircut occurring.

	Excel Column	*Definition*
3	Haircut	The haircut is calculated as the number of standard deviations from the mean. As an example, at -3 standard deviations (STD=15) from the mean (75) the the haircut is 75 - 3 x 15 = 30
4	Percentage Loss	This is the LTV (65%) - Haircut **(col 3)**
5	Percentage Loss x Probability	This is **(col 4)** x **(col 2)**
6	Percentage Loss x Probability if (LTV > Haircut)	This is **(col 5)**, but set to 0 when the LTV is greater than the haircut

Notes:

1. The Mean Haircut is calculated at collateral level from a Mean Haircut Scorecard In this example it is assumed to be 75%
2. The Standard Deviation is assumed to be 15 in this example.
3. The LTV is assumed to be 65%.
4. The Estimated Indexed Sale Price is £100,000.

For each haircut, the probability of loss multiplied by the probability of it occurring is shown in column 5. When we sum column 5 we obtain the total of -10. This is correct because the LTV is 65 and the average haircut is 75 so the answer should be 65-75 = -10. However, in situations where there is a gain (the Haircut is greater than the LTV) we must reimburse the customer. We thus set the value of all these occurrences to zero (column 6) and then recalculate the column sum to give a loss of 2.266% (the total of col 6). When we multiply this by the Estimated Indexed Sale Price of £100,000 we get the Expected Shortfall of £2,266. This is the average shortfall that we would expect if this scenario were repeated indefinitely.

8. Expert Models

Another type of model construction is often used for corporate lending where direct extraction of knowledge from long-time experts is used:

1. To form rules
2. To make adjustments to external models (such as MoodysKMV's RiskCalc or CreditEdge)
3. To form scorecards

We will discuss the scorecard formation here.

Most scorecard construction consists of the following steps:

1. Form the characteristics
2. Decide on the attributes for each characteristic
3. Decide on the points for each attribute so that X points represents a doubling of the odds
4. Provide a weight for each characteristic to represent correlations between the variables.

In this scorecard the final points for an attribute is thus the starting points * the characteristic weight.

This approach is laudable. However, it has a potential drawback:

- When the expert devises the points for the attributes he/she is essentially stating what the expected default rate is for that attribute. However, in the final scorecard the predicted default rate per attribute may differ from that chosen by the expert owing to the correlations in the data. Ideally, the expert needs to review the final result and then make adjustments to ensure that the final assessment is in line with his initial view.

There are two approaches to solving this problem:

- The first is for the modeller to assist the expert in making the necessary adjustments (bearing in mind the correlation matrix).

- The second approach is a mathematical one (that we have checked the logic of but not yet applied in practice). The essence is quite simple. Given the correlations in the data and the final views on the bad rates per attribute from the expert, it is possible to automatically generate the points that would have given these views. The mathematics is just a matrix inversion similar to the one underpinning linear regression.

9. External Models

Generally, when the sample size for one of the outcome categories (e.g. Default) is low a good place to look for assistance is the use of external data. MoodysKMV provide two models that may be used for limited companies and non-limited companies respectively: CreditEdge and RiskCalc. CreditEdge is a Merton-style model measuring 'distance to default' and RiskCalc is a standard scorecard based on information from company accounts. If a lender intends to use either of these models it must convince the regulator that the population on which the models will be applied is similar to that on which they have been developed and/or tested. This is often not an easy task and requires careful arguments and detailed analysis.

Once the models have been applied then conservative confidence intervals can be calculated both for the overall default rate and for the strength of the relationship between score and default.

10. Models using Simulations

One approach when there is no data is to create a structural model, provide distributions for the key variables of interest and then run a simulation, counting up the number of times that the customer defaults. This approach is perfectly valid, is used for specialised lending facilities and we have used it ourselves for LGD modelling and for stress testing. It works even when there are no defaults but its power is related to (a) the extent to which the structural model is a correct representation of reality and (b) the extent to which the parameters underpinning the model represent future situations.

We have used such an approach for stress testing. This was based on calculating changing affordability under future economic scenarios.

11. Markov chains Models

One simple under-rated approach, that is subsumed in the simulation methods, is to use migration matrices (alias 'Markov Chains' or 'Roll-Rates') to observe obligors moving from one state to another and then through to eventual default. The matrices can be constructed from experience, common sense and historical data and then used to provide eventual default rate estimates. In fact, we have used such an approach for a lender to predict expected loss.

12. Comments on Validating and Monitoring

When there are a reasonable number of bad accounts validating ones models at the time of development and monitoring them subsequently can be based on the following principle:

> A model works correctly if it predicts the correct outcome on any subset of the data.

This monitoring principle is independent of the type of model produced and thus is very useful for internal audit departments when checking out the Basel Validation.

If a subset can be found where the predicted default rate is not equal to the actual default rate (by a statistically significant amount) then the model is not optimal and the modeller could be asked to improve it.

The above approach also works when there are few defaults. However, the lower the number of defaults the bigger the requirement for a qualitative validation assessment as well as a quantitative one. Such an assessment should bring in as much external data as possible to support the case.

Note: Any decline in Gini coefficients should, of course, be investigated.

For low default portfolios, as well as examining subsets, the strength of the relationship between score and odds can be tested. If the score x log(odds) relationship is linear then the slope and intercept can be tested based on confidence intervals obtained when the model was developed.

13. Implementation Issues

The following issues are important because they relate jointly to the development sample and the model implementation and are a feature of low default portfolios.

1. If one is testing a Gini coefficient that was obtained from a development sample where the number of defaults was low then it is likely that, on implementation, the number of defaults will also be low - in which case it is insufficient to assume that the development sample Gini value is the true value and then to test against it. Instead one should test for differences (or ratios) in Gini values. Such a test requires a simulation. The table below shows a typical joint distribution of validation (test) sample (vertical) and development (main) sample (horizontal) simulated 10,000 times. To calculate a 95% 1-tailed confidence interval from such a table one has to count the occurrences where (Main Gini – Test Gini)< 5%

Gini Analysis																										
Test sample		Main sample																								
		31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	Total
		0	0	1	3	11	36	104	250	508	874	1272	1563	1625	1428	1061	666	354	159	60	19	5	1	0	0	10000
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
26	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
27	4	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	4
28	7	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	7
29	13	0	0	0	0	0	0	0	0	1	1	2	2	2	2	1	1	0	0	0	0	0	0	0	0	13
30	24	0	0	0	0	0	0	0	1	1	2	3	4	4	3	3	2	1	0	0	0	0	0	0	0	24
31	41	0	0	0	0	0	0	0	1	2	4	5	6	7	6	4	3	1	1	0	0	0	0	0	0	41
32	67	0	0	0	0	0	0	1	2	3	6	8	10	11	10	7	4	2	1	0	0	0	0	0	0	67
33	104	0	0	0	0	0	0	1	3	5	9	13	16	17	15	11	7	4	2	1	0	0	0	0	0	104
34	157	0	0	0	0	0	1	2	4	8	14	20	24	25	22	17	10	6	2	1	0	0	0	0	0	157
35	225	0	0	0	0	0	1	2	6	11	20	29	35	37	32	24	15	8	4	1	0	0	0	0	0	225
36	309	0	0	0	0	0	1	3	8	16	27	39	48	50	44	33	21	11	5	2	1	0	0	0	0	309
37	406	0	0	0	0	0	1	4	10	21	36	52	64	66	58	43	27	14	6	2	1	0	0	0	0	407
38	512	0	0	0	0	1	2	5	13	26	45	65	80	83	73	54	34	18	8	3	1	0	0	0	0	512
39	616	0	0	0	0	1	2	6	15	31	54	78	96	100	88	65	41	22	10	4	1	0	0	0	0	616
40	710	0	0	0	0	1	3	7	18	36	62	90	111	115	101	75	47	25	11	4	1	0	0	0	0	710
41	783	0	0	0	0	1	3	8	20	40	68	100	122	127	112	83	52	28	12	5	2	0	0	0	0	783
42	827	0	0	0	0	1	3	9	21	42	72	105	129	134	118	88	55	29	13	5	2	0	0	0	0	827
43	835	0	0	0	0	1	3	9	21	42	73	106	131	136	119	89	56	30	13	5	2	0	0	0	0	835
44	807	0	0	0	0	1	3	8	20	41	71	103	126	131	115	86	54	29	13	5	2	0	0	0	0	807
45	747	0	0	0	0	1	3	8	19	38	65	95	117	121	107	79	50	26	12	4	1	0	0	0	0	747
46	661	0	0	0	0	1	2	7	17	34	58	84	103	107	94	70	44	23	10	4	1	0	0	0	0	661
47	560	0	0	0	0	1	2	6	14	28	49	71	88	91	80	59	37	20	9	3	1	0	0	0	0	560
48	454	0	0	0	0	0	2	5	11	23	40	58	71	74	65	48	30	16	7	3	1	0	0	0	0	454
49	353	0	0	0	0	0	1	4	9	18	31	45	55	57	50	37	23	12	6	2	1	0	0	0	0	353
50	262	0	0	0	0	0	1	3	7	13	23	33	41	43	37	28	17	9	4	2	1	0	0	0	0	262
51	186	0	0	0	0	0	1	2	5	9	16	24	29	30	27	20	12	7	3	1	0	0	0	0	0	186
52	127	0	0	0	0	0	0	1	3	6	11	16	20	21	18	13	8	4	2	1	0	0	0	0	0	127
53	82	0	0	0	0	0	0	1	2	4	7	10	13	13	12	9	5	3	1	0	0	0	0	0	0	82
54	51	0	0	0	0	0	0	1	1	3	4	7	8	8	7	5	3	2	1	0	0	0	0	0	0	51
55	31	0	0	0	0	0	0	0	1	2	3	4	5	5	4	3	2	1	0	0	0	0	0	0	0	31
56	17	0	0	0	0	0	0	0	0	1	2	2	3	3	2	2	1	1	0	0	0	0	0	0	0	17
57	10	0	0	0	0	0	0	0	0	0	1	1	1	2	1	1	1	0	0	0	0	0	0	0	0	10
58	5	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	5
59	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
60	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
61	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Tot	10000																									10000

2. The second point is a crucial one that, it seems to me, has been missed by the regulators. It applies particularly when there are a handful of defaults – say 2,3,4 etc. - such as one might have for on a corporate portfolio of limited companies. The key point is that, according to current regulations, capital is only required to be calculated based on a conservative view of the default rate that has been obtained from historic data when developing the model. In reality the conservative calculation of default rate should depend on both:
   1. the number of defaults from the development sample
   2. the size of the roll-out sample

   This is because even if one knew with certainty what the population historic default rate was, and even if the future were exactly like the past, implementing on a small sample could involve high default rates purely by chance (based on the binomial distribution) and this extra risk should (but does not at the moment) attract extra regulatory capital.

The second issue here is extremely important and means that most banks that have a small number of large corporate loans will be under-capitalised. This is not a minor adjustment but will have a major impact on capital levels. We welcome views on this issue. Our realisation that there a genuine issue here is supported by our understanding of conservative PD levels allowed for different corporate portfolios in the UK and abroad) .

Appendix A: A "Simple" "Likelihood Ratio" Conservative Default Rate Formula

The Problem

We assume that we have observed "m" defaults out of "M" accounts. We assume that all the accounts have been open for one year so that they have all had an opportunity to default. We also assume that there is no information that can guide us in estimating the default rate other than the observed value.

The requirement is to estimate a conservative default rate.

The Solution

Obviously the most probable estimate given the data is m/M {this is called the Maximum Likelihood estimator}. However, we are required by the Basel Accord to take a conservative view when the number of defaults is low. We recommend that a default rate be chosen which exceeds m/M and is 8 times less likely than it. This will provide a sufficiently conservative value.

Our key assumption is that the observations are independent (that is, the probability of one customer defaulting is not influenced by other customers).

We then solve the equation

$$m*LN[m/d] + (d-m) = 2$$

for d, using trial and error or the Excel goal-seek functionality (*note LN is the natural logarithm to the base e*).

The conservative default rate is then d/M.

Explanation

The solution uses Bayes factors. We construct two hypotheses for the true default rate and we call these hypotheses H1 and H2. Here H2 represents the observed default rate and H1 another default rate that we are testing. We compare the ratio:

$$\frac{Prob(H2 \mid data)}{Prob(H1 \mid data)}$$

where data represents the observed outcomes, and select the default rate for the hypothesis H1 which exceeds that for H2 and makes H1 8 times less likely than H2. This default rate is then taken as the Basel conservative value. 8 is chosen as a sensible, but arbitrary, value.

We then have to find the default rate in H1 for which

$$\frac{Prob(H2 \mid data)}{Prob(H1 \mid data)} = 8$$

We note that using Baye's formula we have:

$Prob(H1 \mid data)\ Prob(data) = Prob(data \mid H1)\ Prob(H1)$ and

$Prob(H2 \mid data)\ Prob(data) = Prob(data \mid H2)\ Prob(H2)$

Substituting these into the ratio we get:

$$\frac{Prob(data \mid H2) \times Prob(H2)}{Prob(data \mid H1) \times Prob(H1)}$$

If we adopt the prior hypothesis (before seeing the data) that H1 and H2 are equally likely then $Prob(H1) = Prob(H2)$ and so our formula becomes:

$$\frac{Prob(data \mid H2)}{Prob(data \mid H1)} = 8$$

Using the above approach we compare the following two hypotheses:

H1: the default rate is a value of d/M

H2: the default rate is the maximum likelihood observed estimate of m/M

Under H1 the actual number of defaults will then follow a binomial distribution with mean d. We are effecively saying that if d/M is the correct default rate then the probability of observing exactly m defaults will be :

$$Prob(data \mid H1) = k\,(d/M)^{m}\,(1-d/M)^{(M-m)}$$

where k is the number of ways of selecting m defaults from M; i.e. K = M!/(m! X (M-m)!)

Under H2, if the default rate were m/M (the observed value) then the probability of observing exactly m defaults would be:

$$Prob(data \mid H2) = k\,(m/M)^{m}\,(1-m/M)^{(M-m)}$$

Dividing the second equation by by the first and manipulating we get:

$$\frac{Prob(data \mid H2)}{Prob(data \mid H1)} = (m/d)^{m}\,[(M-m)/(M-d)]^{(M-m)}$$

As mentioned, we arbitrarily set a confidence limit where this likelihood ratio is 8. Thus, d/M being the true default rate is eight times less likely than m/M being the true default rate. So we have

$$(m/d)^{m}\,[(M-m)/(M-d)]^{(M-m)} = 8$$

and we then have to solve this for d, given m and M.

we take logarithms to get:

$$m*LN[m/d] + (M-m)*LN[(M-m)/(M-d)] = 2.08 \quad (A)$$ (*note LN(8) = 2.08*)

We next note that $LN(1+x) \approx x$ for all x when x is small (B) ($\approx$ means approximately equal)

This relationship is derived from the standard Taylor Series expansion for LN which is

$$LN(1+x) = x - (1/2)x^2 + (1/3)x^3 - \ldots$$

We now express (M-m)/(M-d) as [(M-d) + (d-m)]/(M-d), which, dividing out, we can then write as

1 + (d-m)/(M-d).

We then set x to be (d-m)/(M-d) and apply formula (B) to get:

$$LN[(M-m)/(M-d)] \approx (d-m)/(M-d)$$

Our formula (A) then becomes

$$m*LN[m/d] + (M-m)*(d-m)/(M-d) \approx 2.08$$

Given that d and m are small compared to M we can approximate (M-m)/(M-d) by setting it to 1 to get

$$m*LN[m/d] + (d-m) \approx 2.08$$

Rounding, the formula then becomes:

$$m*LN[m/d] + (d-m) = 2$$

We note that the formula is now independent of M.

Example Uses

Case 1

If we observe 0 defaults (m=0) then the conservative value is 2. we obtain this by solving for d in the above formula, noting that m*LN[m/d] tends to 0 as m approaches 0.

Thus, if we observe 0 defaults from 100 accounts then a conservative default rate (in the absence of any other information) is 2%. If we observe 0 defaults from 1000 accounts then a conservative default rate is 0.2%.

Case 2

If we observe 3 defaults then a conservative number is 8, as when m=2 then d=8

Case 3

If we observe 30 defaults out of 1000 (3% default rate) then a conservative value is 42.6 out of 1000 = 4.26% default rate. This is slightly more conservative than the standard confidence interval approach where, by approximating the binomial distribution by the normal distribution we obtain a default rate of 4.06% (note, this standard approach is only recommended in the statistics books when the number of observed defaults m exceeds 30 as it fails as an approximation for smaller m).

Discussion

The above approach adopted makes the following assumptions:

1. There are no correlations between obligors
2. The correlations between years that are induced by the economy (alias the longitudinal correlations) are adequately catered for by the capital calculation formula embedded within the Basel Accord and hence do not need to be separately examined

The method is founded on the direct assessment of the underlying probabilities. It assesses the probability of the true default rate being X against it being Y given the knowledge that the observed value from the development sample is Z. It examines all values from 0,1,2,... upwards and draws a line where the probability of occurrence of the true rate is one eighth of the probability of occurrence of the most likely default rate.

The method above seems very natural. In fact it is much more natural than the traditional confidence interval approach found in standard statistical textbooks. The traditional approach is as follows:

1. One first assumes that the true default rate is the observed value Z (even though one knows that it is not). One then assesses the probability that the default rate will be greater than X in a new trial
2. One then takes as an answer the minimum value of X such that this probability is less than 5%; This gives the so-called 95% confidence interval (a one-tailed test in this case).

It is immediately obvious what is wrong with this traditional approach:

1. The true default rate is not necessarily the observed value.
2. The method does not work for small samples of defaults. In an extreme, if the observed default rate were 0 (because, by chance, no defaults were observed) then the probability that the default rate is > 0 will be 0! This is an obviously incorrect calculation as, however small the unknown true default rate is, we cannot ascertain with certainty that is zero.

In fact, owing to the nonsensical assumption in the first step, many modern statisticians prefer the likelihood ratio approach as, for each X, it is designed to directly calculate the chance that the probability of default is equal to X.

In the likelihood ratio above we have made an implicit assumption. This is that there we have no prior knowledge that the default rate is more likely to be one value rather than another. We could have incorporated some prior beliefs into the formulation. Doing so would have turned the "likelihood ratio" approach into what is denoted a "Bayesian approach", where judgment and data are combined to produce a 'best' estimate. In this instance we have deliberately avoided incorporating prior beliefs for the following reasons:

1. The beliefs would require justification
2. Incorporating beliefs would lower the conservative estimates; this thus seems against the spirit of conservatism
3. The mathematical calculations would be more complex

Our exclusion of prior beliefs has the implicit impact that, before the observed data is seen, all possibilities are deemed to be equally likely.

In all the calculations we have used, we have adopted, insofar as it is possible, the principle of transparency, in that where a complex solution and a simple solution compete we have selected the simple solution. This is not just because it is less onerous to understand, but also because simpler solutions are more communicable to non-technical senior management, are less error-prone and require less maintenance.

It does not seem sensible to complicate matters unnecessarily - e.g. by picking apart the very capital formula that was specifically designed for simplicity. Our conservative calculation avoids such a deconstruction. The Basel regulatory capital formula is a coarse approach towards assigning regulatory capital. It does not therefore make sense to be too refined in calculating conservative levels.

For Comparison: The Benjamin, Cathcart, Ryan [BCR] Approach

The BCR paper does not determine a fully prescriptive calculation. However, by adopting the logic in the appendix to the paper one can calculate a definite confidence interval based upon a number of parameters.

The BCR paper takes a full look at conservatism. It starts from a position of strength in that there had been a number of CRSG papers that had discussed the issue. Also, Pluto and Tasche from the Deutche Bundesbank [Ref2] had produced a paper examining the subject in depth. This paper is referenced within the BCR paper and the base confidence interval approach proposed by Pluto and Tasche is used by BCR. Another offering referred to is by Alan Forrest , formerly of Dunfermline Building Society, now at HBOS. The Alan Forrest paper [ref 3] is of interest because it is a likelihood approach similar in principle to the one we recommend.

The BCR paper purports to calculate confidence intervals correctly when there few defaults as it does not a calculate confidence interval directly from the observed default rate. It adopts the following calculation approach to determine confidence intervals:

> Assuming that the observed default rate is **d/n** (equal to **d** defaults from **n** accounts), the method finds the maximum default rate **p** such that if this default rate were true then the probability of observing less than or equal to d defaults from n is less than 5% (*for a 95% confidence interval*).

Although this is a twist on a standard confidence interval calculation we still find this rather unnatural. It is not clear, without cogitation, what it is trying to achieve and, when one works it out it is not clear, even with cogitation, why default rates less than **d/n** are relevant in any way. However, this may be somewhat unfair, as we admit that we have a personal bias towards likelihood and Bayesian approaches and we recognise that the BCR and Pluto/Tasche logic, though processes and understanding are deeper than ours.

The BCR paper separates has the following components:

1. It directly addresses the internal asset correlations (the non-longitudinal correlations). It does this by utilizing the Vasicek formula (used in the Basel capital calculation) to adjust PD estimates to cater for the correlations
2. It has a component handling the longitudinal correlations. The correlations between any two consecutive years are assumed to be independent of those between any other two consecutive years
3. It requires a combination of a simulation and goal-seeking to find an answer
4. It requires the following parameters:
    1. A confidence level
    2. An asset correlation value
    3. A longitudinal (year to year) correlation value
    4. A base number of accounts
    5. The number of observed defaults

Comparison of the two methods

Likelihood Method	BCR Method
Simple	Complex
No simulation or goal-seeking required	Simulation and goal-seeking required
Works accurately for very small numbers of defaults	To be ascertained whether it works accurately for small numbers of defaults
Self-contained	Is dependent on knowledge of 1. The Vasicek model 2. The paper by Pluto and Tasche 3. Advance mathematics and statistics
Assumes that there are no direct asset correlations (at a point in time), as the business is consumer-based, and lets the capital calculation formula handle longitudinal correlations	Handles asset/longitudinal correlations by utilising the mathematics behind the capital calculation formula.
The calculation of the conservative estimate can be done in advance and the computer calculation time is independent of the number of defaults. Thus the calculation can be performed on any sample size with no overheads in calculation time.	The calculation is not possible for a large number of defaults and so a barrier has to be set for what is (and is not) a low-default-portfolio. Indeed, even for quite low numbers of defaults, the size of the simulation has to be restricted to a non-ideal level.
The only parameters are: 1. Confidence level 2. Number of defaults The answers can be pre-calculated and the implementation can be table-driven.	There are six main parameters: 1. Confidence level 2. Number of defaults 3. Number of accounts 4. Asset correlation 5. Matrix of longitudinal correlations 6. Number of years This number of inputs precludes pre-calculation of the answers and so the generalized method cannot easily be implemented in a SQL-based environment.

In conclusion we prefer to implement the likelihood ratio approach and to use the BCR method to check against. The likelihood method is much simpler and easier to calculate and convey to senior management. In addition, although the BCR approach is very useful as a checking mechanism, we have reservations about the necessity of implementing it, as it is much more complex than the Capital formula; this is because we believe that the capital formula was derived to make it relatively easy for banks to apply and so it does not make sense to subsequently complicate matters.

Appendix B: Shortfall Tables

Standard Deviations from the Mean	Probability	Haircut	%Loss	%Loss x Probability	%Loss x Probability if (LTV>Haircut)
-4	0.0007%	15.00	50.00	0.00033	0.00033
-3.95	0.0008%	15.75	49.25	0.00040	0.00040
-3.9	0.0010%	16.50	48.50	0.00048	0.00048
-3.85	0.0012%	17.25	47.75	0.00058	0.00058
-3.8	0.0015%	18.00	47.00	0.00069	0.00069
-3.75	0.0018%	18.75	46.25	0.00082	0.00082
-3.7	0.0021%	19.50	45.50	0.00097	0.00097
-3.65	0.0026%	20.25	44.75	0.00114	0.00114
-3.6	0.0031%	21.00	44.00	0.00135	0.00135
-3.55	0.0037%	21.75	43.25	0.00158	0.00158
-3.5	0.0044%	22.50	42.50	0.00185	0.00185
-3.45	0.0052%	23.25	41.75	0.00217	0.00217
-3.4	0.0062%	24.00	41.00	0.00253	0.00253
-3.35	0.0073%	24.75	40.25	0.00294	0.00294
-3.3	0.0086%	25.50	39.50	0.00340	0.00340
-3.25	0.0101%	26.25	38.75	0.00393	0.00393
-3.2	0.0119%	27.00	38.00	0.00453	0.00453
-3.15	0.0140%	27.75	37.25	0.00520	0.00520
-3.1	0.0163%	28.50	36.50	0.00596	0.00596
-3.05	0.0190%	29.25	35.75	0.00681	0.00681
-3	0.0222%	30.00	35.00	0.00776	0.00776
-2.95	0.0257%	30.75	34.25	0.00881	0.00881
-2.9	0.0298%	31.50	33.50	0.00997	0.00997
-2.85	0.0344%	32.25	32.75	0.01125	0.01125
-2.8	0.0396%	33.00	32.00	0.01267	0.01267
-2.75	0.0455%	33.75	31.25	0.01421	0.01421
-2.7	0.0521%	34.50	30.50	0.01589	0.01589
-2.65	0.0596%	35.25	29.75	0.01772	0.01772
-2.6	0.0679%	36.00	29.00	0.01970	0.01970
-2.55	0.0773%	36.75	28.25	0.02182	0.02182
-2.5	0.0876%	37.50	27.50	0.02410	0.02410
-2.45	0.0992%	38.25	26.75	0.02653	0.02653
-2.4	0.1120%	39.00	26.00	0.02911	0.02911
-2.35	0.1261%	39.75	25.25	0.03184	0.03184
-2.3	0.1416%	40.50	24.50	0.03470	0.03470
-2.25	0.1587%	41.25	23.75	0.03769	0.03769
-2.2	0.1774%	42.00	23.00	0.04080	0.04080
-2.15	0.1978%	42.75	22.25	0.04400	0.04400
-2.1	0.2199%	43.50	21.50	0.04729	0.04729
-2.05	0.2440%	44.25	20.75	0.05062	0.05062
-2	0.2700%	45.00	20.00	0.05399	0.05399
-1.95	0.2980%	45.75	19.25	0.05736	0.05736
-1.9	0.3281%	46.50	18.50	0.06070	0.06070
-1.85	0.3603%	47.25	17.75	0.06396	0.06396
-1.8	0.3948%	48.00	17.00	0.06711	0.06711
-1.75	0.4314%	48.75	16.25	0.07010	0.07010
-1.7	0.4703%	49.50	15.50	0.07289	0.07289
-1.65	0.5114%	50.25	14.75	0.07542	0.07542
-1.6	0.5546%	51.00	14.00	0.07765	0.07765
-1.55	0.6001%	51.75	13.25	0.07951	0.07951
-1.5	0.6476%	52.50	12.50	0.08095	0.08095
-1.45	0.6972%	53.25	11.75	0.08192	0.08192

Standard Deviation of Haircut	15.000
Mean Haircut	75.000
LTV at Sale	65.000
Shortfall% [Perfect Information]	0.000
Shortfall% [Future Haircut Unknown]	2.266
Predicted House Price @ Sale	100,000.00
Minimum Haircut	0

-1.4	0.7487%	54.00	11.00	0.08235	0.08235
-1.35	0.8020%	54.75	10.25	0.08220	0.08220
-1.3	0.8569%	55.50	9.50	0.08140	0.08140
-1.25	0.9133%	56.25	8.75	0.07991	0.07991
-1.2	0.9710%	57.00	8.00	0.07768	0.07768
-1.15	1.0297%	57.75	7.25	0.07466	0.07466
-1.1	1.0893%	58.50	6.50	0.07081	0.07081
-1.05	1.1495%	59.25	5.75	0.06609	0.06609
-1	1.2099%	60.00	5.00	0.06050	0.06050
-0.95	1.2704%	60.75	4.25	0.05399	0.05399
-0.9	1.3305%	61.50	3.50	0.04657	0.04657
-0.85	1.3900%	62.25	2.75	0.03823	0.03823
-0.8	1.4485%	63.00	2.00	0.02897	0.02897
-0.75	1.5058%	63.75	1.25	0.01882	0.01882
-0.7	1.5614%	64.50	0.50	0.00781	0.00781
-0.65	1.6150%	65.25	-0.25	-0.00404	0.00000
-0.6	1.6662%	66.00	-1.00	-0.01666	0.00000
-0.55	1.7148%	66.75	-1.75	-0.03001	0.00000
-0.5	1.7604%	67.50	-2.50	-0.04401	0.00000
-0.45	1.8027%	68.25	-3.25	-0.05859	0.00000
-0.4	1.8415%	69.00	-4.00	-0.07366	0.00000
-0.35	1.8763%	69.75	-4.75	-0.08912	0.00000
-0.3	1.9070%	70.50	-5.50	-0.10489	0.00000
-0.25	1.9335%	71.25	-6.25	-0.12084	0.00000
-0.2	1.9553%	72.00	-7.00	-0.13687	0.00000
-0.15	1.9725%	72.75	-7.75	-0.15287	0.00000
-0.1	1.9849%	73.50	-8.50	-0.16871	0.00000
-0.05	1.9923%	74.25	-9.25	-0.18429	0.00000
0	1.9948%	75.00	-10.00	-0.19948	0.00000
0.05	1.9923%	75.75	-10.75	-0.21418	0.00000
0.1	1.9849%	76.50	-11.50	-0.22826	0.00000
0.15	1.9725%	77.25	-12.25	-0.24163	0.00000
0.2	1.9553%	78.00	-13.00	-0.25419	0.00000
0.25	1.9335%	78.75	-13.75	-0.26585	0.00000
0.3	1.9070%	79.50	-14.50	-0.27652	0.00000
0.35	1.8763%	80.25	-15.25	-0.28614	0.00000
0.4	1.8415%	81.00	-16.00	-0.29463	0.00000
0.45	1.8027%	81.75	-16.75	-0.30196	0.00000
0.5	1.7604%	82.50	-17.50	-0.30807	0.00000
0.55	1.7148%	83.25	-18.25	-0.31295	0.00000
0.6	1.6662%	84.00	-19.00	-0.31658	0.00000
0.65	1.6150%	84.75	-19.75	-0.31895	0.00000
0.7	1.5614%	85.50	-20.50	-0.32008	0.00000
0.75	1.5058%	86.25	-21.25	-0.31998	0.00000
0.8	1.4485%	87.00	-22.00	-0.31868	0.00000
0.85	1.3900%	87.75	-22.75	-0.31623	0.00000
0.9	1.3305%	88.50	-23.50	-0.31267	0.00000
0.95	1.2704%	89.25	-24.25	-0.30806	0.00000
1	1.2099%	90.00	-25.00	-0.30248	0.00000
1.05	1.1495%	90.75	-25.75	-0.29599	0.00000
1.1	1.0893%	91.50	-26.50	-0.28867	0.00000
1.15	1.0297%	92.25	-27.25	-0.28060	0.00000
1.2	0.9710%	93.00	-28.00	-0.27188	0.00000
1.25	0.9133%	93.75	-28.75	-0.26257	0.00000
1.3	0.8569%	94.50	-29.50	-0.25278	0.00000

1.35	0.8020%	95.25	-30.25	-0.24259	0.00000
1.4	0.7487%	96.00	-31.00	-0.23209	0.00000
1.45	0.6972%	96.75	-31.75	-0.22136	0.00000
1.5	0.6476%	97.50	-32.50	-0.21048	0.00000
1.55	0.6001%	98.25	-33.25	-0.19953	0.00000
1.6	0.5546%	99.00	-34.00	-0.18858	0.00000
1.65	0.5114%	99.75	-34.75	-0.17770	0.00000
1.7	0.4703%	100.50	-35.50	-0.16695	0.00000
1.75	0.4314%	101.25	-36.25	-0.15639	0.00000
1.8	0.3948%	102.00	-37.00	-0.14607	0.00000
1.85	0.3603%	102.75	-37.75	-0.13603	0.00000
1.9	0.3281%	103.50	-38.50	-0.12632	0.00000
1.95	0.2980%	104.25	-39.25	-0.11696	0.00000
2	0.2700%	105.00	-40.00	-0.10799	0.00000
2.05	0.2440%	105.75	-40.75	-0.09942	0.00000
2.1	0.2199%	106.50	-41.50	-0.09127	0.00000
2.15	0.1978%	107.25	-42.25	-0.08355	0.00000
2.2	0.1774%	108.00	-43.00	-0.07627	0.00000
2.25	0.1587%	108.75	-43.75	-0.06943	0.00000
2.3	0.1416%	109.50	-44.50	-0.06303	0.00000
2.35	0.1261%	110.25	-45.25	-0.05706	0.00000
2.4	0.1120%	111.00	-46.00	-0.05151	0.00000
2.45	0.0992%	111.75	-46.75	-0.04637	0.00000
2.5	0.0876%	112.50	-47.50	-0.04163	0.00000
2.55	0.0773%	113.25	-48.25	-0.03727	0.00000
2.6	0.0679%	114.00	-49.00	-0.03328	0.00000
2.65	0.0596%	114.75	-49.75	-0.02963	0.00000
2.7	0.0521%	115.50	-50.50	-0.02631	0.00000
2.75	0.0455%	116.25	-51.25	-0.02330	0.00000
2.8	0.0396%	117.00	-52.00	-0.02058	0.00000
2.85	0.0344%	117.75	-52.75	-0.01813	0.00000
2.9	0.0298%	118.50	-53.50	-0.01592	0.00000
2.95	0.0257%	119.25	-54.25	-0.01395	0.00000
3	0.0222%	120.00	-55.00	-0.01219	0.00000
3.05	0.0190%	120.75	-55.75	-0.01062	0.00000
3.1	0.0163%	121.50	-56.50	-0.00923	0.00000
3.15	0.0140%	122.25	-57.25	-0.00800	0.00000
3.2	0.0119%	123.00	-58.00	-0.00691	0.00000
3.25	0.0101%	123.75	-58.75	-0.00596	0.00000
3.3	0.0086%	124.50	-59.50	-0.00512	0.00000
3.35	0.0073%	125.25	-60.25	-0.00439	0.00000
3.4	0.0062%	126.00	-61.00	-0.00376	0.00000
3.45	0.0052%	126.75	-61.75	-0.00321	0.00000
3.5	0.0044%	127.50	-62.50	-0.00273	0.00000
3.55	0.0037%	128.25	-63.25	-0.00231	0.00000
3.6	0.0031%	129.00	-64.00	-0.00196	0.00000
3.65	0.0026%	129.75	-64.75	-0.00165	0.00000
3.7	0.0021%	130.50	-65.50	-0.00139	0.00000
3.75	0.0018%	131.25	-66.25	-0.00117	0.00000
3.8	0.0015%	132.00	-67.00	-0.00098	0.00000
3.85	0.0012%	132.75	-67.75	-0.00082	0.00000
3.9	0.0010%	133.50	-68.50	-0.00068	0.00000
3.95	0.0008%	134.25	-69.25	-0.00057	0.00000
4	0.0007%	135.00	-70.00	-0.00047	0.00000
	100.0000%			**-10**	**2.27**

Appendix C: References:

1. Low Default Portfolios: A Proposal for Conservative Estimation of Default Probabilities [Nathanaël Benjamin, Alan Cathcart and Kevin Ryan, Financial Services Authority 3rd April 2006]
2. Estimating Probabilities of Default for Low Default Portfolios, Working Paper [Pluto, Katje and Dirk Tasche (2005)]
3. Likelihood Approaches to Low Default Portfolios, Working Paper [Forrest, Alan (2005)]
4. Small Sample Modelling [A Lucas and J Powell: 1994 Credit Scoring and Credit Control Edinburgh conference]
5. A Constrained Scorecard Development Approach [Presentation by A Lucas at 2003 Credit Scoring and Credit Control Edinburgh conference]

Alan Lucas
Rhino Risk Ltd.