"EXPERT RANKING IN LOW DEFAULT PORTFOLIO MODELLING"

Alistair Paulls and Michelle Greenidge[1]

INTRODUCTION

Probability of Default (PD) modelling for corporate lending portfolios often encounters the situation where the number of defaults for the portfolio, over the available historical period, is low or non-existent. There are well documented methodologies (e.g. Benjamin et. al[2]) for practitioners to establish the central tendency of the PD in such instances to calibrate the portfolio to the right level.

However, little guidance is made available for practitioners to discriminate default risk between obligors in the absence of default data, or in other words, what would constitute the dependent or outcome variable in the PD model build.

One general assumption made in such cases is that external rating agency ratings can be used as the outcome variable, but for niche portfolios external ratings are not always available. In these situations, heritage Bank of Scotland Corporate[3] employs ranking based on expert opinion as the outcome or dependent variable in the PD model. To date three PD models have been built using expert ranks as the target variable and others have used a version of the expert ranking in model validation.

This paper studies various aspects of expert ranking such as the issues of limited data availability and practical challenges faced in conducting an expert ranking. The weaknesses and bias of expert ranking are discussed, and the effectiveness of expert ranking for building PD models is examined in light of validation results as well as the on-going model monitoring.

---

[1] Alistair Paulls and Michelle Greenidge are from Lloyds Banking Group. Address: CityMark, 150 Fountainbridge, Edinburgh EH3 9PE. Email: alistairpaulls@bankofscotland.co.uk , michellegreenidge@hbos.co.uk.

[2] Nathanaël Benjamin, Alan Cathcart and Kevin Ryan, 2006, Low Default Portfolios: A Proposal for Conservative Estimation of Default Probabilities, FSA

[3] Now part of the Lloyds Banking Group

WHAT ARE EXPERT RANKS AND WHAT ARE THEY USED FOR?

WHAT IS AN EXPERT RANK ORDER?

Expert ranking is a process in which a panel of experts rank order a number of deals based on defined criteria. For example, a rank ordering on a portfolio of deals based upon the likelihood that the counterparty will default is performed for PD model development. In this case the estimate must only be based on the likelihood of default, and therefore be independent of the proportion of exposure that would be lost in the event of a default. This however, is not the only basis for a rank ordering by experts, but is the only situation which this paper will address.

WHY DO WE NEED TO PERFORM AN EXPERT RANK?

The main reason for performing an expert rank ordering process is to supplement a lack of hard data, from either internal or external sources, although the situations in which it is used may vary. The data requirements and issues are discussed in the next section.

Rank ordering can also be used to sense check that model outputs are in line with perceived expectation and to check that the statistical model matches real world experience.

Applications of the expert rank process include:-

- Generating the target variable for a PD model build
- Validation of a candidate model
- Annual validation of an existing model

DATA SOURCES AND AVAILABILITY

As mentioned above, the main need for performing an expert rank ordering process is to supplement a lack of hard data, from either internal or external sources. The three main sources of data can be ordered according to their strength as shown below. It should be noted that in terms of practicality, time, resource and logistics, expert judgement would often be superseded by the use of External Credit Assessment Institution (ECAI) ratings.

INTERNAL DATA

The primary data source and the one which should, all other things being equal, carry the greatest weight in decisions throughout the modelling process is internal data.

Internal data will always provide the strongest empirical strength but there are a number of reasons why this may not always be sufficient, these include but are not limited to the following:

- Insufficient time series – Default data may be available for the portfolio required, however there is not enough data captured for modelling purposes. For example where you are trying to model 1 year forward looking PD, but you only have 6 months data

- Insufficient data points – There may be an insufficient number of defaults in your sample for meaningful analysis.

- Missing data – There may be some default data, but it is incomplete and doesn't provide enough weight to show that it is representative of the portfolio as a whole.

- Fundamental change in past, present or future business practice – Any default history becomes of decreasing relevance if the business has changed practices to stop the cause of the default occurring in the future, diminishing further the longer the change has been in place.

EXPERT JUDGEMENT

If insufficient internal data is present then the secondary data source should be expert judgement. This will generally be provided by subject matter experts within the business who have the experience within the organisation or others that have relevance to the particular issue at hand. It is likely these colleagues will be at a level of seniority as well as experience that ensures a relevant knowledge base.

This expert judgement should be used either to validate outcomes where sufficient internal data is available and has produced results or in the absence of sufficient internal data to derive parameters and/or parameter values. By its very nature expert judgement has a degree of subjectivity attached to it and accordingly it is appropriate to try to provide additional context for any results.

In the absence of both internal and external data then expert judgement becomes perhaps the only option.

EXTERNAL DATA

The third source of information is external data or benchmarks. Clearly an important aspect of using either external data sources or benchmarking data is its relevance and appropriateness for the purpose to which we wish to apply it and this needs to be clearly documented. However, expert ranking methodology tends to be applied where external data is difficult to source. In the case where external ratings provide a target variable it must be

acknowledged that the modeller is attempting to reproduce someone else's model without the benefit of that person's logic process.

In addition to the limited availability of external ratings, they can also prove to be prohibitively expensive.

Process Overview & Practical Challenges

This section details the process considerations and highlights the potential problems which may occur with this expert ranking methodology. It links these problems to the stages of the rank ordering exercise so these problems may be minimised, if not negated.

The process to classify the portfolio into expert ranks needs to consider the following:-

- **Data availability:-** Sufficient information on the obligors which are to be ranked must be available. The same level of detail should be available for all obligors.

- **Determining the experts:-** To ensure a comprehensive view of the risk the experts should consist of both individuals with extensive experience of the business area who propose the credit and those who sanction or approve that credit.

- **Consistent method:-** At the outset a detailed briefing should be provided to the experts of what is required for the exercise with clear and concise guidelines.

- **Number of risk buckets:-** The number of buckets to assign the deals to must be chosen, this is largely dependent on the number of deals that need to be ranked. A good rule of thumb is that 20 deals for every category you choose, so that a portfolio of 100 deals should have a maximum of five categories.

- **Reducing bias / subjectivity:-**

    - The two areas (business and sanctioning) performing the ranking should initially work independently to create their respective ranks.

    - Two representatives, one from each team meet and combine their results to provide an overall consensus rank

The process as outlined above has a number of practical challenges which face the ranking procedure:-

It is important to maintain the independence of the groups performing the initial ranking. Ideally an independent view from outside of the business area

with suitable experience within the lending area is sought. As the purpose of the exercise it to define the target variable the final rank should contain the views of those people in the best position to determine the risk of the deal.

This independent view is usually provided by a credit approval or sanctioning function. Independence is also required as the position may arise where the same people who offered the facility to the client are being asked to rank their own deals and judge their efficacy. The objective assessment of their own choices can be difficult to undertake.

One suggestion to counter this issue is to remove the name of the deal to provide anonymity. However, due to the detailed knowledge that the relationship managers often have of their connections they are likely to recognise the deal regardless.

It has been heritage Bank of Scotland practice to recruit a minimum of 4 experts (two from the business area and 2 from the sanctioning area)[4] to perform the exercise. Due to business as usual activities it can be difficult to find expert resource to perform the task. However, the importance of the expert ranking ordering exercise must be made clear at recruitment and further emphasised at the initial meeting. The expert rank order produced is used as the target variable for the PD model build which will differentiate between the risk of deals within the Business' portfolio. This in turn will affect the pricing and capital requirement associated with these deals, therefore the Business must understand this and allocate appropriate resource and time to the task.

The groups ranking the deal should contain sufficient people to ascertain a consensus view of the risk. One person's view should not dominate the final view obtained.

The number of buckets chosen will be driven in part by the ability of the experts to differentiate the deals – the nature of the deals may prevent too much granularity, so a smaller number of buckets may be more apt. An even number of buckets should ideally be used – this prevents deals being entered into the middle bucket or choosing the 'average' rank.

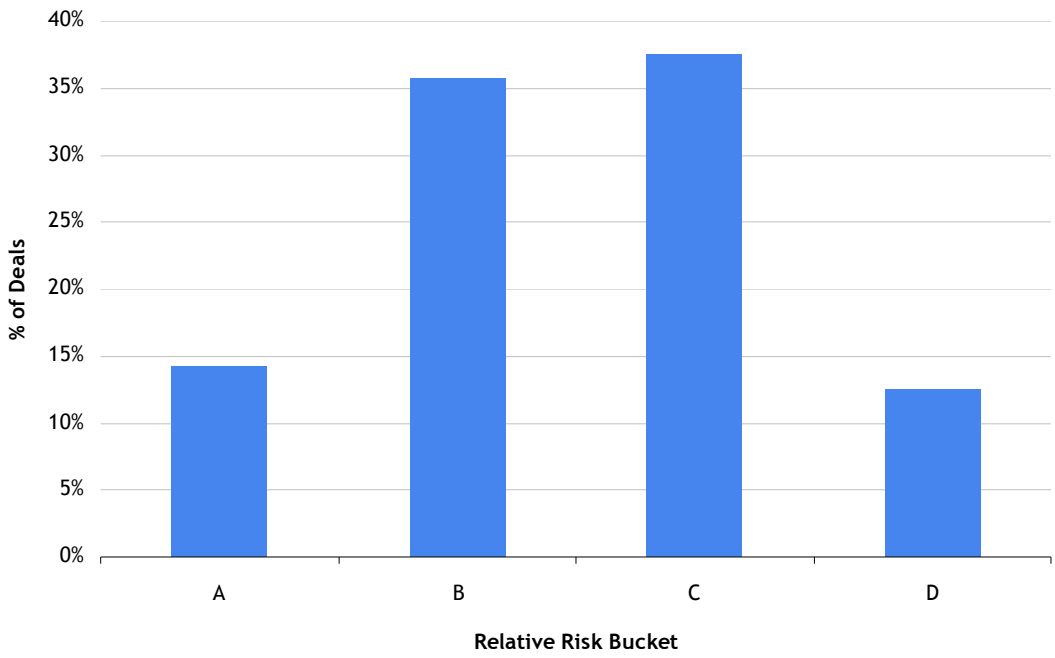
When performing the ranking exercise no direction should be given to the experts as to allocating equal numbers of deals to each bucket, as this would potentially force a distribution not present in the underlying data. The distribution sought should be truly representative of the portfolio, rather than a uniform distribution. The distribution obtained should be within the expectation of those with knowledge of the portfolio for example the distribution below is of the expert ranks of a portfolio for a niche market which managed within a small business area.

---

[4] 3 representatives from each area would be preferential as it would afford a 'casting vote' in areas of disagreement. However, this number should not exceed 3 as this would increase the time taken and probably introduce decisions by 'committee'.

Figure 1 below shows the distribution of risk buckets typically expected from an expert ranking of a portfolio managed by a small number relationship managers from a single area.

Figure 1: Example Distribution of Niche Portfolio Expert Ranks

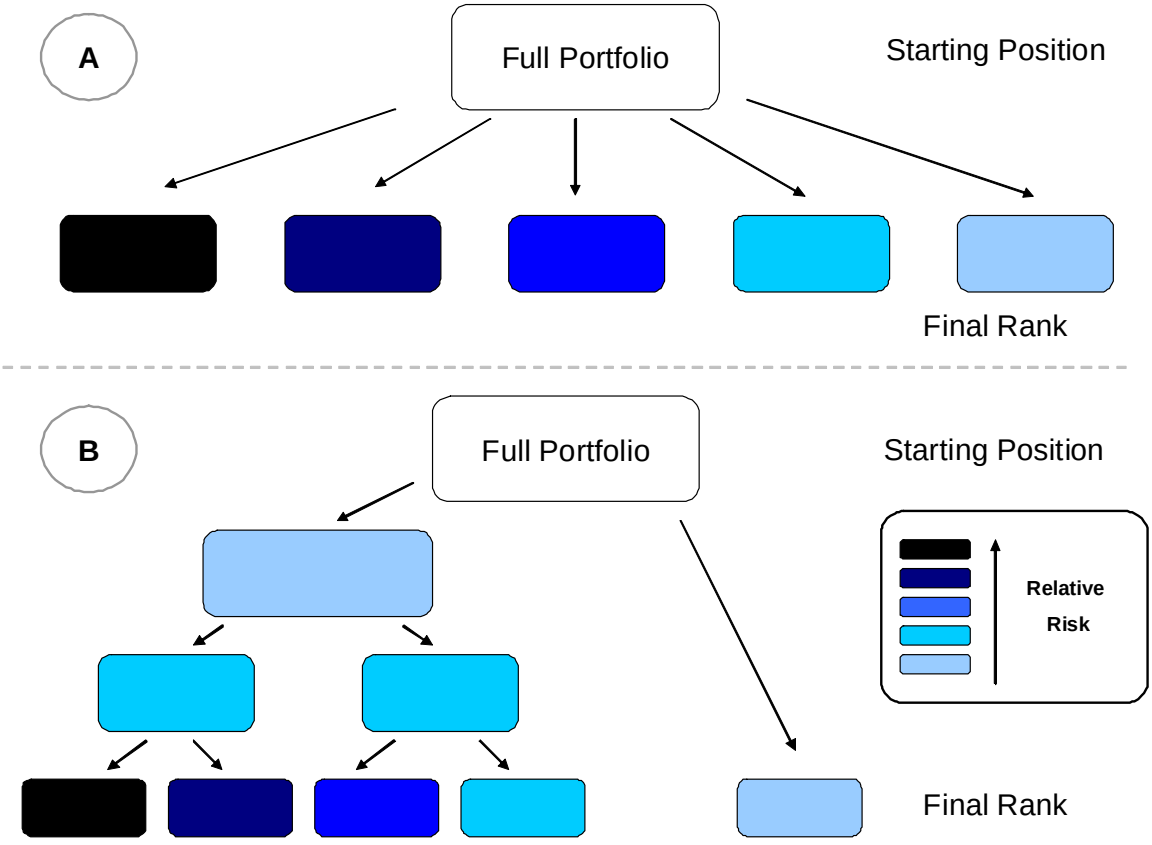


The market size and the size of a niche business area would suggest that there should be consistency of behaviour which would be manifest by a set of deals that have similar risk profiles. The graph above displays this, as most deals are in the centre (B's and C's) with a lesser number that are relatively higher or lower quality than the 'norm'.

One process to allocate deals to ranks which has worked well is the following. First allocate the portfolio to buckets based on a simple 2-way split of the deals to better and worse ranks, then repeating the process until the final rank is obtained. It should be noted that while this approach may result in an equal number of deals per bucket, this should not be the aim. Figure 2 below illustrates the differences between this approach and allocating all of the ranks in a one step process.

This multistep method makes it easier for the experts to conduct the exercise, reduces the time involved by minimising the number of iterations required before a final result is reached and conquers the reluctance to allocate deals to the 'worst' risk bucket.

Figure 2: A – Direct Ranking to Buckets, B – "Split" Method of Ranking to Buckets



Though the experts acknowledge the issue, they are often reluctant or even refuse to categorise any deal within their portfolio to the worst rank, even though it is explained that the buckets are meant to be relative not absolute.

Time must be taken to explain in detail the purpose of the exercise and particularly how the portfolio should be ranked. Many relationship managers form their view of the risk of a deal based upon the expected loss for a connection. In the instance of a PD model build the difference between PD and Loss Given Default should be made explicit at the outset.

The original credit applications and other historical data may take time to source, especially if it is not held centrally. Often the relationship managers who write the deal also store the applications, therefore if a large number of applications need to be sourced a large number of individuals may have to be contacted.

Finally, in order for the whole process to run smoothly and efficiently it needs to be well planned and tightly controlled, particularly around the necessity of having all of the data and the people required available at the same time.

Weakness and Bias in the Process

The following section details some of the weakness and bias which can arise from application of the expert ranking process.

Hindsight risk is a significant vulnerability within the process. Due to the nature of the model build process, the data collected for the model build may be a few years old. This raises the possibility that any of the connections within the portfolio have defaulted subsequent to the credit approval.

This knowledge can influence the judgement of the relationship manager and sanctioner, as they will potentially rank the deal a poor grade, influenced by the knowledge of the default occurring. Therefore it must be made abundantly clear to the experts that their rank must be based solely upon the data available as of that time, and not let any knowledge of any subsequent events alter their perception of risk.

Scorecard in your head - the experts should not be influenced by the modellers as to what the main risk drivers within the portfolio may be, as this can lead to the expert having a 'scorecard in their head'. For example, if an expert is told that turnover is a good driver of predicting credit risk, this direction will influence the rank they allocate deal into by turnover. This inevitably leads to turnover becoming a driver within the final developed model, which in reality it may not be.

Appropriate use of the risk buckets - there can be reluctance of the experts to allocate any of their deals into the lowest rank available. In this instance it must be made clear to them that the ranks are all relative ranks and not absolute i.e. the lowest rank does not correspond to a defaulted case.

In this instance it is important to emphasise that the rank ordering process aims to differentiate the risk between the deals in the portfolio, and that this differentiation does not correlate directly to a rating scale. Therefore time should be taken to explain that the deals given the lowest rank will not necessarily be given a low rating, and that this spread is defined during the model calibration.

Information imbalance - the additional knowledge which a relationship manager may have from managing a particular deal may influence its rank against another. This extra level of detailed knowledge outside of the data provided may bias them towards or against a connection in opposition to another which isn't known to the expert.

Data quality - the amount of data present for the deals within the portfolio may be inconsistent. Scarcity of data for a particular deal could potentially lead to inaccuracy in the final rank assigned.

Resources - as mentioned previously, it can often be difficult to not only find an appropriate expert with knowledge of the area, but also finding their time to

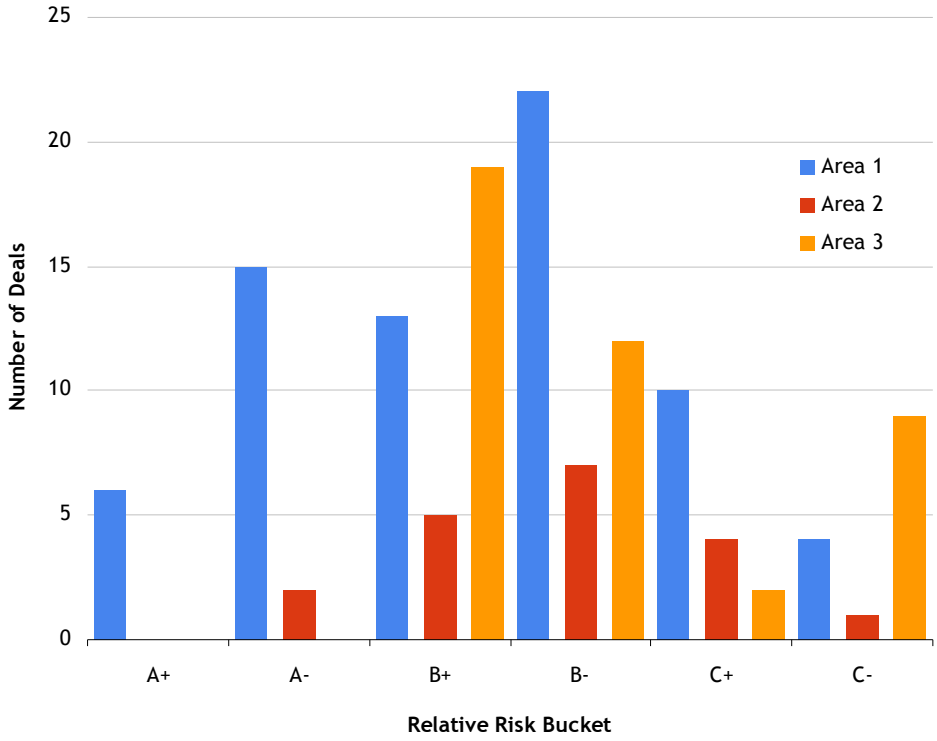
take part in the process can be difficult. Due to their business as usual commitments, many experts find it difficult to justify the time taken to partake in the ranking exercise.

Fatigue / complacency - the length of the ranking process may lead to fatigue or complacency of the experts. If, for example, the portfolio to be ranked consisted of a large number of deals, the time to perform the ranking may take weeks. It should be made clear that the same degree of time and scrutiny should be taken for the initial deals as for the rest of the portfolio.

Inappropriate rank distribution – this paper primarily deals with niche market deals managed within a single business area, however, there are situations where low default portfolios span many asset classes and geographical regions. In allocating deals to risk buckets in such situations there is a real possibility that the end result is a model in which region/asset class is a factor. Using regions as an example, in the worse case scenario all of the deals from a particular region end up in the same worst rank, with the deals from the other regions spread across the remaining ranks.

This may become problematic both in terms of modelling and diplomacy. The former can be effectively dealt with if the there are large number of risk buckets and the region concerned occupies more than one of them in a particular direction, all this requires is a different calibration for this region. However if all of the deals for this region are in one category this is not a viable solution and if the population for that region is large enough it will be necessary to expert rank it on its own. The diplomacy comes with having to manage the expectations of a business area whose deals lie solely in the highest relative risk buckets. Figure 3 below shows a distribution from an expert ranking exercise conducted in December 2008.

Figure 3:- Expert Ranks from a 6 Bucket Exercise (N=131)



The distribution illustrates that area 3 ranks range from B+ to C-, implying that a separate calibration should be considered for it.

Validation - Any model built can be validated by using any of the standard methods e.g. holdout samples, but the ranks themselves cannot be truly validated until defaults start to occur[5]. However, there are two checks available, though both of which are flawed:-

- If there are a few external ratings available for the portfolio then the ranks can be compared to these – agreement between the external ratings and expert ranks should provide a cushion of comfort – the flaw is the assumption that the external ratings are correct

- If there have been any defaults in the portfolio then ranks based on the data available 12 months before should indicate these defaults as being in the worse risk bucket – the flaw is the possible hindsight risk that may be exhibited.

Effectiveness of Expert Ranking for Building PD Models

Expert ranking has been used to define the target variable for a number of low default portfolios within Bank of Scotland Corporate.

[5] This is the same for models built on external ratings

External ratings are often used as a target variable. However, there may be instances where external ratings are only available for a subset of an entire portfolio. In this instance there is the possibility of building the model using the external ratings, then applying it to the whole portfolio.

Care needs to be taken that those deals rated within the subset is representative of the portfolio. Due to the costs involved, often only the large companies can afford external ratings. This may illustrate differences in their makeup and therefore potentially different default drivers may be applicable to them.

Expert ranks only provide the ability to differentiate the risk within a portfolio, but does not provide any help determining the central tendency of the portfolio. Therefore an appropriate estimation of the central tendency needs to be applied in conjunction with the expert ranks.

If a default occurs then the materiality of the default needs to be investigated

Validation and Model Monitoring

Expert ranks have a use in the validation and monitoring of models which are currently in use.

For example, expert ranks can be applied in the situations where there are PDs from an existing model but not enough internal default data to test the efficacy of the model for a given portfolio or sub-portfolio. In this situation, expert ranks can be derived for the portfolio and the correlation between PD and the expert rank can be explored to examine the appropriateness of the model.

This approach validates that:-

- The expert ranks match the direction of the PD estimates from the model – The better ranks have in general have a lower PD than the worse ranks.

- The model can effectively differentiate between the ranks in that the average PD better ranks is lower than the average PD for the worse ranks and that there is a statistical significance between them.

One further technique for validation is to apply a 'soft' definition of default to the portfolio whereby the Basel II definition of default is made more strict. For example, changing the 90 days past due default flag criteria to 30 days past due may increase the number of observed defaults. However, this method still may have its limitations when applied to a small niche portfolio.

Conclusion

The choice of using expert ranks as a target variable will never be a modellers first choice for many reasons, not least of which is the resource requirement in both time and personnel.

However, in the cases where it is employed (i.e. portfolios with little or no default data and no external ratings available) it can be a very effective method if made robust through ensuring that the process is systematic and taking every step possible to reduce subjectivity.

In cases of insufficient internal default data and no external data that can be used as a proxy, the method applied can provide the basis of a PD model and hence act as the determinant of the inputs to the PD model.

Expert rank ordering can also be used to validate or otherwise the outputs of a model which is already in use.

Whilst default data is the ideal basis on which to build a PD model, the validity of expert ranking should not be underestimated. Those performing the ranking usually have knowledge gained over a number of years of underwriting.