Chapter 6

Empirical Results: Validation of Credit Rating system

6.1. Background of the study

The primary purpose of validation is to examine whether the internally constructed scoring model can fully explain the credit status of borrowers. As sampled data used to construct the model can mostly be explained by the model, it is necessary to see whether the model possesses sufficient explanatory power for different samples. Thus out-sample testing should be carried out to observe the tendency of over-learning, which will lower the predictability of model. In addition, improper sampling and omission of relevant information will lead to model bias. Thus external data should be employed to assess the validity of the rating model. Moreover, since the primary objective of the rating model is to make forecast, whether the model works normally under all circumstances, including significant changes of the macroeconomic environment, must also be validated. Below is an introduction to the framework of model validation.
6.2. Analytical framework of the study

The validity of a rating model is judged from three dimensions (a) discriminative power: the accuracy of the model in differentiating non-defaulters and defaulters; (b) homogeneous: does the model provide enough rating grades to classify borrowers with different credit characteristics, while the credit characteristics of all borrowers in the same grade are homogenous; (c) stability: a good rating model must take into account the influence of external economic factors that the model outputs reflect the credit status of individual borrowers and represent long-term trends without being affected by short-term volatility. The validation methods for different dimensions of a rating model are empirically attempted below.

6.3. Data sources

We have taken a small loan portfolio of a bank, whose name is not disclosed for confidentiality. In the said loan portfolio are assigned internal ratings by the said bank, which we have validated. Further, for the default data, we have sourced from Standard & Poor’s from 1981 to 2001.

6.4. Methodology & Derivation of Empirical Results

We applied methods for introducing methods for evaluating discriminatory power (cumulative accuracy profiles and receiver operating characteristics), both discrimination and calibration (Brier
Score) and simple calibration (binominal test and a test allowing for default correlation).

In Cumulative Accuracy Profile (CAP) is a visualizing tool for discriminatory power where it is validates that the rating system discriminates well. The basic inputs for CAP are historical data and ratings and default behaviour.

After we have the observations, rating (A=best to c =worst scale) and default occurrence (default =1), then we computed cumulative accuracy points and computed CAP graph.

**Figure 6.4.1: Cumulative Accuracy Profile (CAP)**

The computed the area under curve which is 0.616667 and area under the perfect ratings which is .76. We computed the Accuracy ratio which is (area under curve -.5)/ (area under perfect rating-.5) and that comes to .448718 or 44.87%.The accuracy ratio makes the information under the CAP into a single number. This is arrived at by relating the area under the CAP curve of an informative rating system.
Then we have computed the area under the curve which is 0.616667 and area under perfect credit rating model is 0.76 and accuracy ratio is 0.448718 i.e. 44.87%.

Generally accuracy ratios of rating system ranges between 50% to 90% and it depends on the portfolio structure specially heterogeneity of the portfolio in respect to default probabilities. Our case accuracy ratio 44.87% allows the discriminatory power of the rating system.

We have set PD estimate for the year 2002 to the average default rate observed in year 1981-2001. The asset correlation value is set to 7%. The we run the Binomial test. We classified three rating-specific PDs as underestimating the true default rate at a significance of 1% and number increased to 4 with normal approximation. When we assumed the asset correlation at 7% the significance level rises prompting not to reject PD at 1% significance level, if the default rate is test brings 0.
### Table 6.4.1: p values - Binomial, Normal and One factor test

<table>
<thead>
<tr>
<th></th>
<th>1981-2001</th>
<th>2002</th>
<th>N\text{2002}</th>
<th>D\text{2002}</th>
<th>Binomial</th>
<th>Normal</th>
<th>One-factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA</td>
<td>0.00%</td>
<td>0.00%</td>
<td>132</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AA</td>
<td>0.01%</td>
<td>0.00%</td>
<td>526</td>
<td>0</td>
<td>0</td>
<td>99.2%</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>0.05%</td>
<td>0.09%</td>
<td>1120</td>
<td>1</td>
<td>42.9%</td>
<td>53.2%</td>
<td>14.5%</td>
</tr>
<tr>
<td>BBB</td>
<td>0.26%</td>
<td>1.02%</td>
<td>1271</td>
<td>13</td>
<td>0.0%</td>
<td>0.0%</td>
<td>1.7%</td>
</tr>
<tr>
<td>BB</td>
<td>1.22%</td>
<td>2.74%</td>
<td>802</td>
<td>22</td>
<td>0.0%</td>
<td>0.0%</td>
<td>6.6%</td>
</tr>
<tr>
<td>B</td>
<td>5.96%</td>
<td>8.09%</td>
<td>754</td>
<td>61</td>
<td>1.1%</td>
<td>0.8%</td>
<td>21.5%</td>
</tr>
<tr>
<td>CCC/C</td>
<td>24.72%</td>
<td>44.12%</td>
<td>170</td>
<td>75</td>
<td>0.0%</td>
<td>0.0%</td>
<td>2.0%</td>
</tr>
</tbody>
</table>

When devising a rating system, it is essential to fit and calibrate the system to empirical data. The quality of a system whether discriminatory power or calibration is higher for higher default data i.e. in-the-sample power is higher than out-of-sample power. The danger of overfitting or data mining increases if more variables are used along with outlier treatment.

### 6.5. Summary

Credit rating model validation covers extensive dimensions, each one of them is hard pressed to take into account all situations. Hence a mix of validation approaches should be more appropriate. Also as statistical figures cannot set a so-called reasonable range, benchmarking using external information and even different models but identical samples should be employed to examine the soundness...
of internal models. In model construction, sometimes it is difficult to match all measures at the same time (e.g. the model results have optimum discriminative power (we have computed CAP), stability and grade distribution at the same time). Thus banks should identify the primary objective of constructing an internal model and prioritize all the dimensions to make sure their internal model achieves the intended results.