

SESSIONAL PAPER

Calibration of transition risk for corporate bonds

[Presented to the Institute & Faculty of Actuaries, Staple Inn Hall, London, 15 May 2023]

J. Sharpe¹, F. Ginghina², G. Mehta³ and A.D. Smith⁴

¹Sharpe Actuarial Limited, United Kingdom

²Milliman, United Kingdom

³Eva Actuarial and Accounting Consultants Limited, United Kingdom

⁴University College Dublin, Ireland

Corresponding author: J. Sharpe; Email: james@sharpeactuarial.co.uk

Abstract

Under the European Union's Solvency II regulations, insurance firms are required to use a one-year VaR (Value at Risk) approach. This involves a one-year projection of the balance sheet and requires sufficient capital to be solvent in 99.5% of outcomes. The Solvency II Internal Model risk calibrations require annual changes in market indices/term structure/transitions for the estimation of the risk distribution for each of the Internal Model risk drivers.

Transition and default risk are typically modelled using transition matrices. To model this risk requires a model of transition matrices and how these can change from year to year. In this paper, four such models have been investigated and compared to the raw data they are calibrated to. The models investigated are:

- A bootstrapping approach – sampling from an historical data set with replacement.
- The Vašíček model was calibrated using the Belkin approach.
- The K-means model – a new non-parametric model produced using the K-means clustering algorithm.
- A two-factor model – a new parametric model, using two factors (instead of a single factor with the Vašíček) to represent each matrix.

The models are compared in several ways:

1. A principal components analysis (PCA) approach that compares how closely the models move compared to the raw data.
2. A backtesting approach that compares how each model's extreme percentile compares to regulatory backtesting requirements.
3. A commentary on the amount of expert judgement in each model.
4. Model simplicity and breadth of uses are also commented on.

Keywords: Credit risk; Transition and default risk; Vašíček credit model; solvency II; K-means

1. Introduction

In this paper transition and default risk for credit assets in the context of the Solvency II requirement to develop a 1-in-200 VaR and full risk distribution has been looked at. This primarily involves developing a model of transition matrices. Relative to other risks on an insurance company balance sheet, this risk is more complex with a wider range of considerations.

- In Section 2 we outline the key risk drivers associated with this risk and introduce the core modelling component – the transition matrix.
- In Section 3 we look at the primary sources of historical transition matrix data with a discussion of how this data is treated.
- Section 4 analyses of the data, presenting the key features any credit model should aim to capture.
- Section 5 discusses the different types of credit models and then presents a range of models split between parametric and non-parametric model types. The parametric models explored are the Vašíček model (Vašíček, 1987, 2002) calibrated using an approach described in Belkin *et al.* (1998); and a new two-factor model introduced in this paper. Two non-parametric models are also explored; a model (known as the K-means model) which uses the K-means algorithm to group historical matrices; as well as a simple bootstrapping approach of simulating historical transition matrices with replacement.
- Section 6 includes a quantitative and qualitative comparison of the various credit models. For the quantitative comparison, principal components analysis (PCA) is used to identify the directions of most variance in historical data, which is then compared for each of the models. The first and second principal components (PC1 and PC2) are focused on. A second quantitative comparison involves comparing the 99.5th percentile with that expected from Solvency II regulations. For the qualitative comparison, there is a discussion of strengths and weaknesses of each model. The simpler models are easier to calibrate and explain to stakeholders but at the cost of not explaining as many of the key features seen in the data in practice. The more complex models can allow a closer replication of the key data features but with a greater challenge in explaining them to stakeholders.

The key questions this paper has sought to answer when comparing the models are:

- Does the model output move in a consistent way compared to the historical data (i.e. are PC1 and PC2 from the underlying data consistent with PC1 and PC2 from the model)?
- Does the model produce stress transition matrices that are sufficient to meet reasonable back-testing requirements?
- Is the model calibration largely objective (i.e. based on prescribed calibration methods/data) or is there significant scope for expert judgement in the model calibration?

In this paper we find that the Vašíček model does not move consistently with the raw data; PC1 from the raw data is more consistent with PC2 from the Vašíček; and PC1 from the Vašíček is more consistent with PC2 from the raw data. The other models explored in this paper do move more consistently with the raw data.

The Vašíček and two-factor model require additional strengthening to ensure their 99.5th percentile exceeds the 1932 transition matrix. A bootstrapping approach can never exceed the worst event in its data set, which is a significant issue for models used for future events as the worst case can never be worse than the worst event in history (an example of the Lucretius fallacy). The K-means model is specified to pass back-testing as required and includes events worse than the worst event in history.

The K-means model as implemented in this paper has significant expert judgement required. This allows for flexibility in model development but is also less objective. The bootstrapping approach has no requirement for expert judgement at all beyond the choice of data. The Vašíček and two-factor models can be applied with varying amounts of expert judgement depending on the purpose for which the model is designed.

Table 1. S&P average transitions from 1981 to 2018.

From/to	AAA	AA	A	BBB	BB	B	CCC/C	D
AAA	89.82%	9.42%	0.55%	0.05%	0.08%	0.03%	0.05%	0.00%
AA	0.52%	90.63%	8.17%	0.51%	0.05%	0.06%	0.02%	0.02%
A	0.03%	1.77%	92.30%	5.40%	0.30%	0.13%	0.02%	0.06%
BBB	0.01%	0.10%	3.64%	91.63%	3.86%	0.49%	0.12%	0.18%
BB	0.01%	0.03%	0.12%	5.35%	85.80%	7.36%	0.61%	0.72%
B	0.00%	0.02%	0.09%	0.19%	5.63%	85.09%	5.05%	3.93%
CCC	0.00%	0.00%	0.13%	0.24%	0.70%	15.63%	51.49%	31.82%
D	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100%

2. Risk Driver Definition

Transition and default risk apply to both the modelling of assets and liabilities.

- On the asset side, credit ratings are given to individual assets and the movement between different rating classes can impact the asset price. Default on any asset also means significant loss of value on that asset. It might be possible to use credit spreads as opposed to credit ratings to model credit risk, however historical time series of credit spreads are largely split by credit ratings, so it is difficult to avoid the use of credit rating. Hence transition and default risk are modelled using transition matrices.
- On the liability side, many solvency regulations have a link between the discount rate used to discount liabilities and the assets held to back the liabilities. In the case of the matching adjustment in the Solvency II regime, the credit rating of the assets is explicitly used to define default allowances.

Transition matrices are used to capture probabilities of transitioning between credit ratings and default (an absorbing state). They are produced from the number of corporate bonds that moved between credit ratings or defaulted over a given time period.

An S&P transition matrix is shown below. This gives the one-year transition and default probabilities based on averages over 1981–2018.

The transition matrix itself is the data item that is being modelled. A historical time series of transition matrices can be obtained and this time series of 30–100 transition matrices is used to gain an understanding of the risk. Each matrix is itself 7*7 data points (i.e. the default final column is simply 100% minus the sum of the other columns in that row; the bottom default row is always a row of 0% with 100% in the final column (as above)).

The complexity of this data source makes transition and default risk one of the most complex risks to model.

3. Data Sources

For historical transition matrices, there are three main data sources for modelling:

1. Moody's Default and Recovery Database (DRD) and published Moody's data.
2. Standard and Poor (S&P) transition data via S&P Credit Pro.
3. Fitch's transition data.

We present a qualitative comparison of the data sources in Section 4.

Table 2. Comparison of transition risk data sources considered.

Criteria	Moody's default and recovery database	Moody's published data	S&P credit pro	Fitch transition data
Data length and Frequency	Annual transition data available from 1920's for US. For Europe and UK, the data starts from 1980's	Annual transition matrix data available from 1920's for the global portfolio of bonds	Annual transition matrix data available from 1981 for US, UK, and global portfolio of bonds	Annual transition matrix data available from 1973 for a global portfolio of bonds
Granularity of Data	Sector specifics: transitions data available specific to Corporates, Financial Institutions, Insurers, Covered Bonds, Structured Finance and US Public Finance Rating Specifics: 22 different ratings available Currency Specifics: GBP, EUR, and USD	Sector specifics: transitions data available specific to Corporates, Financial Institutions, Insurers, Covered Bonds, Structured Finance and US Public Finance Rating Specifics: 22 different ratings available Currency Specifics: GBP, EUR, and USD	Sector specifics: transitions data available specific to Corporates, Structured Finance and Sovereigns Rating Specifics: 16 different ratings available Currency Specifics: GBP, EUR, Canadian Dollar, Yen, and USD	Sector specifics: transitions data available specific to Corporates, Financial Institutions, Insurers, Covered Bonds, Structured Finance and US Public Finance Rating Specifics: 20 different ratings available Currency Specifics: GBP, EUR, and USD

Table 3. S&P one-year corporate transition rates by region (2018).

From/to	AAA	AA	A	BBB	BB	B	CCC/C ¹	D	NR
AAA	88.89	11.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00
AA	0.00	92.05	5.50	0.00	0.00	0.00	0.00	0.00	2.45
A	0.00	0.87	91.97	3.69	0.00	0.00	0.00	0.00	3.47
BBB	0.00	0.00	2.81	90.79	1.43	0.00	0.06	0.00	4.91
BB	0.00	0.00	0.00	4.08	82.93	4.15	0.00	0.00	8.84
B	0.00	0.00	0.00	0.00	2.80	78.13	3.99	0.98	14.09
CCC/C	0.00	0.00	0.00	0.00	0.51	12.82	46.15	27.18	13.33

Note that as well as the main credit ratings, this data contains a category called "Not Rated (NR)". We have removed the NR category by reallocating it to all other ratings by dividing by $(1-p(\text{NR}))$.

We have used published S&P transition matrices as the key market data input for the corporate downgrade and default risk calibration in the models analysed in this paper. This data is freely available for the period 1981–2019 in published S&P indices and using this data combined with transition matrices from the Great Depression (Varotto, 2011) can be used to calibrate transition matrix models. A sample matrix is shown in Table 3.

Some key points to note about transition matrices are:

1. Each row sums to 1 (100%), as this represents the total probability of where a particular rated bond can end up at the end of the year.
2. The leading diagonal of the transition matrix is usually by far the largest value, representing bonds that have remained at the same credit rating over the year.

¹These ratings are grouped together in the S&P data.

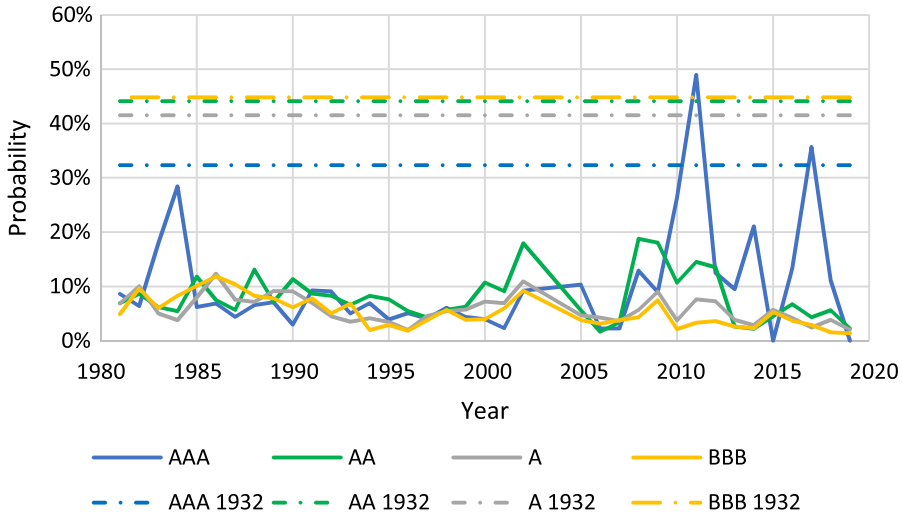


Figure 1. Downgrade rates for investment-grade assets.

3. A transition matrix multiplied by another transition matrix is also a valid transition matrix with the rows summing to 1; and the calculated matrix containing transition probabilities over two periods.
4. For completeness, there is also a row for the default state with zero in every column, except for the default state itself which has value 1.

4. Stylised Facts of Data

The data set used is a series of transition matrices, one for each year. This makes it the most complex data set most Internal Models will use. There are upgrades, downgrades, and defaults which each have complex probability distributions and relations between each other. Downgrade and defaults tend to be fat tailed with excess Kurtosis (i.e. a higher than for a Normal distribution). The probabilities of each of these events can vary significantly over time.

For the purpose of detailed empirical data analysis, we have used publicly available data:

- 1932 Moody's transition matrix.
- 1931–1935 Moody's average transition matrix during the Great Depression.
- 1981–2019 S&P transition matrix data.

Figures 1, 2, 3 and 4 show the 1932 values compared to the 1981–2019 data. The 1931–1935 transition matrix has been used in the model calibrations in this paper but is not shown in the plot below.

The above analysis shows:

- When comparing the types of transitions:
 - For investment grade ratings, the probability of downgrade is more significant, with defaults forming a very small percentage of transitions (although note the scale of the asset loss is much more material for defaults than for transitions.)
 - Defaults are shown to be much more material at the sub-investment grade ratings.

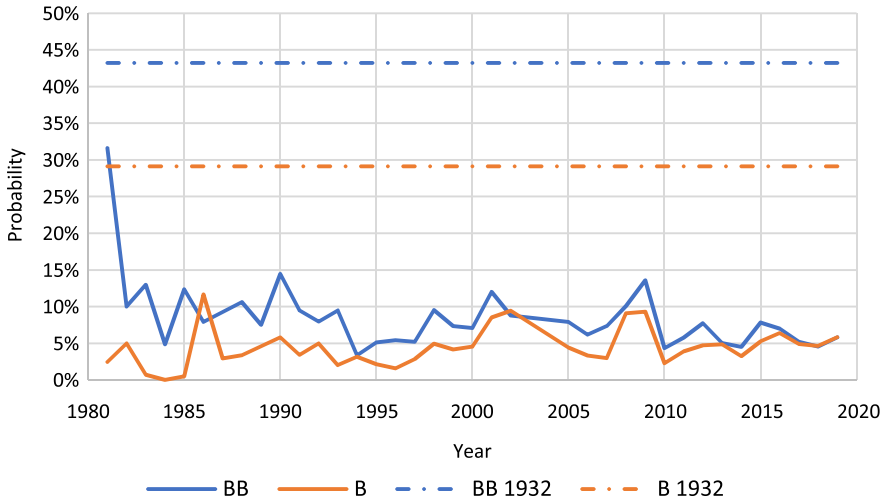


Figure 2. Downgrade rates for sub-investment-grade assets.

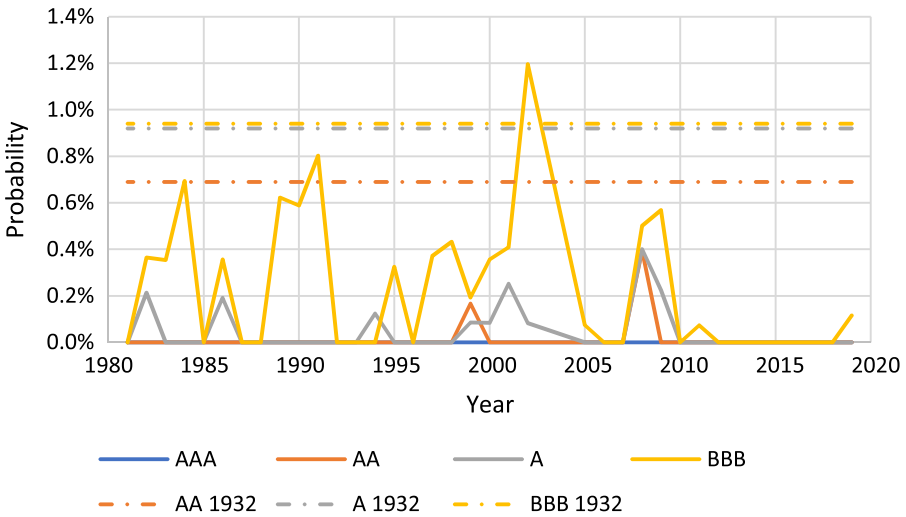


Figure 3. Default rates for investment-grade assets.

- When comparing across years:
 - The 1932 matrix is shown as straight lines across all the plots for each rating to compare with the 1981–2019 period. The 1932 matrix was worse than any in the more recent period 1981–2019.
 - 2009 and 2001 show relatively high levels of default and downgrade, which is expected given the financial crisis and dot com bubble respectively.

Table 4 shows the mean, standard deviation, skewness and excess kurtosis for the upgrades, downgrades and defaults based on data from 1981 to 2019 including the 1932 transition matrix.

Table 4. First four moments for downgrades, upgrades, and defaults.

Upgrades	AAA	AA	A	BBB	BB	B	CCC/C
Mean		0.5%	2.0%	4.6%	6.3%	6.2%	15.1%
Standard Deviation		0.6%	1.3%	2.6%	2.9%	3.2%	9.7%
Skew		1.1	1.2	1.3	1.0	1.4	1.0
Excess kurtosis (above 3)		-0.0	1.0	1.6	1.8	3.2	0.7
Defaults	AAA	AA	A	BBB	BB	B	CCC/C
Mean		0.03%	0.07%	0.25%	1.10%	4.88%	25.74%
Standard Deviation		0.13%	0.17%	0.31%	1.43%	3.87%	12.83%
Skew		4.38	3.87	1.25	2.12	1.18	0.07
Excess kurtosis (above 3)		19.90	17.60	1.06	4.63	0.75	0.25
Downgrades	AAA	AA	A	BBB	BB	B	CCC/C
Mean		10.9%	9.0%	6.7%	6.3%	9.4%	5.1%
Standard Deviation		10.7%	7.3%	6.3%	7.0%	7.4%	4.7%
Skew		2.0	3.2	4.7	4.8	3.5	3.7
Excess kurtosis (above 3)		4.0	14.2	26.0	26.2	13.6	18.0

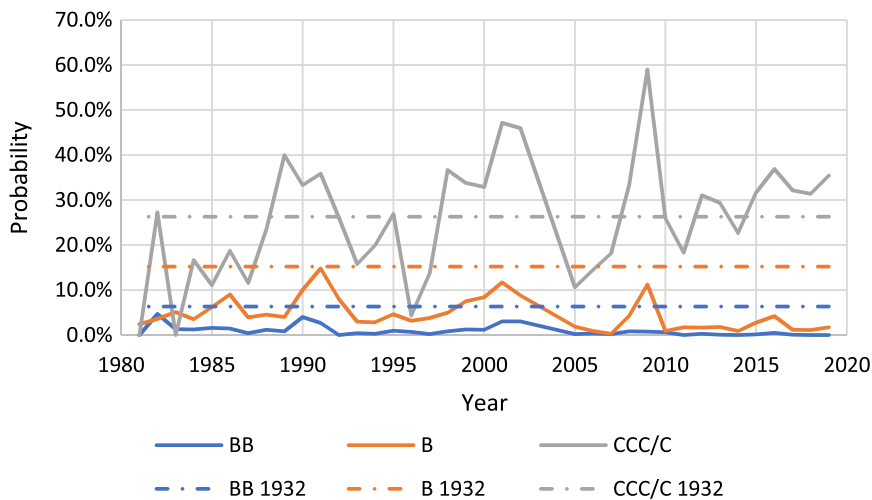


Figure 4. Default rates for sub-investment-grade assets.

The main comments on the first four moments for upgrades, downgrades and defaults for each credit rating are:

- For upgrades the mean and standard deviation increase as the ratings decrease. Each rating has a slightly positive skew; excess kurtosis is either close to zero or slightly above zero.
- For downgrades, the mean and standard deviation decrease as the ratings decrease. The positive skewness is higher than for upgrades and the excess kurtosis is very high, indicating non-normal characteristics.

- For defaults, the mean and standard deviation rise significantly as the ratings fall, with the mean default for AAA at zero, and by CCC/C at 25.7% of bonds defaulting within a year. The higher-rated assets have a more positive skewness, which gradually falls from AA to CCC/C. The AA and A ratings have a very high excess kurtosis with occasional defaults and long periods of no default from these ratings.
- The ratings above CCC are more likely to downgrade/default than to upgrade. This feature is specifically captured in the two-factor model later in the paper with the “Optimism” parameter.

5. Models Explored

Four credit models are described in detail, split between parametric and non-parametric models.

5.1. Parametric Models

For parametric models, the systemic components of transition matrices are expressed as a function of a small number of parameters. In this Section two parametric models are discussed:

- Vašíček (calibrated using Belkin approach)
- Two-factor model (a model introduced in this paper)

5.1.1. The Vašíček model

Oldrich Vašíček first considered the probability of loss on loan portfolios in 1987 (Vašíček, 1987). Starting from Merton’s model of a company’s asset returns (Merton, 1974), the question Vašíček was seeking to answer was relatively simple: what is the probability distribution of default for a portfolio of fixed cashflow assets?

Vašíček required several assumptions for the portfolio of assets:

- All asset returns are described by a Wiener process. In other words, all asset values are lognormal distributed, similar to Merton’s approach.
- All assets have the same probability of default p .
- All assets are of equal amounts.
- Any two of the assets are correlated with a coefficient ρ (rho).

The starting point in Vašíček’s model was Merton’s model of a company’s asset returns, defined by the formula:

$$\ln A(T) = \ln A + \mu T - \frac{1}{2} \sigma^2 T + \sigma \sqrt{T} X \quad (1)$$

where T is the maturity of the asset, $W(t)$ is standard Brownian motion, asset values (denoted $A(t)$) are lognormal distributed, μ and σ^2 are the instantaneous expected rate and instantaneous variance of asset returns respectively, and X represents the return on a firm’s asset. In this setting, X follows a standard normal distribution, given by $X = \frac{W(T) - W(0)}{\sqrt{T}}$.

The next step in Vašíček’s model was to adapt Merton’s single-asset model to a portfolio of assets. For a firm denoted i (with $i = 1, \dots, n$), Equation (1) can be rewritten as:

$$\ln A_i(T) = \ln A_i + \mu_i T - \frac{1}{2} \sigma_i^2 T + \sigma_i \sqrt{T} X_i \quad (2)$$

Given the assumptions above, and the equi-correlation assumption for variables X_i , it follows that the variables X_i belong to an equi-correlated standard normal distribution (equi-correlation

means all assets in the portfolio are assumed to have the same correlation with one another). Any variable X_i that belongs to an equi-correlated standard normal distribution can be represented as a linear combination of jointly standard normal random variables Z and Y_i such that:

$$X_i = Z\sqrt{\rho} + Y_i\sqrt{1 - \rho}, \text{ where } i = 1, \dots, n \tag{3}$$

Equation (3) is a direct result of statistical properties of jointly equi-correlated standard normal variables, which stipulates that any two variables X_i and X_j are bivariate standard normal with correlation coefficient ρ if there are two independent standard normal variables Z and Y for which $X_i = Z$ and $X_j = \rho Z + \sqrt{1 - \rho^2}Y$, with ρ a real number in $[-1, 1]^2$.

Note that it can be shown that the common correlation of n random variables has a lower bound equal to $-\frac{1}{n-1}$. As n tends to infinity, ρ will have a lower bound of 0, also known as the zero lower bound limit of common correlation. In other words, for (very) large portfolios, firms' assets can only be positively correlated, as is their dependence on systematic factors.

With each firms' asset return X_i of the form $X_i = Z\sqrt{\rho} + Y_i\sqrt{1 - \rho}$, variable Z is common across the entire portfolio of assets, while Y_i is the i^{th} firm's specific risk, and independent from variable Z and variables Y_j , where $j \neq i$.

As a parenthesis, the covariance between two firms' asset returns X_i and X_j is determined by $\rho_{ij} = \frac{\text{cov}(X_i, X_j)}{\sigma(X_i)\sigma(X_j)}$, where $\sigma(X_i)$ and $\sigma(X_j)$ are the standard deviations of each firm's asset returns. For a fixed ρ , a higher variance of asset returns requires a higher covariance of asset returns and vice-versa. For standard normal variables, $\sigma(X_i) = \sigma(X_j) = 1$, and hence $\rho_{ij} = \text{cov}(X_i, X_j)$.

The final step in Vašíček's model is the derivation of a firm's probability of default, conditional on the common factor Z . This is relatively straightforward:

$$P(\text{firm } X_i \text{ defaults} \mid Z) = \Phi\left(\frac{x_i - Z\sqrt{\rho}}{\sqrt{1 - \rho}}\right) \tag{4}$$

Finally, although Vašíček has not considered credit ratings in his setting other than the default state, from Equation (4) it follows that, conditional on the value of the common factor Z , firms' loss variables are independent and equally distributed with a finite variance.

The loss of an asset portfolio (e.g. a portfolio represented in a transition matrix) can thus be represented by a single variable on the scaled distribution of variable X_i . Although overly simplistic, this setup is helpful for analysing historical data (such as historical default rates or transition matrices) and understanding the implications of asset distributions and correlations in credit risk modelling. In the following Section, we consider a method that applies a firm's conditional probability of default to historical transition matrices.

5.1.1.1. Vašíček model calibration - Belkin. Belkin et al. (1998) introduced a statistical method to estimate the correlation parameter ρ and common factors Z in Vašíček's model based on historical transition matrices. The starting point in their model is the one-factor representation of annual transition matrices (denoted by variables X_i , with i representing year):

$$X_i = Z\sqrt{\rho} + Y_i\sqrt{1 - \rho}$$

where Z , Y_i and ρ are as per Vašíček's framework, and X_i is the standardised asset return of a portfolio in a transition matrix.

The method proposed in Belkin employs a numerical algorithm to calibrate the asset correlation parameter ρ and systematic factors Z subject to meeting certain statistical properties (e.g. the unit variance of Z on the standard normal distribution), using a set of historical transition matrix data.

²See example 5.36 at https://www.probabilitycourse.com/chapter5/5_3_2_bivariate_normal_dist.php.

This approach allows a transition matrix to be represented by a single factor, representing that transition matrix's difference from the average transition matrix. Matrices with more downgrades and defaults are captured with a high negative factor; matrices with relatively few defaults and downgrades have a positive factor.

Representing a full transition matrix (of 49 probabilities) with a single factor inevitably leads to loss of information. More information can be captured in a two-factor model, which is introduced in Section 5.1.2.

The Belkin implementation uses the standard normal distribution. If this distribution was replaced with a fatter-tailed distribution, it could be used to strengthen the calibration of extreme percentiles (e.g. the 99.5th percentile). However, other moments of the distribution are also expected to be impacted (e.g. the mean) which would need to be carefully understood before implementation. This approach has not been explored in this paper, but it would not be expected to change the directional results seen in Section 6.

5.1.2. The two-factor model

5.1.2.1. Two-factor model description. The Vašíček model includes just a single parameter used to model transition matrices and some of the limitations of the model arise from an oversimplification of the risk. In this Section, a two-factor model is described to capture two important features of the way transition matrices change over time, particularly in stress. This model is based on a description given by Rosch and Scheule (2008).

Rosch and Scheule (2008) use two defined features of a transition matrix they describe as "Inertia" and "Bias." In this paper we use the terms "Inertia" and "Optimism" as the term bias is widely used in statistics, potentially causing confusion with other uses such as statistical bias in parameter estimation.

Inertia is defined as the sum of the leading diagonal of the transition matrix. For a transition matrix in our setting with probabilities p_{ij} , where i denotes row and j denotes column:

$$Inertia = \sum_{i=1}^7 p_{ii}$$

Optimism is defined as the ratio between the upgrade probabilities and downgrade probabilities summed over all seven rows and weighted by the default probabilities in each row:

$$Optimism = \sum_{i=2}^6 \left(\frac{\sum_{i < j} p_{ij}}{\sum_{i > j} p_{ij}} * p_{iD} \right) / \sum_{i=2}^6 (p_{iD})$$

p_{iD} is the default probability for each row and this value is summed across. $\sum_{i < j} p_{ij}$ is the sum of all upgrades in each row. $\sum_{i > j} p_{ij}$ is the sum of all downgrades in each row. Appendix B gives a sample calculation of Inertia and Optimism for a given matrix.

Any historical matrix can be characterised by these two factors and the base transition matrix (a long-term average transition matrix used as a mean in the best estimate) can be adjusted so that its Inertia and Optimism correspond to those in an historical transition matrix. This allows the generation of an historical time series for Inertia and Optimism based on our historical data set; then to fit probability distributions to the values of these parameters that can then be combined with a copula.

It would also be possible to extend the model by weighting the values of Inertia and Optimism by the actual assets held in the portfolio (but this is not explored in this paper). In this paper, Optimism has only been calculated based on AA-B ratings; but in practice, this could be changed to be closer to the actual assets held.

This means that with probability distributions for Inertia, Optimism and a copula all calibrated from historical data, a full probability distribution of transition matrices can be produced.

The base transition matrix can be adjusted so that its Inertia and Optimism are consistent with the parameters of a historical matrix or parameters outputted from a probability distribution by using the following steps (BaseInertia and StressInertia and BaseOptimism and StressOptimism are used for the base Inertia and Optimism and the Inertia and Optimism from the matrix the base matrix is adjusted to have the same values as):

1. Multiply each of the diagonal values by StressInertia/BaseInertia.
2. Adjust upgrades and downgrades so the rows sum to 1, by dividing them by a single value.
3. Adjust upgrades and downgrades so that their ratio is now in line with StressOptimism.

The adjustments required for steps 2 and 3 above are now defined. To do this the matrices required to calculate the adjustments are first defined.

- Elements of the base transition matrix are defined as $p_{ij}^{(1)}$ for the i^{th} row and j^{th} column from this matrix.
- Elements of the matrix after step 1 (having the diagonals adjusted by StressInertia/BaseInertia) are defined by $p_{ij}^{(2)}$.
- Elements of the matrix after step 2 (upgrades and downgrades adjusted to sum to 1) are defined by $p_{ij}^{(3)}$.
- Elements of the matrix after step 3 (upgrades and downgrades adjusted so optimism is equal to StressOptimism) are defined by $p_{ij}^{(4)}$.

Following step 1, upgrades and downgrades in step 2 are given by:

$$p_{ij}^{(3)} = p_{ij}^{(2)} * \frac{1 - p_{ii}^{(2)}}{\sum_{k=1(k \neq i)}^7 p_{ik}^{(2)}}$$

This gives a matrix which has the same Inertia as the StressInertia, but the Optimism is still not the same as the StressOptimism.

To carry out step 3 no change is needed for the AAA or CCC/C categories, which have no upgrades nor downgrades respectively. For the other ratings, a single factor is found to add to the sum of the upgrades and subtract from the sum of the downgrades to give new upgrades and downgrades, which have the same ratio as the value of StressOptimism. This factor is:

$$Factor = \frac{(StressOptimism * \sum_{j=i+1}^7 p_{ij}^{(3)} - \sum_{j=1}^{i-1} p_{ij}^{(3)})}{(1 + StressOptimism)}$$

The final matrix can now be found by:

$$\begin{aligned}
 \text{if } j > i \text{ (i.e. downgrades)} \quad p_{ij}^{(4)} &= \frac{p_{ij}^{(3)}}{\sum_{j=i+1}^7 p_{ij}^{(3)}} * \left(\sum_{j=i+1}^7 p_{ij}^{(3)} - Factor \right) \\
 \text{if } j < i \text{ (i.e. upgrades)} \quad p_{ij}^{(4)} &= \frac{p_{ij}^{(3)}}{\sum_{j=1}^{i-1} p_{ij}^{(3)}} * \left(\sum_{j=1}^{i-1} p_{ij}^{(3)} - Factor \right)
 \end{aligned}$$

5.1.2.2. *Two-factor model calibration.* This Section describes an approach to calibrating the two-factor model.

Using historical data for historical transition matrices from 1981 to 2019, as well as the 1932 matrix and a matrix from 1931 to 1935, a time series of Inertia and Optimism can be constructed. It is then possible to fit probability distributions from this data and simulate these probability

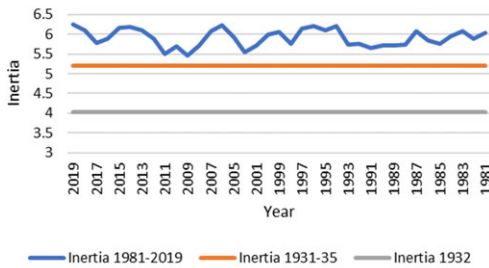
Table 5. First four moments of inertia and optimism.

Moments	Inertia	Optimism
Mean	5.84	0.89
Standard Deviation	0.383	0.67
Skew	-2.79	1.27
Excess kurtosis (above 3)	10.8	1.9

Table 6. Correlation between inertia and optimism.

Correlation	Value
Pearson	0.33
Spearman	0.43
Kendall Tau	0.30

(a) Inertia 1981-2019 compared to 1932 and 1931-35



(b) Optimism 1981-2019 compared to 1932 and 1931-35

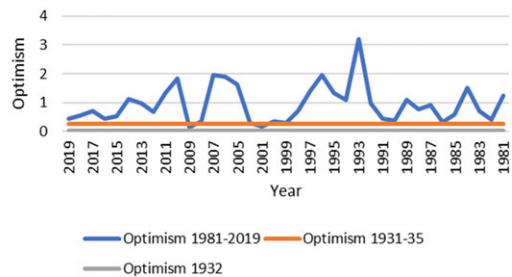


Figure 5. (a) Inertia and (b) Optimism, historical values compared to 1932 and 1931-1935.

distributions, combined with the approach of adjusting the base transition matrix for given Inertia and Optimism to give a full probability distribution of transition matrices.

The data is shown in Figure 5 with the time series compared to the values for 1932 and 1931-1935.

The first four moments for this data are shown in Table 5 (including the 1932 and 1931-1935 matrices).

The correlations between the two data series are given in Table 6:
The main comments on these data series are:

- Inertia is negatively skewed and has a fat-tailed distribution.
- Optimism is slightly positively skewed and with a slightly fatter tail than the normal distribution.
- The two data series are correlated based on the Pearson, Spearman or Kendall Tau measure of correlation. In the most extreme tail event, both Inertia and Optimism were at their lowest values. This means that this year had the most amount of assets changing rating as well as the greatest number of downgrades relative to upgrades in that year.

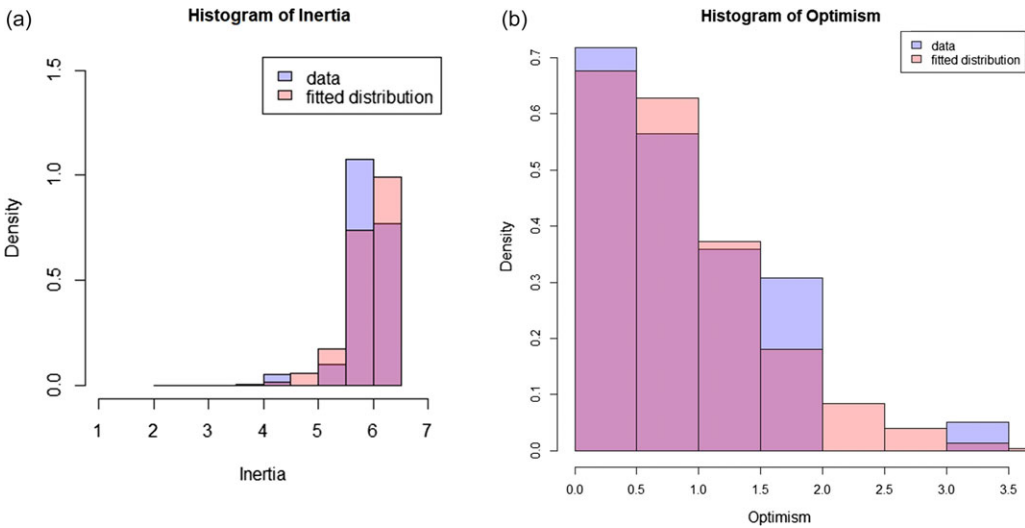


Figure 6. Historical plots of (a) Inertia and (b) Optimism compared to fitted distributions.

To simulate from these data sets probability distributions have been fitted to the two data series using the Pearson family of probability distributions. The Pearson Type 1 distribution produces a satisfactory fit to the two data sets as shown in Figure 6, comparing the raw data (“data”) against 10,000 simulated values from the fitted distributions (“fitted distribution”).

The plot above shows the actual historical data values for Inertia and Optimism from each of the historical transition matrices compared to the distributions fitted to this data. Note that the darker pink indicates where both the data (blue) and fitted distribution (pink) overlap.

As well as probability distributions for Inertia and Optimism, a copula is also needed to capture how the two probability distributions move with respect to one another. For this purpose, a Gaussian Copula has been selected with a correlation of 0.5. This correlation is slightly higher than found in the empirical data. The model could be extended to use a more complex copula such as the t-copula instead of the Gaussian copula. An alternative to a one or two-factor model is to use non-parametric models and, rather than fit a model to the data, use the data itself directly to generate a distribution for the risk.

5.2. Non-Parametric Models

While the Vašiček and two-factor models involve a specific functional form for transition matrices, there are also non-parametric methods for constructing distributions of transition matrices. These are multivariate analogues to the concept of an empirical distribution function, in contrast to formulas such as Vašiček’s, which are akin to the fitting of a parametric distribution family. Two different non-parametric models are considered; the first involves dimension reduction using the K-means algorithm; the second is a bootstrapping approach, whereby historical transition matrices are simulated many times with replacement, to get a full risk distribution.

5.2.1. The K-means model

Under the K-means model the key steps are:

1. Apply the K-means algorithm to the data to identify a set of groups within the data set and decide how many groups are required for the analysis.

2. Assign each of the groups to real-line percentiles manually, e.g. assign a group containing the 1932 matrix at the 0.5th percentile, assign the average matrix of the 1931–1935 matrix at 0.025th percentile, put an identity matrix at the 100th percentile, put a square of the 1932 matrix at the 0th percentile, etc.
3. Interpolate any percentiles needed in between using a matrix interpolation approach.

5.2.1.1. K-means to transition data. K-means clustering is an unsupervised clustering algorithm that is used to group different data points based on similar features or characteristics. K-means clustering is widely used when un-labelled data (i.e. data without defined categories or groups) needs to be organised into a number of groups, represented by the variable K (Trevino, 2016). The algorithm works iteratively to assign each data point to one of K groups based on the features that are provided. The results of the K-means clustering algorithm are:

1. The means of the K clusters can be used to label new data.
2. Labels for the training data (each data point is assigned to a single cluster).
3. Rather than defining groups before looking at the data, clustering allows you to find and analyse groups that have formed organically. The “Choosing K” Section below describes how the number of groups can be determined.

Based on the transition risk data, we apply the K-means algorithm to group each of the transition year data points into groups with a similar profile of transitions from one rating to another. We present the K-means visualisations in Figure 7. In K-means clustering, as we increase the number of clusters, generally the sum of squares within and between groups reduces. The key idea is to optimise the number of groups (i.e. value of K) such that reduction in the sum of squares within and between the groups stops reducing substantially for higher numbers of groups.

Further details of the K-means clustering algorithm are given in Appendix A.

As shown in Figure 7:

- 1931 and 1931–1935 average matrices are in separate groups as we increase the groups from 7 onwards.
- Below $K = 7$ groups, the average 1931–1935 matrix does not come out as a separate group in its own right.

As shown in Figure 8, the total within clusters sum of squares reduces significantly as we increase the number of groups from $K = 2$ to $K = 6$. The reduction in total within clusters sum of squares does not reduce materially after $K = 8$. We have applied equal weights to each of the transition matrices.

5.2.1.2. Manually assign the matrix group to percentiles. Once K-means has been run, we get each of the transition matrices assigned to a group, as shown in Figure 7. For our example, we have selected $K = 8$. The next step is to manually assign each of the groups to a percentile value on the empirical distribution. This is set via an expert judgement process as follows:

- Square of 1932 matrix is assigned to the 0th percentile.
- 1932 matrix is assigned to the 0.5th percentile.
- 1932–1935 average matrix is assigned to the 1.25th percentile.
- A Group containing the 2002 matrix is assigned to the 8.75th percentile.
- A Group containing the 2009 matrix is assigned to the 27.5th percentile.
- A Group containing the 1981 matrix is assigned to the 48.75th percentile.
- A Group containing the 2016 matrix is assigned to the 66.25th percentile.
- A Group containing the 2017 matrix is assigned to the 81.25th percentile.

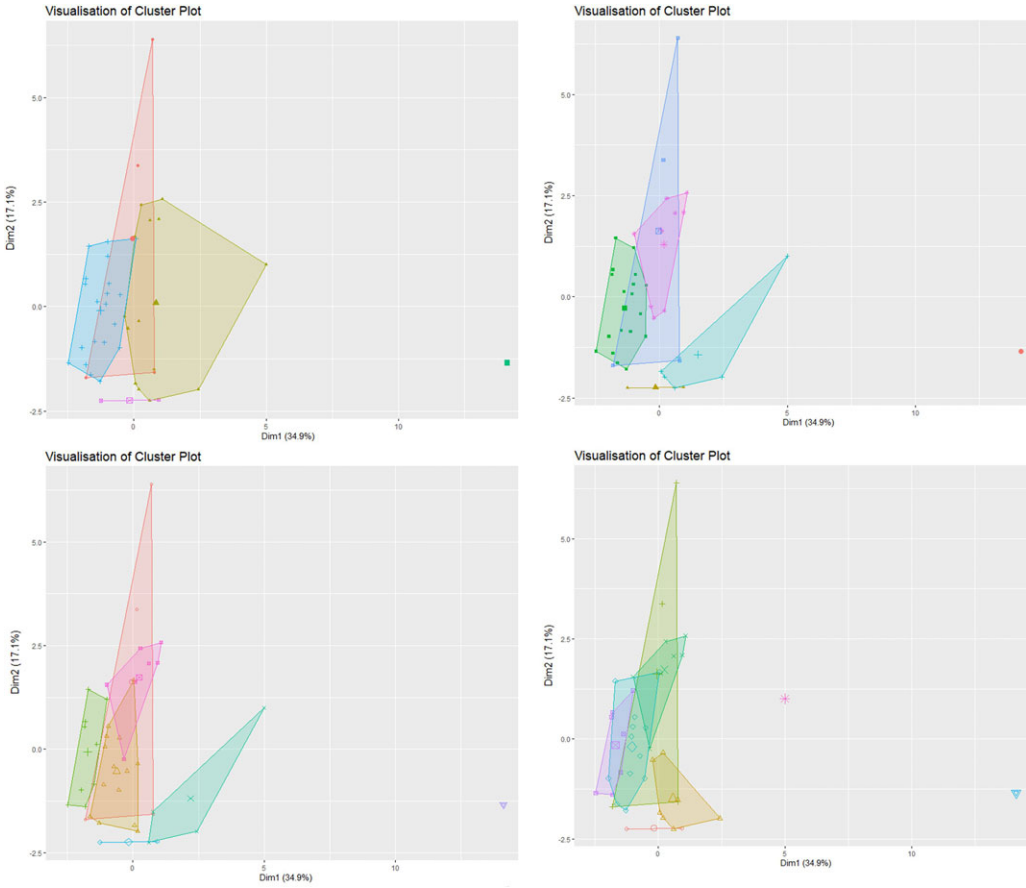


Figure 7. K-Means clustering. Examples where $K = 5, 6, 7$ and 8 .

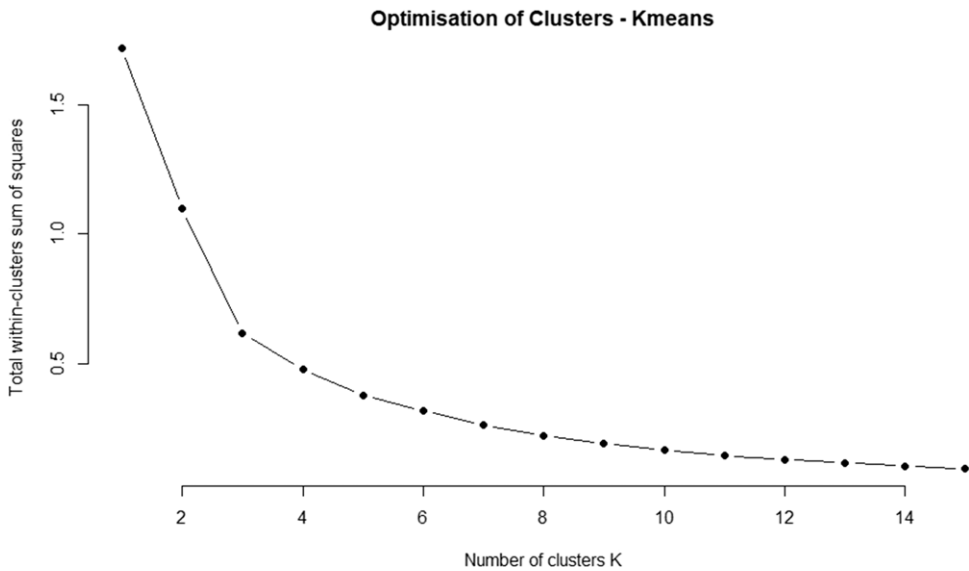


Figure 8. K-means clustering examples with different K values – sum of squares within clusters.

- A Group containing the 2019 matrix is assigned to the 93.75th percentile.
- An identity matrix is assigned to the 100th percentile.

All other percentiles are derived using a matrix interpolation approach as follows, where P_1 is the known matrix at percentile p_1 , P_2 is the known matrix at percentile p_2 , and Q is the interpolated matrix at percentile q .

$$\text{Interp} = \left(\frac{p_2 - q}{p_2 - p_1} \right)$$

$$E_1 = \text{Eigen}(P_1) * \text{Diag}(\text{EigenVal}(P_1))^{\text{Interp}} * \text{Eigen}(P_1)^{-1}$$

$$E_2 = \text{Eigen}(P_2) * \text{Diag}(\text{EigenVal}(P_2))^{(1-\text{Interp})} * \text{Eigen}(P_2)^{-1}$$

$$Q = E_1 * E_2$$

5.2.2. Bootstrapping

The bootstrapping model refers to the relatively simplistic approach of sampling from the original data set with replacement. In this case, there are n transition matrices from which 100,000 random samples are taken with replacement to give a distribution of transitions and defaults.

This is a very simple model and the main benefit of being true to the underlying data without many expert judgements and assumptions. A significant downside of this model is it cannot produce scenarios worse than the worst event seen in history; this means it is unlikely to be useful for Economic Capital models where the extreme percentiles are a crucial feature of the model. Nevertheless, this model is included for comparison purposes as it is very close in nature to the underlying data.

6. Comparison of the Models Explored

In this Section, several metrics are used to compare the models described in Section 5:

1. Do the models show movements of a similar nature to historical data capturing the types of stresses seen historically?
2. Do the models produce extreme percentiles sufficiently extreme to pass back-testing?
3. Are the models calibrated in an objective, easily definable way?

The first of these metrics requires an assessment of how the model transition matrices compare to historical transition matrices. In particular, how do the model transition matrices move over time compared to how the historical transition matrices move?

Sections 6.1, 6.2 and 6.3 describe how the movements of transition matrices generated by the models can be compared to historical data. Section 6.4 shows how the various models described in Section 5 compare in terms of this metric. Section 6.5 shows how the models compare to historical back-tests expected under UK regulatory frameworks. Section 6.6 compares the models in terms of the amount of expert judgement required to calibrate them.

6.1. Dimension Reduction and Visualisation

6.1.1. The need to reduce dimensions

Transition matrix modelling is a high-dimensional exercise. It is almost impossible to visualise the 56-dimensional distribution of a random 7×8 matrix. To make any progress comparing models

we need to reduce the number of dimensions while endeavouring to mitigate the loss of information.

Dimension reduction is even more necessary because commonly used transition models employ a low number of risk drivers, in order to limit calibration effort, particularly the need to develop correlation assumptions with other risks within an internal model. Vašíček's model, for example, has a single risk driver when a portfolio is large. If we have 7 origin grades and 8 destination grades (including default, but excluding NR), we could say that the set of feasible matrices under Vašíček's model is a one-dimensional manifold (i.e. a curve) in 56-dimensional space. This is a dramatic dimension reduction relative to the historical data.

We can simplify matters to some extent by modelling the transition matrices row-by-row, considering different origin grades separately. This is possible because the investment mandates for many corporate bond portfolios dictate a narrow range of investment grades most of the time, with some flexibility to allow the fund time to liquidate holdings that are re-graded out with the fund mandate. In that case, we are dealing with a few 8-dimensional random variables; still challenging but not as intimidating as 56 dimensions.

Popular transition models generally calibrate exactly to a mean transition matrix so that the means of two alternative consistently calibrated models typically coincide. It is the variances and covariances that distinguish models.

6.1.2. *Principal components analysis*

Principal components analysis is a well-known dimension-reduction technique based on the singular value decomposition of a variance-covariance matrix. A common criticism of PCA, valid in the case of transition modelling, is that it implicitly weights all variances equally, implying that transitions (such as defaults) with low frequency but high commercial impact have little effect on PCA results. We propose a weighted PCA approach which puts greater weight on the less frequent transitions.

Standard PCA can also be distorted by granulation, that is lumpiness in historical transition rates caused by the finiteness of the number of bonds in a portfolio. We now describe granulation in more detail and show how a weighted PCA approach, applied one origin grade at a time, can reveal the extent of granulation.

6.2. *Granulation*

6.2.1. *Systematic and granulated models*

Some theoretical models start with transition probabilities for an infinitely large portfolio (the systematic model), and then use a granulation procedure (such as a multinomial distribution) for bond counts so that, for example, each destination contains an integer number of bonds. Other models such as that of Vašíček are specified at the individual bond level and then the systematic model emerges in the limit of diverse bond portfolios.

It is possible that two transition models might have the same systematic model, differing only in the extent of granulation. It could also be that discrepancies between a proposed model and a series of historical matrices are so large that granulation cannot be the sole explanation. It is important to develop tests to establish when model differences could be due to granulation.

6.2.2. *Granulation frustrates statistical transformations*

When a theoretical model puts matrices on a low-dimensional manifold, granulation can cause noise in both historical matrices and simulated future matrices, which are scattered about that systematic manifold. Granulation complicates naïve attempts to transform historical transition data. For example, under Vašíček's model, the proportion of defaults (or transitions to an X-or-worse set of grades) is given by an expression involving a cumulative normal distribution whose

argument is linear in the risk factor. We might attempt to apply the inverse normal distribution functions to historical default rates and then reconstruct the risk factor by linear regression. However, when the expected number of bond defaults is low, the observed default rate in a given year can be exactly zero, so that the inverse normal transformation cannot be applied.

6.2.3. *Mathematical definition of granulation*

For $n \geq 2$, let S_n denote the n -simplex, that is the set of ordered $(n+1)$ -tuples $(x_0, x_1, x_2 \dots x_n)$ whose components are non-negative and sum to one.

We define a granulation to be a set of probability laws $\{\mathbb{P}_x: x \in S_n\}$ taking values in S_n such that if a vector Y satisfies $Y \sim \mathbb{P}_x$ then:

$$\begin{aligned} \mathbb{E}(Y_i) &= x_i \\ \text{Cov}(Y_i, Y_j) &= hx_i(\delta_{ij} - x_j) = \begin{cases} hx_i(1 - x_i) & i = j \\ -hx_ix_j & i \neq j \end{cases} \end{aligned}$$

The parameter h , which must lie between 0 and 1 is the Herfindahl index (Herfindahl 1950) of the granulation.

6.2.4. *Granulation examples*

One familiar example of a granulation is a multinomial distribution with n bonds and probabilities x_i in which case the Herfindahl index is n^{-1} . In the extreme case where $n = 1$, this is a categorical distribution where all probability lies on the vertices of the simplex. In the other extreme, as n becomes large, the law \mathbb{P}_x is a point mass at x .

Other plausible mechanisms for individual matrix transitions also conform to the mathematical definition of a granulation. For example, if bonds have different face values, we might measure transition rates weighted by bond face values. Provided the bonds are independent, this is still a granulation with the standard definition of the Herfindahl index. In a more advanced setting, we might allocate bonds randomly to clusters, with all bonds in each cluster transitioning in the same way, but different clusters transitioning independently. This too satisfies the covariance structure of a granulation. Transition models based on Dirichlet (multivariate-beta) distributions are granulations, with h^{-1} equal to one plus the sum of the alpha parameters. Finally, we can compound two granulations to make the third granulation, in which case the respective Herfindahl indices satisfy:

$$1 - h_3 = (1 - h_1)(1 - h_2)$$

6.2.5. *Granulation effect on means and variances of transition rates*

We now investigate the effect of granulation on means and variance matrices of simplex-valued random vectors. Suppose that X is a S_n -valued random vector, representing the systematic component of a transition model and that Y is another random vector with $Y|X \sim \mathbb{P}_X$ for some granulation.

Let us denote the (vector) mean of X by

$$\pi = \mathbb{E}(X)$$

and the variance (-covariance) matrix of X by

$$\mathbf{V}^{\text{sys}} = \text{Var}(X)$$

Then it is easy to show that the mean of Y is the same as that of X

$$E(Y) = \pi$$

And the variance(-covariance) matrix of Y is:

$$V_{ij}^{gran} = (1 - h)V_{ij}^{sys} + h\pi_i(\delta_{ij} - \pi_j)$$

6.3. Weighted Principal Components

6.3.1. Weighting proposal

We are now able to propose a weighted principal components approach for models of simplex-valued transition matrices.

Suppose then that we have a model with values in S_n . Its mean vector π is, of course, still in S_n . Let us denote the variance matrix by V . Our proposed weighted PCA method is based on a singular value decomposition of the matrix.

$$diag(\pi)^{-\frac{1}{2}}Vdiag(\pi)^{-\frac{1}{2}}$$

As this is a real positive-semidefinite symmetric matrix, the eigenvalues are real and non-negative. The simplex constraint in fact implies no eigenvalue can exceed 1 (which is the limit of a categorical distribution). We can without loss of generality take the eigenvectors to be orthonormal. We fix the signs of eigenvectors such that the component corresponding to the default grade is non-positive so that a positive quantity of each eigenvector reduces default rates (and a negative quantity increases defaults). This is consistent with our definitions of Optimism and Inertia in Section 5.1.2.1.

As the components of a simplex add to 1, it follows that $V\mathbf{1} = \mathbf{0}$ where $\mathbf{1}$ is a vector of 1s. This implies that the weighted PCA produces a trivial eigenvector e with eigenvector zero and

$$e^{triv} = \pi^{\frac{1}{2}} = diag(\pi)^{\frac{1}{2}}\mathbf{1}$$

6.3.2. Weighted PCA and granulation

Suppose now that we have a non-trivial eigenvector e of the weighted systematic matrix with eigenvalue λ^{sys} so that

$$diag(\pi)^{-\frac{1}{2}}V^{sys}diag(\pi)^{-\frac{1}{2}}e = \lambda^{sys}e$$

It is easy to show that e is also an eigenvector of any corresponding granulated model, with transformed eigenvalue shrunk towards 1.

$$\lambda^{gran} = (1 - h)\lambda^{sys} + h$$

Thus, if one model is a granulation of another, the weighted eigenvectors are the same and the eigenvalues are related by a shrinkage transformation towards 1. This elegant result is the primary motivation for our proposed weighting.

PCA usually focuses on the most significant components, that is those with the largest associated eigenvalues. In the context of a transition matrix, the smallest (non-zero) eigenvalue of a granulated model has a role as an upper bound for the Herfindahl index of any granulation. Where the systematic model inhabits a low-dimensional manifold, the smallest non-zero eigenvalue is typically close to zero, which implies that the smallest non-zero eigenvalue of a granulated model is a tight upper bound for the Herfindahl index.

Knowing the Herfindahl index allows us to strip out granulation effects to reconstruct the variance matrix of an underlying systematic model.

Table 7. Eigenvalues for each of the four models and the raw data.

Single A	Eigenvalues					Single A	% Variance explained by each PC				
	PC1	PC2	PC3	PC4	PC5		PC1	PC2	PC3	PC4	PC5
Raw data	0.075	0.008	0.003	0.002	0.001	Raw data	84%	9%	3%	3%	1%
Bootstrap	0.073	0.007	0.003	0.002	0.001	Bootstrap	84%	9%	3%	3%	1%
K-means	0.049	0.003	0.001	0.001	0.000	K-means	90%	6%	3%	1%	1%
Twofactor	0.056	0.012	0.000	0.000	0.000	Twofactor	82%	18%	0%	0%	0%
Vašíček	0.029	0.004	0.000	0.000	0.000	Vašíček	88%	11%	0%	0%	0%
BBB	Eigenvalues					BBB	% Variance explained by each PC				
	PC1	PC2	PC3	PC4	PC5		PC1	PC2	PC3	PC4	PC5
Raw data	0.085	0.016	0.004	0.003	0.002	Raw data	77%	15%	4%	3%	2%
Bootstrap	0.083	0.016	0.004	0.003	0.002	Bootstrap	77%	15%	4%	3%	2%
K-means	0.048	0.007	0.004	0.001	0.000	K-means	81%	11%	6%	1%	0%
Twofactor	0.064	0.007	0.000	0.000	0.000	Twofactor	90%	10%	0%	0%	0%
Vašíček	0.031	0.005	0.000	0.000	0.000	Vašíček	87%	13%	0%	0%	0%

6.3.3. PCA and independence

PCA decomposes a random vector into components whose coefficients are uncorrelated.

It is well known that lack of correlation does not imply independence. Nevertheless, in some contexts, the component loadings emerging from PCA might be analysed separately and then recombined as if they were independent. In this way, the PCA is sometimes used as a step in a model construction procedure.

Model construction via PCA does not work well for simplex-valued transition models. As the simplex is bounded, typically the component loadings have distributions on closed intervals. Recombining component models as if they were independent then produces distributions in a hyper-cuboid. The hyper-cuboid cannot represent a simplex; either some points of the hyper-cuboid poke beyond the simplex, leading to infeasible negative transition rates, or some feasible parts of the simplex poke beyond the hyper-cuboid rendering those events inaccessible to the model.

For these reasons, we do not propose PCA as a way of constructing transition models. Rather, we advocate the use of PCA to analyse historical transition matrices and to compare models that have been constructed by other means.

6.4. Weighted PCA Applied to Our Fitted Models

In this Section we present the eigenvectors and eigenvalues from the PCA approach applied to the four models and the raw data. To apply PCA to the models, 10,000 simulated transition matrices were generated for each model and PCA was applied to this data. The eigenvectors are shown for the single A and BBB-rated movements.

The eigenvalues can be converted to the proportion of the model variance explained by each principal component by dividing that eigenvalue by the sum of the eigenvalues for that model. The raw data shows that most of the variance is in the first principal component with a smaller amount in PC2 and much smaller amounts in the final three components.

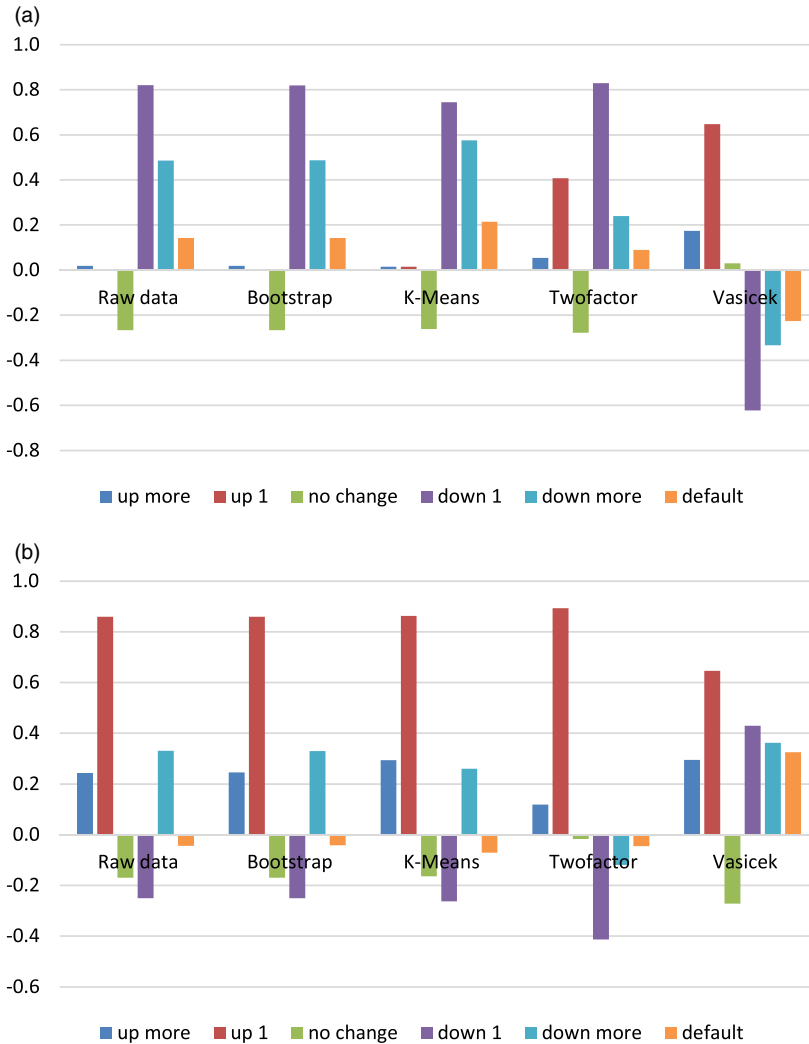


Figure 9. Plots of the eigenvectors of four models and raw data for (a) BBB for PC1 and (b) BBB for PC2.

- The bootstrap model has movements most like the raw data, which is in line with expectations as it is simply the raw data sampled with replacement.
- The K-means model eigenvalues are also like the raw data because again it is effectively a closely summarised version of the raw data grouping the raw data into eight groups and interpolating between them for intermediate percentiles.
- The two-factor and Vašíček models both have just two components. This is to be expected as these models summarise full transition matrices with 2 or 1 parameters respectively.
- Based on the eigenvalues, none of the models appears to be significantly different from the data to make them inappropriate for use.

The eigenvectors for each model are now compared in Figure 9 for BBB assets:

Figure 9 shows the first two eigenvectors for BBB-rated assets, which have a similar pattern to other ratings.

Table 8. The 1932 transition matrix.

From/to	AAA	AA	A	BBB	BB	B	CCC/C	D
AAA	68%	21%	10%	1%	0%	0%	0%	0.0%
AA	2%	53%	33%	7%	3%	0%	0%	0.7%
A	0%	1%	56%	32%	9%	1%	0%	0.9%
BBB	0%	0%	1%	53%	36%	8%	0%	0.9%
BB	0%	0%	0%	0%	50%	39%	4%	6.3%
B	0%	0%	0%	0%	1%	54%	29%	15.2%
CCC/C	0%	0%	0%	1%	1%	3%	68%	26.3%

For PC1 there is a clear similarity between the raw data, bootstrapping, and K-means. The PC1 direction for these assets is a fall in the assets staying unchanged and a rise in all other categories (with a small rise for upgrades). The two-factor model is similar in nature to these non-parametric approaches, albeit with a larger rise in the upgrades increasing by one rating. However, the Vašíček model is structurally different for PC1 in that, for PC1, the upgrades are moving in the same direction as the assets staying unchanged; and the opposite direction from the downgrades/defaults.

For PC2, the raw data, bootstrapping and K-means models are all very similar in nature. The two-factor model is also similar but notable with the down more than one rating category in the opposite direction to the non-parametric models. The Vasicek model is again structurally different from the other models with the “no change” group moving in the opposite direction to all the other categories. PC2 for the Vašíček is perhaps like PC1 for the other models in that the “no change” category is moving in the opposite direction to all other categories.

In this comparison, clearly, the non-parametric models are moving most closely in line with historical data. The Vašíček model is structurally different from the raw data, with PC2 of the Vasicek being more akin to PC1 for the other models. This suggests that Vašíček is not capturing the movement in historical data. The two-factor model is an improvement on the Vašíček in that it is a closer representation of the movement underlying raw data, which is what might reasonably be expected from the additional parameter.

6.5. Backtesting Comparison

A key requirement for transition and default models is that they meet any back-testing requirements. For example, the UK-specific requirement for Matching Adjustment Internal models is given in Supervisory Statement 8/18 point 4.3.4 as “compare their modelled 1-in-200 transition matrix and matrices at other extreme percentiles against key historical transition events, notably the 1930s Great Depression (and 1932 and 1933 experience in particular). This should include considering how the matrices themselves compare as well as relevant outputs”.

In this Section, the 99.5th percentile from the models is compared to the 1932 matrix. The 1932 matrix itself (Varotto, 2011), is shown in Table 8.

The four models being compared are:

1. Bootstrapping
2. The K-means model
3. The Vašíček model
4. The two-factor model

Table 9. The 99.5th transition matrix from the Vašíček model.

From/to	AAA	AA	A	BBB	BB	B	CCC/C	D
AAA	71%	25%	3%	0%	0%	0%	0%	0.0%
AA	0%	74%	22%	2%	0%	0%	0%	0.2%
A	0%	0%	81%	17%	1%	1%	0%	0.5%
BBB	0%	0%	0%	83%	12%	2%	1%	1.2%
BB	0%	0%	0%	1%	74%	20%	2%	3.7%
B	0%	0%	0%	0%	1%	73%	12%	14.2%
CCC/C	0%	0%	0%	0%	0%	4%	36%	60.6%

Table 10. The 99.5th transition matrix with strengthened Vašíček calibration.

From/to	AAA	AA	A	BBB	BB	B	CCC/C	D
AAA	43%	45%	7%	1%	1%	1%	1%	0.0%
AA	0%	47%	43%	7%	1%	1%	0%	0.6%
A	0%	0%	57%	34%	4%	2%	0%	1.5%
BBB	0%	0%	0%	63%	26%	6%	2%	3.7%
BB	0%	0%	0%	0%	48%	36%	6%	10.8%
B	0%	0%	0%	0%	0%	47%	19%	33.9%
CCC/C	0%	0%	0%	0%	0%	0%	13%	86.9%

The bootstrapping model is simply using the raw data, sampled with replacement. Thus, the most extreme percentile is simply the worst data point; in this case the 1932 matrix. On this basis, it might be concluded that the bootstrapping model passes the backtest, however, it also has no scenarios worse than the 1932 matrix. This means that scenarios worse than the worst event in history are not possible to model, which is a significant model weakness.

The K-means model produced as part of this paper had the 99.5th percentile specifically set to be the 1932 matrix. On this basis, the model passes the backtest by construction. The model is flexible enough so that the percentiles of the various K-means clusters are selected by the user for the specific purpose required. The K-means model also has transition matrices stronger than the 1932 event with the 100th percentile set at the 1932 matrix multiplied by itself (effectively two such events happening in a single year).

The Vašíček model calibrated to the data set described in Section 1 gives a rho value of just over 8%. The 99.5th percentile from this model is shown in Table 9:

Compared to the 1932 matrix, it can be clearly seen that:

- The defaults are lower for most ratings.
- The transitions one rating lower for AA, A and BBB assets are lower.
- The leading diagonal values are higher.

This shows this matrix is not as strong as the 1932 matrix. However, it would be possible to strengthen the calibration of the Vašíček model, moving the rho parameter to say 30% as an expert judgement loading – specifically targeted at passing the backtesting requirements. The updated 99.5th percentile for this strengthened Vašíček model is shown in Table 10.

Table 11. The two-factor 99.5th percentile transition matrix.

From/to	AAA	AA	A	BBB	BB	B	CCC/C	D
AAA	66%	31%	2%	0%	0%	0%	0%	0.0%
AA	2%	67%	29%	2%	0%	0%	0%	0.1%
A	0%	2%	68%	28%	2%	1%	0%	0.3%
BBB	0%	0%	2%	67%	26%	3%	1%	1.2%
BB	0%	0%	0%	2%	63%	30%	2%	2.9%
B	0%	0%	0%	0%	2%	63%	20%	15.5%
CCC/C	0%	0%	0%	0%	0%	3%	38%	58.8%

Compared to the 1932 matrix, it can be clearly seen that:

- The defaults are now largely higher than the 1932 matrix.
- The transitions one rating lower for AA, A and BBB assets are higher or comparable.
- The leading diagonal values are lower or more comparable.

The two-factor model has a range of 99.5th percentiles that could be used, depending on the portfolio of assets it is applied to; but for the purposes of this paper, the 99.5th percentile has been taken as the 99.5th percentile of the sum of investment grade default rates. Using this approach, the below 99.5th percentile model has been produced:

Compared to the 1932 matrix, it can be seen that:

- The defaults are higher only for BBB investment grade assets; with the 1932 matrix higher for other investment grades.
- The transitions one rating lower for AA, A and BBB assets are more comparable to the 1932 matrix than the unadjusted Vašíček calibration, but slightly lower than the 1932 matrix.
- The leading diagonal values are slightly higher than the 1932 matrix.

Overall the two-factor transition matrix at the 99.5th percentile is slightly weaker than the 1932 matrix; but stronger than the Vašíček 99.5th percentile. This model could also be strengthened in a similar way to the strengthening of the Vašíček model – with an expert judgement uplift to one of the parameters. There are a few more options where just an adjustment might be applied:

1. In the probability distributions used to model Inertia and Optimism. These could be fatter-tailed distributions than those chosen in the calibration in this paper.
2. The copula used to model Inertia and Optimism could be made a t-copula rather than a Gaussian copula. This is potentially more appropriate as in practice the 1932 matrix has the most extreme values for both variables, indicating a tail dependence perhaps more in line with the t-copula than the Gaussian copula.
3. Specific adjustments could be made to the parameters of the calibrated risk distributions.

6.6. Objectivity Comparison

Each of the models has varying levels of expert judgement applied, and this Section compares each.

Table 12. Summary of model comparisons.

Comparison	Bootstrapping	K-means model	Vašíček	Two-factor
Principle component	Near identical replication of underlying data movements	Very close replication of underlying data movements	Poor replication of underlying data movements	Good replication of underlying data movements
1932 Backtest	Would pass a backtest if backtest level is in the historical data; but cannot produce stress worse than anything in the data	Pass by construction	Requires significant additional expert judgement strengthening to pass	Limited expert judgement strengthening to pass
Objectivity	Objective – no expert judgement	Heavy expert judgement in distribution construction	Expert judgement to strengthen to pass backtest	Expert judgement in choice of distributions and copula
Simplicity	Simple	Complex	Complex	Complex
Breadth of uses	Less appropriate for extreme percentiles as cannot produce scenarios worse than worst historical event.	Highly flexible and can be set to the required use with appropriate judgements	Is used widely; but does not capture historical movements in the data well	Flexible model for a range of uses; additional parameters to Vašíček allows better replication of historical data movements

6.6.1. Bootstrapping

Bootstrapping involves sampling many data points from the historical data with replacement. This model is the simplest method to apply without any inclusion of expert judgement. The expert judgement is now exclusively in the selection of the data set.

6.6.2. K-means model

The K-means model involves the most amount of expert judgement of the four models compared. This is mainly due to the choice of percentiles for each of the k-means groups being selected by expert judgement. The model could easily be updated to a different choice of percentiles for each k-means group which would give different results and could be used for a different purpose. There are other judgements required including how many groups to use.

6.6.3. Two-factor model

The main expert judgements are around the distribution used to model the two factors and the copula used to model their interaction.

6.6.4. Vašíček model

The assumptions behind the Vašíček model include several expert judgements around the way the risk behaves. To pass the 1932 backtest an expert judgement overlay is required to strengthen the calibration.

6.6.5. Summary of model comparisons

The comparisons between the models are summarised in Table 12:

The models explored all have strengths and weaknesses and all could be used for a variety of purposes. The Vašíček model is used extensively in the insurance industry, but it does not replicate historical movements in the data as well as the K-means model or the two-factor model.

7. Conclusions

Transition and default modelling is one of the most complex risks to be modelled by insurance companies. The use of transition matrices creates a large modelling challenge due to the large number of data items contained in each matrix and how these interact with each other.

This paper has reviewed four models for assessing this risk, two of which (the K-means and two-factor model) are not previously captured in the literature. The four models have been compared with several metrics, including a new test using PCA to compare model movements to historical data movements for transition matrix models. The PCA-based test has highlighted a deficiency in the Vašíček model in that it does not replicate the way historical data moves. The first principal component of the Vašíček model is not well matched to the first principal component of the underlying data, and the second principal component of the Vašíček model is perhaps closer to the first principal component in the underlying data. The other three models shown in this paper capture the historical movement of the underlying data more accurately than the Vašíček model.

The non-parametric models have the advantage of having historical movements very close to the historical data, but the bootstrapping approach has a limitation of not producing stresses worse than the worst historical data point. The K-means model in the form presented in this paper has a relatively significant amount of expert judgement in its construction. The two-factor model has the advantage of being relatively simple to apply, with an improved representation of the historical data (relative to the Vašíček model); but not as close to the historical data as the non-parametric models.

Acknowledgements. The authors are grateful to helpful and insightful discussions from Mr Alan Reed FFA, on the use of principal components analysis with respect to transition matrices and for introducing them to the K-means clustering model described later in the paper.

Disclaimer. The views expressed in this publication are those of invited contributors and not necessarily those of the Institute and Faculty of Actuaries. The Institute and Faculty of Actuaries do not endorse any of the views stated, nor any claims or representations made in this publication and accept no responsibility or liability to any person for loss or damage suffered as a consequence of their placing reliance upon any view, claim or representation made in this publication. The information and expressions of opinion contained in this publication are not intended to be a comprehensive study, nor to provide actuarial advice or advice of any nature and should not be treated as a substitute for specific advice concerning individual situations. On no account may any part of this publication be reproduced without the written permission of the Institute and Faculty of Actuaries.

References

- Belkin, B., Suchower, S., & Forest, Jr., L.** (1998). The effect of systematic credit risk on loan portfolio value-at-risk and loan pricing. *Credit Metrics Monitor*, 1st quarter, 17–28.
- Herfindahl, O.** (1950). Concentration in the U.S. Steel Industry. Dissertation, Columbia University.
- Merton, R.C.** (1974). On the pricing of corporate debt: the risk structure of interest rates. *The Journal of Finance*, **29**(2), 449–470. Papers and Proceedings of the Thirty-Second Annual Meeting of the American Finance Association, New York, December 28–30, 1973. Wiley.
- Moody's.** (2006). *Measuring Corporate Default Rates*, available at <https://www.moody.com/sites/products/DefaultResearch/2006200000425249.pdf>
- Rosch, D., & Scheule, H.** (2008) *Stress Testing for Financial Institutions, Applications Regulations and Techniques*. Risk Books - a division of Incisive Financial Publishing Ltd.
- S&P.** (2016). *2015 Annual Global Corporate Default Study and Rating Transitions*. New York: S&P.
- Trevino, A.** (2016). Introduction to K-means Clustering; Oracle AI & Science Data blog, available at <https://blogs.oracle.com/ai-and-datascience/post/introduction-to-k-means-clustering#:~:text=K%2Dmeans%20clustering%20is%20a,represented%20by%20the%20variable%20K>
- Varotto, S.** (2011). Stress testing credit risk: The great depression scenario. *Journal of Banking & Finance*, **36**(12), 3133–3149.
- Vašíček, O. A.** (1987). *Probability of Loss on Loan Portfolio*. San Francisco: KMV Corporation. *Journal of Financial Risk Management*, 5(4).
- Vašíček, O. A.** (2002). The distribution of loan portfolio value. *Risk*, **15**(12), 160–162.

Appendix A. Details of K-means model

Technical Details on Grouping

1. We have used the standard K-means clustering algorithm (an unsupervised learning algorithm) to perform the grouping.
2. The K-means clustering algorithm tries to classify a data set through a number of clusters (assumed K clusters) fixed a priori within an unlabelled multidimensional data set. It accomplishes this using a conception of what the optimal clustering looks like:
 - a. The “cluster centre” is the arithmetic mean of all the points belonging to the cluster.
 - b. Each point is closer to its own cluster centre than to other cluster centres.
3. The objective function also known as squared error function is given by:

$$J(V) = \sum_{i=1}^c \sum_{j=1}^{c_i} (||x_i - v_j||)^2$$

where $||x_i - v_j||$ is a Euclidian distance between x_i and v_j , c_i is the number of data points in i^{th} cluster and c is the number of cluster centres.

Algorithm steps

Let $X = \{x_1, x_2, \dots, x_n\}$ be the set of data points (in our case transitions for the key ratings with weights) and $V = \{v_1, v_2, \dots, v_c\}$ be the set of centers.

1. Randomly select ‘c’ cluster centres.
2. Calculate the distance between each data point and each of the cluster centres.
3. Assign the data point to the cluster centre whose distance from the cluster centre is the minimum distance versus all the available cluster centres.
4. Recalculate the new cluster centre using:

$$v_i = \left(\frac{1}{c_i}\right) * \sum_{j=1}^{c_i} x_j$$

where c_i is the number of data points in the i^{th} cluster.

5. Recalculate the distance between each data point and the newly obtained cluster centres.
6. If no data point was reassigned then stop, otherwise repeat from step 3 above.

Appendix B. Details of two-factor model

This appendix gives an example for how the two-factor model Inertia and Optimism values are calculated for a specific matrix.

The Inertia is calculated as the sum of the diagonal values (other than for the Default category). These are coloured yellow in Table B1.

Table B1. Highlighted example of inertia.

From/to	AAA	AA	A	BBB	BB	B	CCC/C	D
AAA	89.82%	9.42%	0.55%	0.05%	0.08%	0.03%	0.05%	0.00%
AA	0.52%	90.63%	8.17%	0.51%	0.05%	0.06%	0.02%	0.02%
A	0.03%	1.77%	92.30%	5.40%	0.30%	0.13%	0.02%	0.06%
BBB	0.01%	0.10%	3.64%	91.63%	3.86%	0.49%	0.12%	0.18%
BB	0.01%	0.03%	0.12%	5.35%	85.80%	7.36%	0.61%	0.72%
B	0.00%	0.02%	0.09%	0.19%	5.63%	85.09%	5.05%	3.93%
CCC	0.00%	0.00%	0.13%	0.24%	0.70%	15.63%	51.49%	31.82%
D	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100%

For this example,

$$\text{Inertia} = 89.82\% + 90.63\% + 92.3\% + 91.63\% + 85.8\% + 85.09\% + 51.49\% = 5.87$$

The Optimism is calculated as the default weighted sum of upgrades divided by the sum of downgrades and defaults. The defaults used in the weighting are shown in yellow. The upgrades are in blue, the downgrades are in green.

Table B2. Highlighted example of optimism.

From/to	AAA	AA	A	BBB	BB	B	CCC/C	D
AAA	89.82%	9.42%	0.55%	0.05%	0.08%	0.03%	0.05%	0.00%
AA	0.52%	90.63%	8.17%	0.51%	0.05%	0.06%	0.02%	0.02%
A	0.03%	1.77%	92.30%	5.40%	0.30%	0.13%	0.02%	0.06%
BBB	0.01%	0.10%	3.64%	91.63%	3.86%	0.49%	0.12%	0.18%
BB	0.01%	0.03%	0.12%	5.35%	85.80%	7.36%	0.61%	0.72%
B	0.00%	0.02%	0.09%	0.19%	5.63%	85.09%	5.05%	3.93%
CCC	0.00%	0.00%	0.13%	0.24%	0.70%	15.63%	51.49%	31.82%
D	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100%

$$\begin{aligned} \text{Optimism} = & (0.52\% / (8.17\% + 0.51\% + 0.51\% + 0.06\% + 0.02\% + 0.02\%)) * 0.02\% \\ & + (0.03\% + 1.77\%) / (5.4\% + 0.3\% + 0.13\% + 0.02\% + 0.06\%) * 0.06\% \\ & + (0.01\% + 0.1\% + 3.64\%) / (3.86\% + 0.49\% + 0.12\% + 0.18\%) * 0.18\% \\ & + (0.01\% + 0.03\% + 0.12\% + 5.35\%) / (7.36\% + 0.61\% + 0.72\%) * 0.72\% \\ & + (0.00\% + 0.02\% + 0.09\% + 0.19\% + 5.63\%) / (5.05\% + 3.93\%) \\ & * 3.93\% / (0.02\% + 0.06\% + 0.18\% + 0.72\% + 3.93\%) = 0.655 \end{aligned}$$

Note, in this version of this model, Optimism has only been calculated on AA-B ratings; but could also be calculated including CCC ratings or just investment grade ratings.