Institute and Faculty of Actuaries

Calibration of Transition Risk for Corporate Bonds

by J Sharpe*, F Ginghina, G Mehta, A D Smith

Presented to the Institute & Faculty of Actuaries

15 May 2023
Staple Inn Hall, London

Contents

1 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. There is a discussion of the available data sources and some of the key features of the data. 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 historic data set with replacement
- The Vasicek model calibrated using the Belkin approach
- The K-means model – a new non-parametric model produced using the K-mean clustering algorithm
- A Two Parameter model – a new parametric model, using two parameters (instead of a single parameter with the Vasicek) to represent each matrix

The models are compared in a number of ways:

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

2 Acknowledgements

The authors are grateful to helpful and insightful discussions from Mr Alan Reed FFA, on the use of Principal Component Analysis with respect to Transition matrices and for introducing them to the K-mean cluster model described later in the paper.

3 Keywords

Credit risk; Transition and Default risk; Vasicek credit model; Solvency II; K-means

4 Correspondence details

*Correspondence to: James Sharpe email: james@sharpeactuarial.co.uk

5 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 models and considerations.

- In section 4 we outline the key risk drivers associated with this risk and introduce the core modelling component – the transition matrix
- In section 5 we look at the primary sources of historic transition matrix data with a discussion of how this data is treated.
- Section 6 has an analysis of the data presenting the key features any credit model should aim to capture.
- Section 7 discusses the different types of credit model and then presents a range of models split between parametric and non-parametric model types. The parametric models explored are the Vasicek model (Vasicek 1987) calibrated using an approach described in Belkin (1998); and a new two parameter 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 historic matrices; as well as a simple bootstrapping approach of simulating historic transition matrices with replacement.
- Section 8 includes a quantitative and qualitative comparison of the various credit models. For the quantitative comparison, Principal Component Analysis (PCA) is used to identify the directions of most variance in historic data, which is then compared for each of the models. 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 being easier to calibrate and explain to stakeholders, but at the cost of not explaining as much 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 to the historic 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 sufficient to meeting 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 Vasicek model does not move consistently with the raw data; PC1 from the raw data is more consistent with PC2 from the Vasicek; and PC1 from the Vasicek 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 Vasicek and Two Parameter model require additional strengthening to ensure their 99.5th percentile exceeds the 1932 transition matrix. 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 (i.e. an example of the Lucretius fallacy). The K-means model is specified to pass the 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 the model development but is also less objective. The bootstrapping model has no requirement for expert judgement at all beyond the choice of data. The Vasicek and Two Parameter models can be applied with varying amounts of expert judgement depending on the purpose the model is designed for.

6 Risk Driver Definition

Transition and default risk applies 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 historic time series of credit spreads are largely split by credit ratings, so it is difficult to avoid the use of credit rating and so transition and default risk 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 to 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 give the one-year transition and default probabilities based on averages over 1981-2018.

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%

Table 1 – S&P average transitions from 1981-2018

The transition matrix itself is the data item that is being modelled. An historic time series of transition matrices can be obtained and this time series of perhaps 30 or 100 transitions 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 simple 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.

7 Data Sources

For historic 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 8.

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 2: Comparison of Transition Risk Data Sources Considered

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. An example matrix is shown below.

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

Table 3: S&P One-Year Corporate Transition Rates by Region (2018)

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(NR)).

Some key points to note about transition matrices are:

1. Each row sums up to 1, as this represents the total probability for 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 which have remained at the same credit rating over the year

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

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

8 Stylised Facts of Data

Key Messages:

- The data set used is a series of transition matrices which varies 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 to 2019 S&P transition matrix data.

The plot immediately below shows the 1932 values compared to the 1981-2019 data. The 1931-35 transition matrix has been used in the model calibrations in this paper, but is not shown in the plot below.

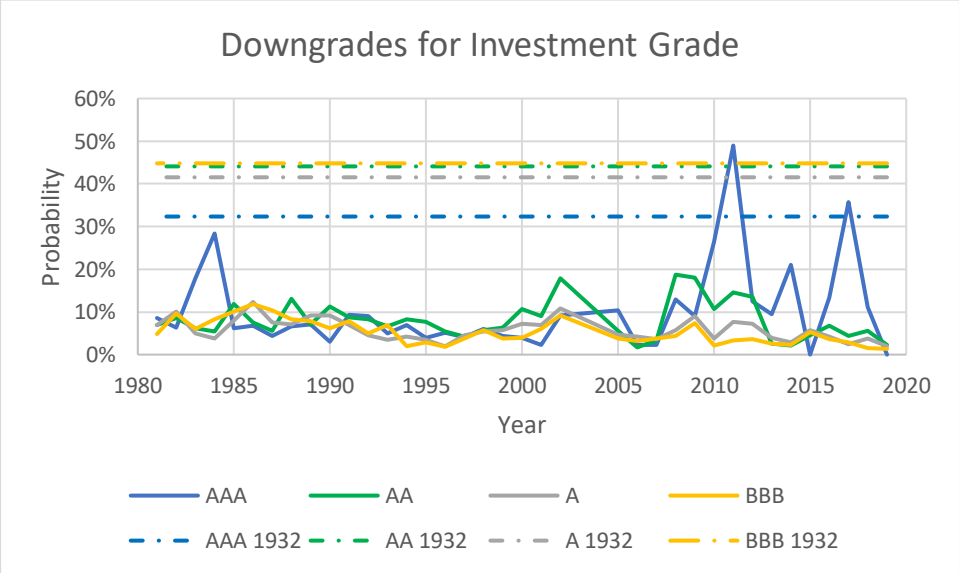


Figure 1 – downgrade rates for investment grade assets

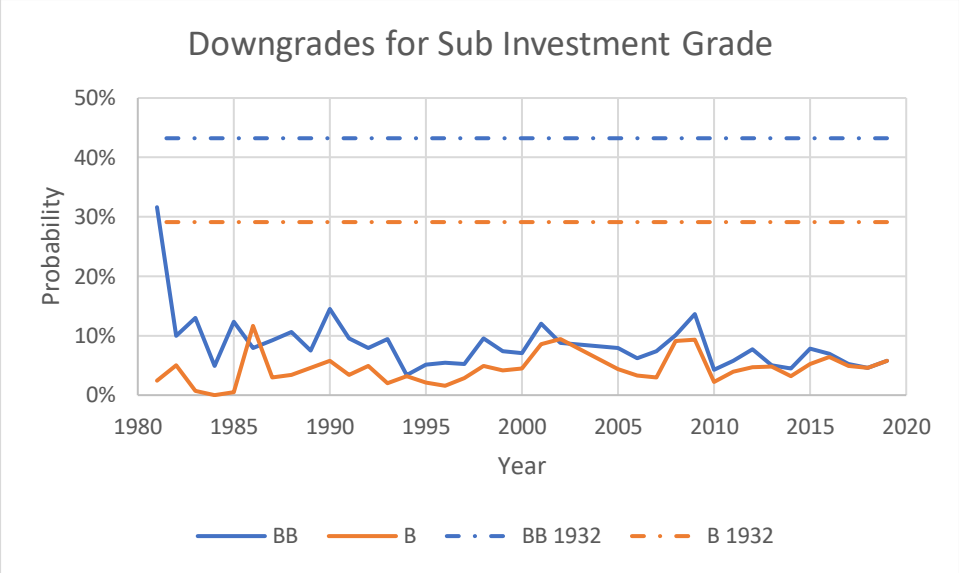


Figure 2 – downgrade rates for sub-investment grade assets

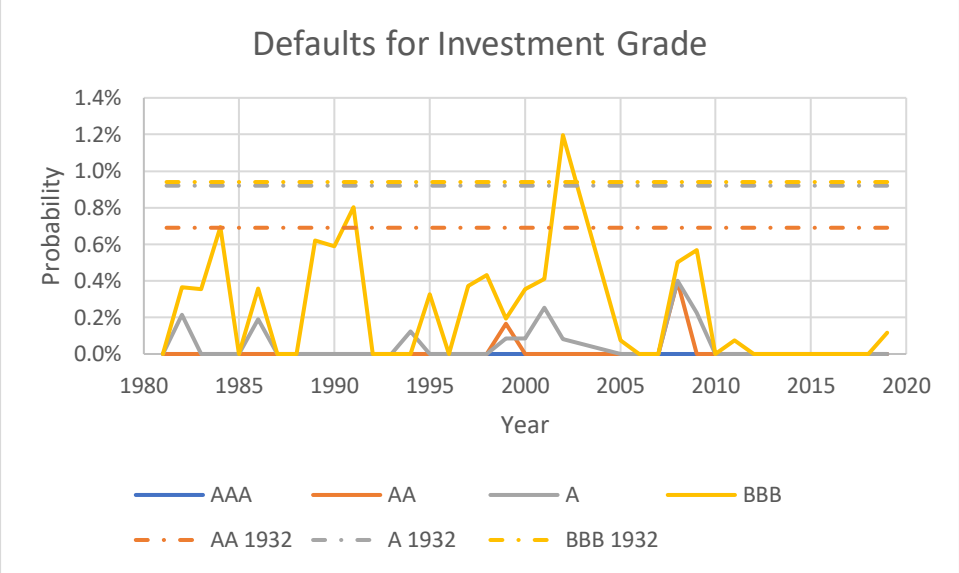


Figure 3 – default rates for investment grade assets

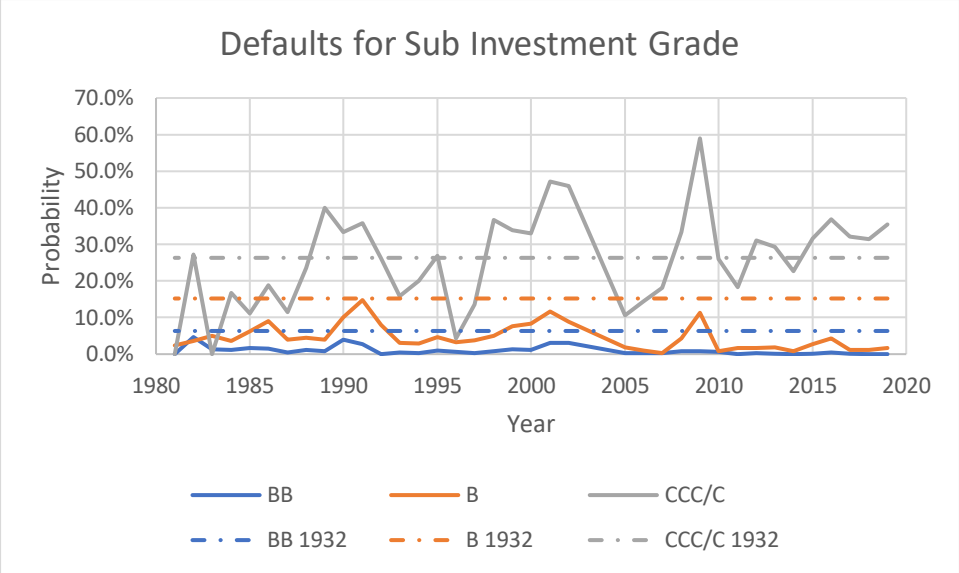


Figure 4 – default rates for sub-investment grade assets

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 much more material as the sub-investment grade ratings.
- 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. It is clear that the 1932 matrix was worse than and in the more recent period 1981-2019
    - 2009 and 2001 shown relatively high levels of default and downgrade which is expected given the financial crisis and dot.com bubble respectively.

The tables below show the mean, standard deviation, skewness and excess kurtosis for the upgrades, downgrades and defaults based on data from 1981-2019 and including the 1932 transition matrix.

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

+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	

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

Table 4 – first four moments for downgrades, upgrades and defaults

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 and 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 Parameter model later in the paper with the “Optimism” parameter.

9 Models Explored

In the next sections four credit models are described in detail which we have split between parametric and non-parametric models.

9.1 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:

- Vasicek (calibrated using Belkin approach)
- Two parameter model (a model introduced in this paper)

9.1.1 The Vasicek Model

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

Vasicek required several assumptions for the portfolio of assets:

- All asset returns are described by a Wiener process. In other words, all asset values are Lognormally 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*.

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

$$lnA(T) = lnA + \mu T - \frac{1}{2}\sigma^2 T + \sigma\sqrt{T}X \qquad [1]$$

where *T* is the maturity of the asset, *W(t)* is a standard Brownian motion, asset values (denoted *A(t)*) are Lognormally distributed, μ and σ^2 are the instantaneous expected rate and instantaneous variance of asset return 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 Vasicek's model was to adapt Merton's single asset model to a portfolio of assets. For a firm denoted *i* (with *i* = 1, …, n), equation [1] can be re-written as:

$$lnA_i(T) = lnA_i + \mu_i T - \frac{1}{2}\sigma_i{}^2 T + \sigma_i\sqrt{T}X_i \quad [2]$$

Given the assumptions above, and in particular the equi-correlation assumption of variables X_i, it follows that the variables X_i belong to an equi-correlated Standard Normal distribution. (Note equi-correlation means all assets in the portfolio are assumed to have the same correlation between each other). Any variable X_i that belongs to an equi-correlated Standard Normal distribution can be represented as linear combination of jointly Standard Normal random variables Z and Y_i such that:

$$X_i = Z\sqrt{\rho} + Y_i\sqrt{1-\rho}, \text{ where i =1,…,n.} \quad [3]$$

Equation [3] above is a direct result of statistical properties of jointly equi-correlated Standard Normal variables, which stipulates that any two variable X_i and X_j are bivariate Standard Normal with

corelation coefficient ρ if there are two independent Standard Normal variables Z and Y for which $X_{\mathrm{i}} = Z$ and $X_{\mathrm{j}} = \rho Z + \sqrt{1-\rho^2}Y$, 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 to 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 asset, whilst Y_i are ith firm's specific, and independent from variable *Z* and variables Y_j, where *j* <> *i*. Next, Vasicek denotes variable *Z* as a portfolio common factor (for instance, some measure of the state of the economy), and variable Y_i as each firm's specific risk.

As a parenthesis, the covariance between two firm's asset returns X_{i} and X_{j} is determined by $\rho_{ij} = \frac{cov(X_i,X_j)}{\sigma(X_i)\,\sigma(X_j)}$, 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 $\rho_{ij} = cov(X_i, X_j)$.

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

$$P(firm\ X_i\ defaults \mid Z) = \Phi\left(\frac{x_i - Z\sqrt{\rho}}{\sqrt{1-\rho}}\right), \quad [4]$$

Finally, although Vasicek has not considered credit ratings in his setting other than the default state, from Equation [4] above 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 (for example, 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 of 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.

9.1.1.1 Vasicek Model calibration - Belkin

Belkin (1998) introduced a statistical method to estimate the correlation parameter ρ and common factors Z in Vasicek'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 years):

$$X_i = Z\sqrt{\rho} + Y_i\sqrt{1-\rho}, \text{ (with i denoting year)} \quad [X]$$

where Z, Y_i and ρ are as per Vasicek'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.

[2] See 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 parameter model which is introduced in the next section.

The Belkin implementation uses the Standard Normal distribution. If this distribution were replaced with fatter tailed distributions, 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 8.

9.1.2 The Two Parameter Model

9.1.2.1 Two Parameter Model description

The Vasicek 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 parameter model is described to capture two important features about the way transition matrices change over time and particularly in stress. This model is based on a description given in Rosch (2008).

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

Inertia is defined as the sum of the leading diagonal of the transition matrix. For a transition matrix with probabilities p_{ij} where i is the row and j the 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.

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

Where $p_{iD}$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 a each row. $\sum_{i>j} p_{ij}$ is the sum of all downgrades in each row. Appendix B gives an example calculation of Inertia and Optimism for a give matrix.

Any historic matrix can be characterized by these two parameters 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 historic transition matrix. This allows the generation of an historic time series for Inertia and Optimism based on our historic data set; then to fit probability distributions to the values of these parameters which 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 historic 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 an historic matrix or parameters output from a probability distribution by using the following steps (in these 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 ith row and jth 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:

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

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:

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

The final matrix can now be found by:

$$if\ j > i\ (i.e. downgrades)\ then\ p_{ij}^{(4)} = \frac{p_{ij}^{(3)}}{\sum_{j=i+1}^{7} p_{ij}^{(3)}} * \left(\sum_{j=i+1}^{7} p_{ij}^{(3)} - Factor\right)$$

$$if\ j < i\ (i.e. upgrades)\ then\ p_{ij}^{(4)} = \frac{p_{ij}^{(3)}}{\sum_{j=1}^{i-1} p_{ij}^{(3)}} * (\sum_{j=1}^{i-1} p_{ij}^{(3)} - Factor)$$

9.1.2.2 Two Parameter Model Calibration

This section describes an approach to calibrating the two parameter model.

Using historic data for historic transition matrices from 1981-2019 as well as the 1932 matrix and a matrix from 1931-35 a time series of Inertia and Optimism can be constructed. It is then possible to fit probability distributions from this data and then simulate from these probability 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 below with the time series compared to the values for 1932 and 1931-35.

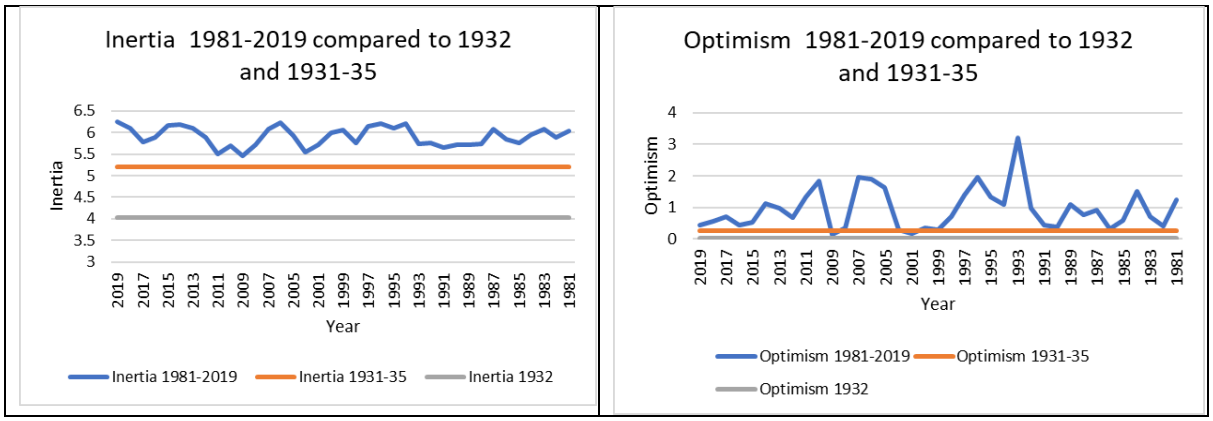


Figure 5 – Inertia and Optimism, historic values compared to 1932 and 1931-35

The first four moments for this data are shown below (including the 1932 and 1931-35 matrices).

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 – first four moments of Inertia and Optimism

The correlations between the two data series are:

Correlation	Value
Pearson	0.33
Spearman	0.43
Kendall Tau	0.30

Table 7 – correlation between Inertia and Optimism

The main comments on these data series are:

- Inertia is negatively skewed and very 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 most amount of downgrades relative to upgrades in that year.

In order to simulate from these data sets probability distributions have been fit to the two data series using the Pearson family of probability distributions. The Pearson Type 1 distribution produces a

satisfactory fits to the two data sets as shown below comparing the raw data ("data") against 10,000 simulated values from the fitted distributions ("fitted distribution").

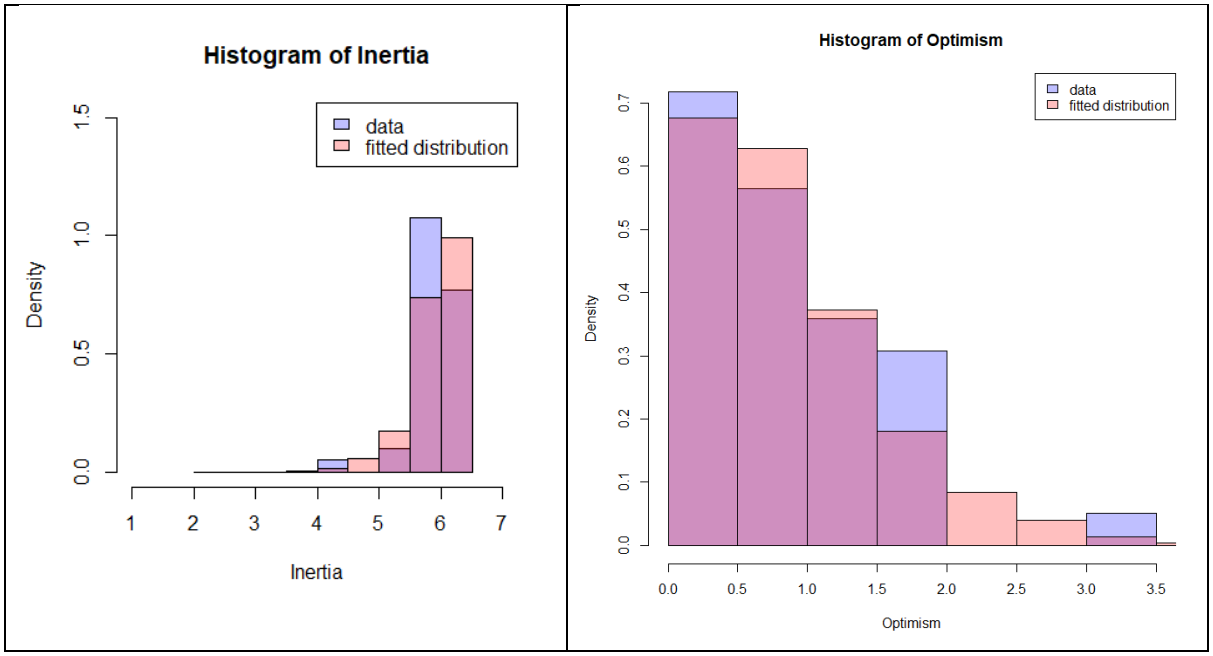


Figure 6 – historic plots of Inertia and Optimism compared to fitted distributions

The plot above shows the actual historic data values for Inertia and Optimism from each of the historic 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 each other. For this purpose a Gaussian Copula has been picked 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 parameter models are 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.

9.2 Non-Parametric Models

While the Vasicek and Two Parameter 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 Vasicek 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 (known as the K-means model); the second is a bootstrapping approach, whereby historic transitions matrices are simulated a large number of times with replacement, to get a full risk distribution.

9.2.1 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. We then assign each of the groups to real-line percentiles manually, e.g. assign a group containing the 1932 matrix at 0.5th percentile, assign the average matrix of the 1931-35 matrix at 0.025 percentile, put an identity matrix at the 100th percentile, put a square of 1932 matrix at 0th percentile, etc.

3. Interpolate any percentiles in between we need using a matrix interpolation approach.

9.2.1.1 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 groups. The goal of this algorithm is to find groups in the data, with the 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. Data points are clustered based on feature similarity. The results of the K-means clustering algorithm are:

1. The centroids 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 the 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-mean algorithm to group each of the transition year data into groups with a similar profile of transitions from one rating to the other. 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 the 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.

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

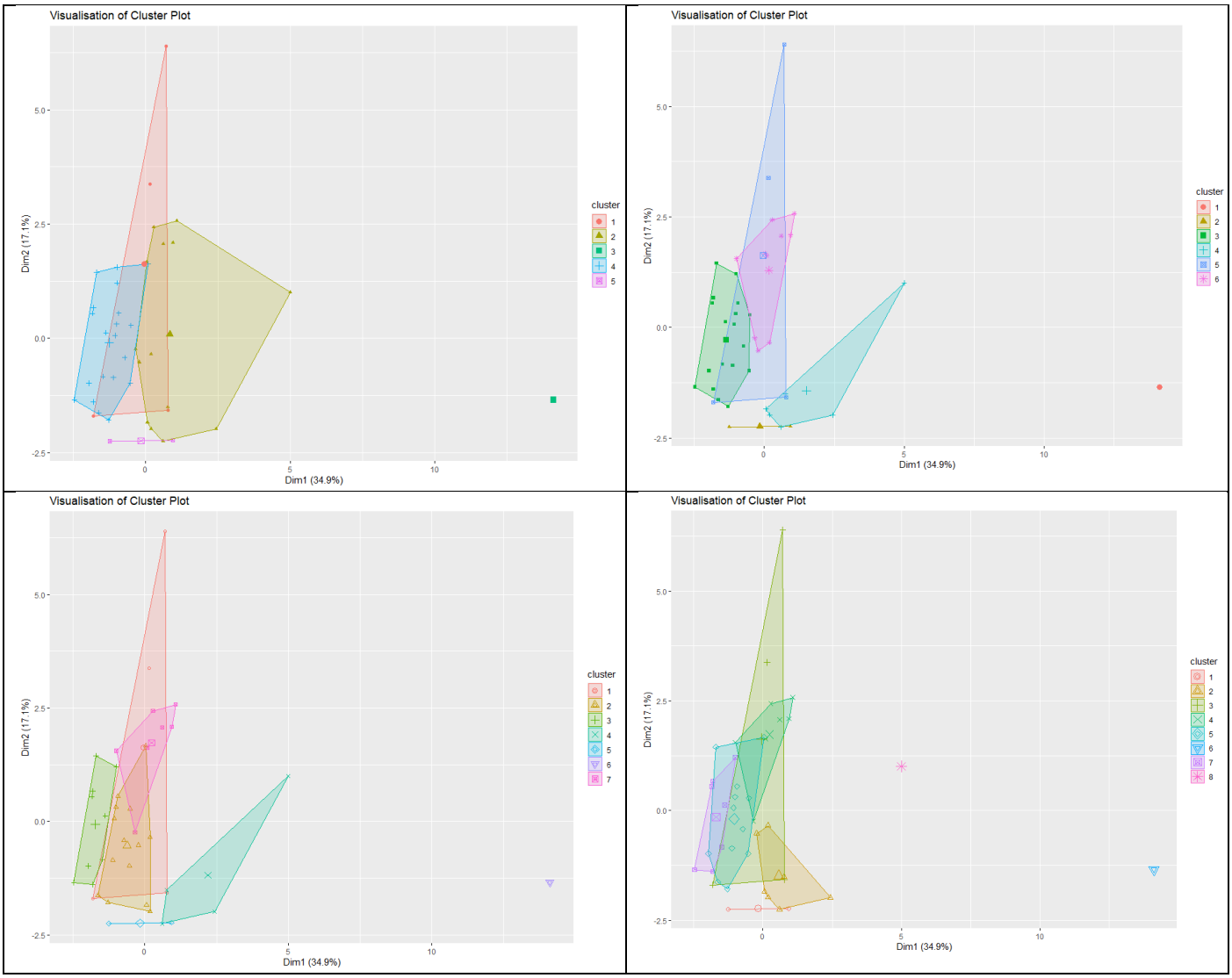


Figure 7: K-Means Clustering – Examples where K=5, 6, 7 and 8

As shown above

- 1931 and 1931-35 average matrices are in separate groups as we increase the groups from 7 onwards
- Below K=7 groups, the average 1931-35 average matrix is not coming out as a separate group in its own right.

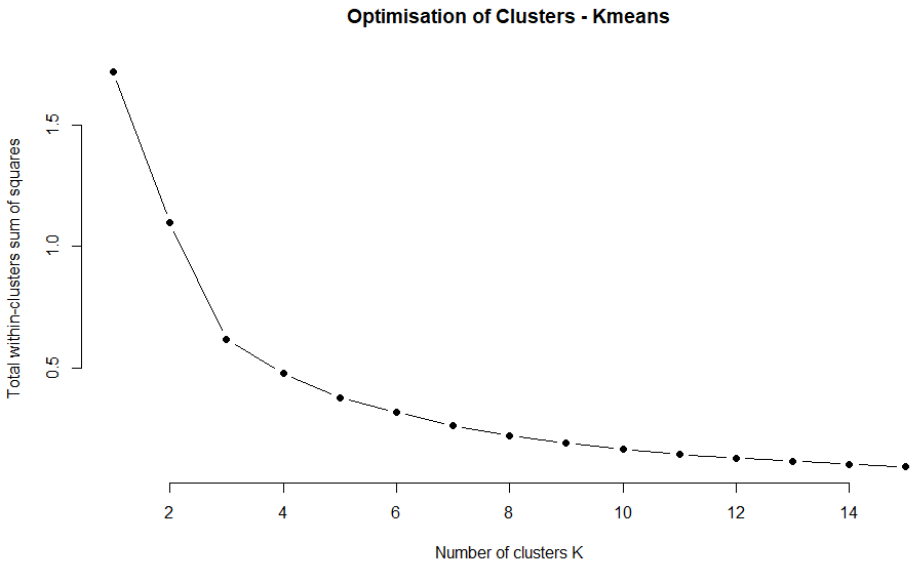


Figure 8 : K-Means Clustering – Examples with different K Values – Sum of Squares within and between the Groups

- As shown in **Error! Reference source not found.**, 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 .

Note: Please note we have applied equal weights to each of the transition matrices.

9.2.1.2 Manually Assign the matrix group to Percentiles

Once K-means steps is completed, we get each of the transition matrix assigned to a group as shown in *Figure 7*. For our example, we have selected K=8. The next step is manually to assign each of the groups to a percentile value on the empirical distribution. This is set via an expert judgment process and we have set these as follows:

- Square of 1932 matrix is assigned to 0th percentile;
- 1932 matrix is assigned to 0.5th percentile;
- 1932-1935 average matrix is assigned to 1.25th percentile;
- A Group containing 2002 matrix to 8.75th percentile
- A Group containing 2009 matrix to 27.5th percentile
- A Group containing 1981 matrix to 48.75th percentile
- A Group containing 2016 matrix to 66.25th percentile
- A Group containing 2017 matrix to 81.25th percentile
- A Group containing the 2019 matrix to 93.75th percentile.
- A identity matrix is assigned to 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.

$$Interp = \left(\frac{p_2 - q}{p2 - p_1}\right)$$

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

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

$$Q = E_1 * E_2$$

9.2.2 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 a large number of 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.

10 Comparison of the Models Explored

In this section several metrics are used to compare the models described in the previous section:

1. Do the models show movements of a similar nature to historic data capturing the types of stresses seen historically
2. Do the models produce extreme percentiles sufficiently extreme to pass backtesting
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 historic transition matrices. In particular how do the model transition matrices move over time compared to how the historic transition matrices move.

Sections 8.1, 8.2 and 8.3 describe how the movements of transition matrices generated by the models can be compared to historic data. Section 8.4 show how the various models described in section 7 compare in this metric. Section 8.5 shows how the models compare to historic backtests expected under UK regulatory frameworks. Section 8.6 compares the models in terms of the amount of expert judgement required to calibrate them.

10.1 Dimension Reduction and Visualisation

10.1.1 The need to reduce dimensions.

Transition matrix modelling is a high-dimensional exercise. It is almost impossible to visualise the 63-dimensional distribution of a random 7×9 matrix. To make any progress comparing models we need to reduce the number of dimensions while endeavouring to mitigate the loss of information.

The dimension reduction is all the more necessary because commonly used transition models employ a low number of risk drivers, in order to contain calibration effort, particularly the need to develop correlation assumptions with other risks within an internal model. Vasiček'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 Vasiček's model is a one-dimensional manifold (ie a curve) in 56-dimensional space. This is a dramatic dimension reduction relative to the historical data.

We can simplify matters to some extent 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 63 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.

10.1.2 Principal Components Analysis.

Principal component analysis (PCA) 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.

10.2 Granulation

10.2.1 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 Vasiček model 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.

10.2.2 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 Vasiček'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.

10.2.3 Mathematical Definition of Granulation

For $n \geq 2$, Let S_n denote the n-simplex, that is the set of ordered $(n+1)$-tuples $(x_0, x_1, x_2 \ldots x_n)$ 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 a vector Y satisfies $Y \sim \mathbb{P}_x$ then

$$\mathbb{E}(Y_i) = x_i$$

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

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

10.2.4 Granulation Examples.

One familiar example of a granulation is a multinomial distribution with n bonds and probabilities x, in which caser the Herfindahl index is n^{-1}. In the extreme case n=1, this a categorical distribution where all the 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 a given 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

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

10.2.5 Granulation Effect on Means and Variances of Transition Rates

We now investigate the effect of granulation on means and variance matrix 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

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

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

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

And the variance(-covariance) matrix of Y is

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

10.3 Weighted Principal Components

10.3.1 Weighting Proposal

We are now in a position 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 $\mathbf{V}$. Our proposed weighted PCA method is based on a singular value decomposition of the matrix

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

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 is reduces default rates (and a negative quantity increases defaults). This is consistent with our definitions of optimism and inertia in section ???

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

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

10.3.2 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

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

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

10.3.3 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 re-combined 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 hypercuboid 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.

10.4 Weighted PCA Applied to Our Fitted Models

In this section we present the eigen vectors and eigen values from the PCA approach applied to the four models and the raw data. To apply the PCA to the models, 10,000 simulated transition matrices were generated for each model and the PCA applied to this data.

The eigen vectors are shown for the single A and BBB rated movements.

Single A	Eigen values				
	PC1	PC2	PC3	PC4	PC5
Raw data	0.075	0.008	0.003	0.002	0.001
Bootstrap	0.073	0.007	0.003	0.002	0.001
K-means	0.049	0.003	0.001	0.001	0.000
Twofactor	0.056	0.012	0.000	0.000	0.000
Vasicek	0.029	0.004	0.000	0.000	0.000

Single A	% Variance explained by each PC				
	PC1	PC2	PC3	PC4	PC5
Raw data	84%	9%	3%	3%	1%
Bootstrap	84%	9%	3%	3%	1%
K-means	90%	6%	3%	1%	1%
Twofactor	82%	18%	0%	0%	0%
Vasicek	88%	11%	0%	0%	0%

BBB	Eigen values				
	PC1	PC2	PC3	PC4	PC5
Raw data	0.085	0.016	0.004	0.003	0.002
Bootstrap	0.083	0.016	0.004	0.003	0.002
K-means	0.048	0.007	0.004	0.001	0.000
Twofactor	0.064	0.007	0.000	0.000	0.000
Vasicek	0.031	0.005	0.000	0.000	0.000

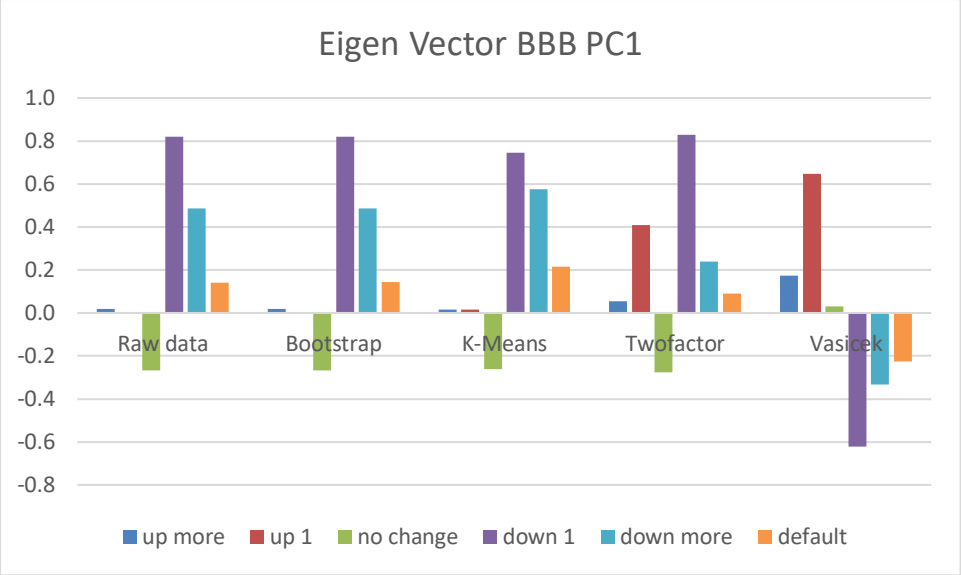
BBB	% Variance explained by each PC				
	PC1	PC2	PC3	PC4	PC5
Raw data	77%	15%	4%	3%	2%
Bootstrap	77%	15%	4%	3%	2%
K-means	81%	11%	6%	1%	0%
Twofactor	90%	10%	0%	0%	0%
Vasicek	87%	13%	0%	0%	0%

Table 8 – Eigenvalues for each of the four models and the raw data

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

- The Bootstrap model has movements most similar to the raw data, which is in line with expectations as it is simply the raw data sampled with replacement.
- The K-means model eigen values are also similar to the raw data; again because 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 TwoFactor and Vasicek both have just two components. This is to be expected as these models summarise full transition matrices with 2 or 1 parameters respectively
- From the eigen values none of the models appear to be significantly different from the data to make them inappropriate for use

The eigen vectors for each model are now compared below for just BBB assets:



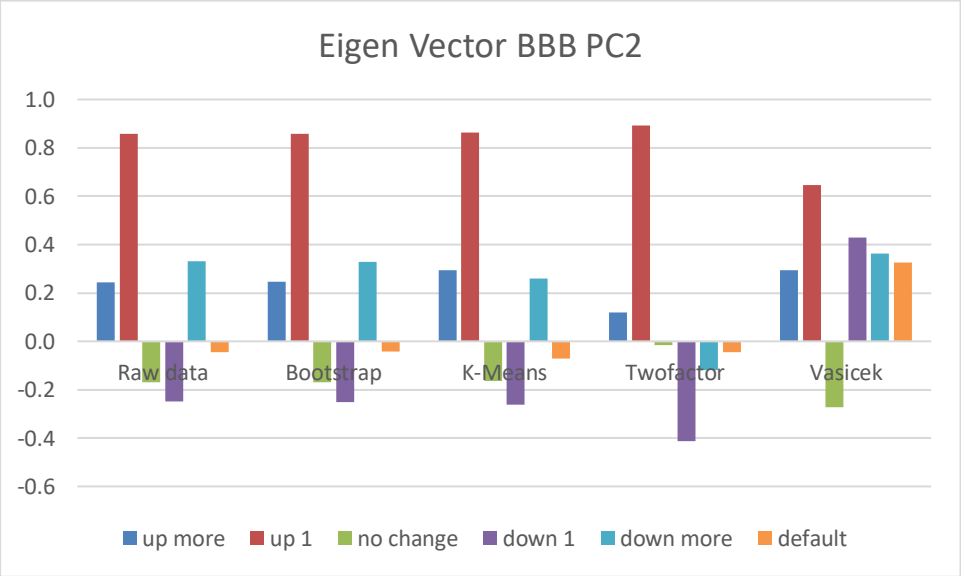


Figure 9 – plots of the Eigenvectors of four models and raw data

Commentary on the Eigen Vectors

The plots above show the first two eigen vectors for BBB rated assets which have a similar pattern to other ratings.

For PC1 there is a clear similarity between the raw data, the bootstrapping and the 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 parameter model is similar in nature to these non-parametric approaches, albeit with a larger rise in the upgrades increasing by one rating. However, the Vasicek 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 Twofactor model is also similar but notable with the “down more” than one rating category in the opposite direction to the non-parametric models. The Vasciek 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 Vasicek is perhaps similar to 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 historic data. The Vasicek model is clearly structurally different from the raw data, with PC2 of the Vasciek being more akin to PC1 for the other models. This suggests the Vasicek is not capturing the movement in historic data. The Twofactor model is an improvement on the Vasicek in that it is a closer representation of the movement underlying raw data, which is what might reasonably be expected from the additional parameter.

10.5 Backtesting comparison

A key requirement for transition and default models is that they meet any back testing requirements. For example, UK specific requirement for Matching Adjustment Internal models is given in Supervisory Statement 8/18 4.3.4 "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 below.

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%

Table 9 – the 1932 transition matrix

The four models being compared are:

1. Bootstrapping
2. The K-means model
3. The Vasicek model
4. The Two Parameter model

The bootstrapping model is simply using the raw data, sampled with replacement. And so the most extreme percentile is simple 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 transitions 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 Vasicek model calibrated to the data set described in section 5, gives a rho value of just over 8%. The 99.5th percentile from this model is shown below

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 from the Vasicek Model

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

- The defaults are lower
- 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 Vasicek model, moving the rho parameter to say 30% as an expert judgement loading – specifically targeted at passing the backtesting requirements. The updated 99.5th for this strengthen Vasicek model is shown below.

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%

Table 11 – the 99.5th transition matrix with strengthened Vasicek calibration

Comparing 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 parameter 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:

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%

Table 12- the Two parameter 99.5th transition matrix

Comparing 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 Vasicek calibration, but slightly lower than the 1932 matrix
- The leading diagonal values are slightly higher than the 1932 matrix

Overall the two parameter transition matrix at the 99.5th percentile is slightly weaker than the 1932 matrix; but stronger than the Vasicek 99.5th percentile. This model could also be strengthened in a similar way to the strengthening of the Vasicek 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

10.6 Objectivity comparison

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

10.6.1 Bootstrapping

Bootstrapping involves sampling a large number of data points from the historic data with replacement. This model is the most simple method to apply without any inclusion of expert judgement.

10.6.2 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.

10.6.3 Two parameter model

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

10.6.4 Vasicek model

The assumptions behind the Vasicek model include a number of expert judgements around the way the risk behaves. In order to pass the 1932 backtest an expert judgement overlay is required to strength the calibration.

10.6.5 Summary of model comparisons

The comparisons between the models are summarised in the table below:

Comparison	Bootstrapping	K-means model	Vasicek	Two parameter
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 historic 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	Highly flexible and can be set to the required use with appropriate judgements	Is used widely; but does not capture historic movements in the data well	Flexible model for a range of uses; additional parameters to Vasicek allows better replication of historic data movements

Table 13 – summary of model comparisons

The models explored all have strengths and weaknesses and all could be used for a variety of uses. The Vasicek model is used very extensively in the insurance industry, but it does not replicate historic movements in the data as well the K-means model or the Two parameter model.

11 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 parameter model) are not previously captured in the literature. The four models have been compared with a number of metrics including a new test of using PCA to compare model movements to historic data movements for transition matrix models. The PCA based test has highlighted a deficiency in the Vasicek model in that it does not replicate the way historic data moves. The PC1 of the Vasicek is not well matched to the PC1 of the underlying data, and the PC2 of the Vasicek is perhaps closer to PC1 in the underlying data. The other three models shown in this paper capture the historic movement of the underlying data significantly more accurately than the Vasicek model.

The non-parametric models have the advantage of having historical movements very close to the historic data; but the bootstrapping approach has a limitation of not producing stresses worse than the worst historic 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 Parameter model has an advantage of being relatively simple to apply, with an improved representation of the historic data (relative to the Vasicek model); but not as close as the non-parametric models.

12 References

Amemiya&Wu. (1972). The effect of aggregation on prediction in the autoregressive model. Journal of American Statistical Association, 628-632.

Belkin, Barry, Stephan J. Suchower, and Lawrence R. Forest, Jr. (1998) The effect of systematic credit risk on loan portfolio value-at-risk and loan pricing. CreditMetrics Monitor, 1st quarter 1998.

Cont, R. (2005). Volatility Clustering in Financial Markets: Empirical Facts and Agent-Based Models. In A. K. Teyssiere, Long Memory in Economics. Springer.

D. Kwiatkowski, P. C. (1992). Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root. Journal of Econometrics, 1590178.

Feike_Drost_Nijman. (1993). Temporal Aggregation of Garch Processes. Econometrica, 909-927.

Ljung&Box. (1978). On a measure of lack of fit in time series models. Biometrika, 297-303

Merton, R.C. (1974) On the Pricing of Corporate Debt: The Risk Structure of Interest Rate. Journal of Finance, 29, 449-470.

Moody's. (2006). Measuring Corporate Default Rates. https://www.moodys.com/sites/products/DefaultResearch/2006200000425249.pdf: Moody's.

Philip, J. R. (2004). Structural vs reduced Form Models. Journal of Investment Management, 2(2).

PhillipsPerron. (1988). Testing for a unit root in time series regression. Biometrica, 335-346.

Protter, R. A. (2004). Structural vs Reduced Form Models: A New Information Based Perspective. Journal of Investment Management.

Rosch D, Scheule H (2008) Stress Testing for Financial Institutions, Applications Regulations and Techniques. Publisher 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 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; Volume 36, Issue 12, December 2012, Pages 3133-3149

Vasicek, O. A. (1987). Probability of Loss on Loan Portfolio. San Francisco: KMV Corporation. Journal of Financial Risk Management, Vol.5 No.4, December 16, 2016

Vasicek, O. (2002). The distribution of loan portfolio value. Risk, 15(12), 160-162.
Wang, Y. (2009). Structural Credit Risk Modeling: Merton and Beyond. Risk Management, 16.

13 Appendix A – Details of K-means model

Technical Details on Grouping

1. We have used standard K-Means clustering algorithm (un-supervised learning algorithm) to perform the grouping.
2. K-Means clustering algorithm tries to classify a given data set through a certain number of clusters (assumed k clusters) fixed apriori within an unlabelled multidimensional dataset. It accomplishes this using a conception of what the optimal clustering looks like:
    a. The “cluster center” is the arithmetic mean of all the points belonging to the cluster;
    b. Each point is closer to its own cluster center than to other cluster centers.
3. K-means is a generally implemented using expectation maximisation procedure as follows:
    a. Guess some cluster centres
    b. Repeat step a until converged
        i. E-Step (Expectation Step): Assign points to the nearest cluster center. It is named as expectation step because it involves updating our expectation of which cluster each point belongs to.
        ii. M-Step (Maximisation Step): Set the cluster centers to the mean. It is named as maximisation step because it involves maximising some fitness function that defines the location of the cluster centres – in this case, that maximisation is accomplished by taking a simple mean of the data in each cluster.
4. The objective function also known as squared error function is given by:

$J(V) = \sum_{i=1}^{c}\sum_{j=1}^{c_i}\left(\left|\left|x_i - v_j\right|\right|\right)^2$ where $\left|\left|x_i - v_j\right|\right|$ 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 $X = \{x_1, x_2, .., x_n$ 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 centers.
2. Calculate the distance between each data point and cluster centers.
3. Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers.
4. Recalculate the new cluster center using:

$v_i = \left(\frac{1}{c_i}\right) * \sum_{j=1}^{c_i} x_i$ c_i is the number of data points in i[th] cluster .

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

14 Appendix B – Details of Two Parameter model

This appendix gives an example for how the Two Parameter 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 Default category). These are coloured yellow in the below matrix.

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,

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 (downgrades plus defaults). The defaults used in the weighting and downgrades are shown in yellow. The upgrades are in blue, the downgrades are in green.

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%

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

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.

Institute and Faculty of Actuaries

London

7th Floor · Holborn Gate · 326-330 High Holborn · London · WC1V 7PP
Tel: +44 (0) 20 7632 2100 · Fax: +44 (0) 20 7632 2111

Edinburgh

Level 2 ·Exchange Crescent · 7 Conference Square · Edinburgh ·EH3 8RA
Tel: +44 (0) 131 240 1300 · Fax +44 (0) 131 240 1311

Oxford

1st Floor · Park Central · 40/41 Park End Street · Oxford · OX1 1JD
Tel: +44 (0) 1865 268 200 · Fax: +44 (0) 1865 268 211

Beijing

Level 14 · China World Office · No.1 Jianguomenwai Avenue · Chaoyang District · Beijing, China 100004
Tel: + +86 (10) 6535 0248

Hong Kong

1803 Tower One · Lippo Centre · 89 Queensway · Hong Kong
Tel: +11 (0) 852 2147 9418

Singapore

5 Shenton Way · UIC Building · #10-01 · Singapore · 068808
Tel: +65 8778 1784

www.actuaries.org.uk