SESSIONAL MEETING DISCUSSION

"The calibration of transition risk for corporate bonds" by the Extreme Events Working Party

[Institute and Faculty of Actuaries, Sessional Webinar, Monday 15 May 2023]

This discussion relates to the paper presented by Andrew Smith, Florin Ginghina, James Sharpe and Gaurang Mehta at the IFoA sessional event held on 15 May 2023.

Moderator (Mr P. C. Jakhria): The Extreme Events Working Party has written papers and commented on many aspects of capital markets modelling, starting with equities, and moving on to interest rates, credit spreads, correlations, and even economic scenario generators. This time the exploration is of one of the lesser explored frontiers; the calibration of transition risk for corporate bonds. Part of the reason for there being reduced exploration of this topic in the past is because it is a particularly challenging topic, with numerous viewpoints, and yet is quite critical for valuing annuity funds and pension schemes.

The working party members are:

Andrew Smith – a capital markets modelling specialist. Andrew worked in the insurance industry for around 30 years and has been in an academic role for the last five years.

Florin Ginghina – a consulting actuary with 15 years of experience. Recently, Florin has been developing plans for annuity writers and working on major model changes for credit risk and longevity risk, as well as on mergers and acquisitions (M&A). Over the last year Florin has worked extensively on asset and liability management (ALM) and capital management for Solvency II.

Guarang Mehta – very experienced in the Solvency II methodology, risk calibrations and volatility adjustments, and has been working on a number of implementation areas for IFRS 17.

James Sharpe – the Chair of the Extreme Events Working Party and has 25 years of consulting experience in the industry. Recently, James has been mainly involved in calibrating internal models, matching adjustment portfolios and various stress testing projects.

Mr J. A. Sharpe, F.I.A.: Firstly, we will discuss the data and modelling requirements. The data we have is essentially a series of transition matrices. Figure 1 has an example of a transition matrix.

This is a complete mapping of probabilities from one rating to another over a given time period. The time period in which we are interested is one year. That is because the transition matrices we have each cover a period of one year. We had a data set that was freely available. It is the S&P data set, and it contains just over 40 transition matrices, going back to 1981. We also includes some transition matrices from the 1930s that give an extreme event.

We want to model this transition risk in a full probability distribution. The key challenge is to have a relatively simple, intuitive model that captures the volatility we see in historical transition matrices. The first and simplest model is the bootstrapping model and, as in statistics, you have a straightforward sample with replacement. If you want to create a bootstrap model, you would perhaps take 10,000 samples from your small data set with replacement. You would take one transition matrix, put it into your calculation engine, run the calculation, and then place it back in the data pool again and sample another one. You would repeat that 10,000 times. It is very similar to using the underlying data itself, but instead you are sampling from the data repeatedly. It has a

- The data used to model default and transition risk is the transition matrix

- A matrix with a full range of probabilities for any rated asset moving to any other rating within a defined timeframe

- Historical data has a different transition matrix for each year giving and one of the most difficult risks to model for most insurers

- The challenge is to produce a model that captures the variability in historical transition matrices whilst being relatively simple to implement

- Simplest model – bootstrapping – sampling with replacement

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%

1981
1982
1983
…
…
2018
2019

Figure 1. Data and modelling requirements.

A statistical model of a transition matrix. Each transition matrix can be defined by two parameters – Inertia and Optimism which cover the main sources of historical variability.

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%

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%

Inertia

- The sum of the leading diagonal of the transition matrix
- Gives a measure of how much movement there has been from existing credit ratings

"Stress Testing for Financial Institutions" Rosch et al

Optimism

- The ratio of upgrades to downgrades/ defaults, weighted by the default amount
- Gives a measure of how upgrades have compared to downgrades

Figure 2. Two-factor model – description.

number of strengths and weaknesses. The real strength is that it is as close as possible to the actual data, because you are resampling the actual data. The main weakness is that it does not give you anything worse than the 1-in-200 stress. Here, the worst thing that can happen is effectively the 1932 matrix in the data set. It does not give you anything worse than the worst event in history, which is a fundamental flaw with any capital model.

The two-factor model is a purely statistical model of a transition matrix. It is not one of the models based on economic theory. It just looks at the features of the transition matrix, such as the way it has changes over time, and models them directly.

We used the two-factor model in the paper because it has a couple of features, which are explained in Figure 2, that tend to explain historic variability in the data quite nicely.

Both features are detailed in a book called "Stress Testing for Financial Institutions" edited by Harald Scheule and Daniel Roesch (Scheule and Roesch, 2008). They are termed "inertia" and "optimism." Inertia provides a measure of how many bonds have changed rating over a time period. It has a very intuitive feel to it. The lower the value, the more transitions you have had during the time period. If they are all equal to one, then nothing has changed rating during the time period. The second measure, optimism, is also very intuitive. It is the ratio of upgrades and downgrades. In the paper we have weighted that ratio by the probability of default, in line with Scheule and Roesch, but there are other ways that you could implement the weighting. You could

- Calibration is based on historic transition matrices. Each historic matrix is converted to two parameters

- This gives two time series to which a probability distribution or time series model can be fit

- The probability distributions can be combined into a single joint probability distribution using a copula

- This gives a full risk distribution for each of the two parameters

- A base transition matrix is adjusted by simple scaling of the transition matrix elements so that the adjusted matrix has the same Inertia / Optimism as required
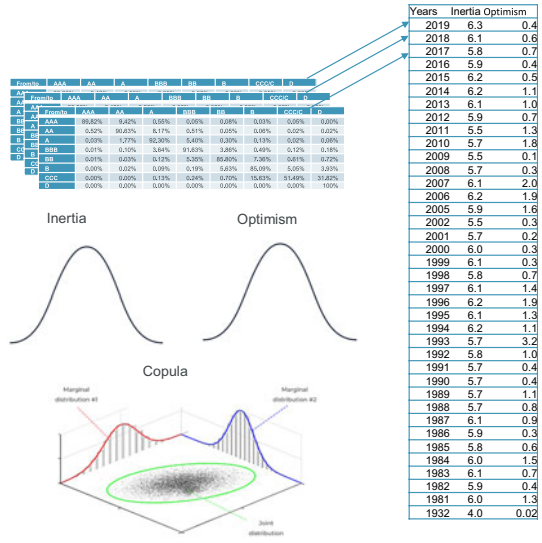
From/To	AAA	AA	A	BBB	BB	B	CCC/C	D
AAA	89.82%	9.42%	0.55%	0.05%	0.08%	0.03%	0.06%	0.20%
AA	0.52%	90.63%	8.17%	0.51%	0.05%	0.06%	0.02%	0.02%
A	0.03%	1.77%	92.30%	5.40%	0.30%	0.13%	0.02%	0.06%
BBB	0.01%	0.10%	3.64%	91.63%	3.86%	0.49%	0.12%	0.16%
BB	0.01%	0.03%	0.12%	5.20%	85.60%	7.30%	0.61%	0.72%
B	0.00%	0.02%	0.09%	0.19%	5.63%	86.09%	5.06%	3.33%
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%

Inertia   Optimism

Copula

Years	Inertia	Optimism
2019	6.3	0.4
2018	6.1	0.6
2017	5.8	0.7
2016	5.9	0.4
2015	6.2	0.5
2014	6.2	1.1
2013	6.1	1.0
2012	5.9	0.7
2011	5.5	1.3
2010	5.7	1.8
2009	5.5	0.1
2008	5.7	0.3
2007	6.1	2.0
2006	6.2	1.9
2005	5.9	1.6
2002	5.5	0.3
2001	5.7	0.2
2000	6.0	0.3
1999	6.1	0.3
1998	5.8	0.7
1997	6.1	1.4
1996	6.2	1.9
1995	6.1	1.3
1994	6.2	1.1
1993	5.7	3.2
1992	5.8	1.0
1991	5.7	0.4
1990	5.7	0.4
1989	5.7	1.1
1988	5.7	0.8
1987	6.1	0.9
1986	5.9	0.3
1985	5.8	0.6
1984	6.0	1.5
1983	6.1	0.7
1982	5.9	0.4
1981	6.0	1.3
1932	4.0	0.02

Figure 3. Two-factor model – calibration.

use a weighting that does not put as much weight on the lower-rated bonds, or you could have separate measures for different ratings, and so on. The way we have implemented it is to use probability of default as the weighting for upgrades and downgrades.

The strength of this model is that you can calculate these two parameters for any transition matrix that has occurred in the past. The table on the right-hand side of Figure 3 shows these two parameters calculated for one of our historical data sets.

We have modified the problem from having a string of transition matrices to having two time series to which we can fit probability distributions. That substantially reduces the complexity of the problem. The time series are correlated and we used a copula to combine them and create a full risk distribution for the two parameters. For any percentile we want, we can read off these two parameters. Finally, we use that data to create a modelled transition matrix. That was done using the long term average matrix, which we call the base matrix. Some very simple scaling is applied to scale the inertia and the values in the off-diagonal piece to get the correct value for optimism, so you can effectively have a model transition matrix. We have a full distribution of our parameters that leads to a full distribution of our transition matrix values.

That is the model we talk about in the paper. It does have two factors and this method of fitting distributions is a new approach, but there is a more embedded, one factor, model that has been more widely used and that is the Vašíček model.

Mr L. F. Ginghina, F.I.A.: We can describe a company's asset returns using geometric Brownian motion and then by applying Itô's lemma and setting time to maturity we get the well-known formula from Merton. That is the first equation on the left-hand side of Figure 4.

Vašíček was interested in a portfolio of identical assets. He moved over from Merton's formula to the same formula applied to identical assets with a portfolio of identically distributed assets. That is the second equation on the left-hand side of Figure 4.

Vašíček noticed that the distribution of the company's asset returns, variable x_i, belongs to the class of multivariate normal distributions. Assuming equi-correlations between all companies, the third equation on the left-hand side of Figure 4 is a description of a multivariate normal distribution. Interestingly, Vašíček made a number of assumptions. These are that the companies within the portfolio are identical and are subject to the same probability of default. In addition, the

- Merton's model of a company's asset return (1974):

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

Where X represents a firm's asset return, and X follows a Standard Normal distribution.

- Oldřich Vašíček (1987) extended the Merton's model to a portfolio of assets (i = 1, …, n):

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

- Vašíček also observed the following property of equi-correlated Standard Normal variables:

$$X_i = Z\sqrt{\rho} + Y_i\sqrt{1-\rho}, \text{ i = 1,…, n}$$

Where $Z, Y_1, Y_2, …, Y_n$ are mutually independent Standard Normal variables, n the number of firms in a portfolio

- Vašíček replaced X_i in Merton's model with $X_i = Z\sqrt{\rho} + Y_i\sqrt{1-\rho}$
  Where:
  » Variable Z is as common across the entire portfolio
  » Variables Y_i are i^{th} firm's specific variables
  » Parameter ρ is the asset portfolio correlation (and it's as an important driver of credit risk as it gives a measure of joint probability of default).
  » A portfolio asset correlation is like Modern Portfolio Theory.

- A key result in Vašíček's framework follows − a firm's probability of default conditional on Z in a large portfolio of assets:

$$P(firm\ i\ defaults\ |\ Z) = \Phi\left(\frac{x_i - Z\sqrt{\rho}}{\sqrt{1-\rho}}\right)$$

- For the avoidance of doubt, Vašíček's framework introduced in our slides is different to Vašíček's model for interest rates.

Figure 4. Vašíček's model – description.

correlation parameter between them, parameter ρ, is restricted to an interval between 0 and 1. In other words, all the correlations between the companies are positive.

Another innovation introduced by Vašíček was to denote variable **Z** as common across the entire portfolio. In other words, it is the state of the economy. Variable **Z**, again, is a standard normal variable, so it can be negative or positive. A negative value means the economy is in a bad state, and a positive value means the economy is in a good state. The random variable **Y** is the idiosyncratic risk for each company. Importantly, there are some statistical properties necessary here, which is that **Z** and all the variables **Y** are independent. With this result in mind, Vašíček derived the conditional probability of default conditional on variable **Z**, which is the equation on the right-hand side of Figure 4. This is the key result from the work of Vašíček (1987), which simply answered his question, "What is the loss probability for a portfolio of identical assets?"

For the avoidance of doubt, when we talk about the Vašíček model, we mean the Vašíček credit risk model, not the Vašíček interest rate model. The Vašíček model is quite widely used in other industries, for example in the banking industry where it is part of the credit risk calculation underlying the Basel framework. Some firms use the Vašíček model for International Financial Reporting Standards 9 (IFRS 9) when projecting expected losses for credit risk portfolios, as it is a requirement to disclose the expected losses for a corporate bond portfolio. There are some other uses, particularly in North America. Vašíček developed his model back in 1987. This was at a time when a lot of corporate bond transactions took place in North America and people were looking for methods to produce a pricing framework for sub-prime assets, in particular bonds.

Belkin derived the method to apply the Vašíček model to a transition matrix in 1998. He introduced a number of simplifications, that is, in a transition matrix all firms have associated bonds, they are all identical, and they are all subject to the same probability of default. With that in mind, Vašíček's result can be used to derive a transition rate within a transition matrix via the formula on the left-hand side of Figure 5.

Belkin also introduced a numerical algorithm to calibrate the parameters within the Vašíček framework, particularly the ρ parameter, and to find values for the **Z** parameter that correspond to all historical years.

We have about 40 years of historical transition matrices. The key restriction in Belkin's paper is that the average transition matrix becomes the benchmark, that is, the average matrix corresponds to the value of the **Z** parameter being 0. Going through the numerical algorithm repeated for small steps applied to ρ you then obtain various series of **Z** values. You then pick up the ρ that corresponds to the series of **Z** values, which gives you a unit variance, because we know that the **Z** values are standard normal variables.

- A firm's probability of default conditional on Z can be applied to transition rates in a transition matrix.

- Fitted transition rates for the 'G to g' credit rating can be written as:

$$\Phi\left(\frac{x_{g+1}^G - Z\sqrt{\rho}}{\sqrt{1-\rho}}\right) - \Phi\left(\frac{x_g^G - Z\sqrt{\rho}}{\sqrt{1-\rho}}\right)$$

Where:
  » $\Phi(\cdot)$ represents the standard normal cumulative distribution function, with $\Phi(\infty) = 1$ and $\Phi(-\infty) = 0$
  » G is the credit rating at the beginning of the year (G = AAA, AA, …, CCC/C)
  » g is the credit rating at the end of the year (g = AA, A, …, CCC/C, and g+1 = AAA, AA, …, B).

- 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 subject to Z restricted to unit variance on the Standard Normal distribution.
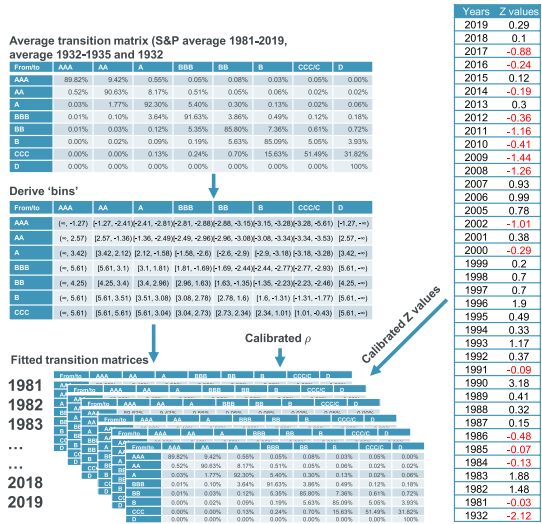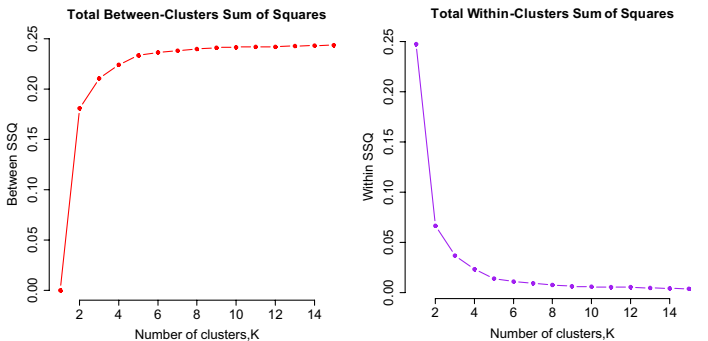
Average transition matrix (S&P average 1981-2019, average 1932-1935 and 1932)

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%

Derive 'bins'

From/to	AAA	AA	A	BBB	BB	B	CCC/C
AAA	(∞, -1.27)	[-1.27, -2.41)	[-2.41, -2.81)	[-2.81, -2.88)	[-2.88, -3.15)	[-3.15, -3.28)	[-3.28, -5.61) [-1.27, ∞)
AA	(∞, 2.57)	[2.57, -1.36)	[-1.36, -2.49)	[-2.49, -2.96)	[-2.96, -3.08)	[-3.08, -3.34)	[-3.34, -3.53) [2.57, ∞)
A	(∞, 3.42)	[3.42, 2.12)	[2.12, -1.58)	[-1.58, -2.6)	[-2.6, -2.9)	[-2.9, -3.18)	[-3.18, -3.28) [3.42, ∞)
BBB	(∞, 5.61)	[5.61, 3.1)	[3.1, 1.81)	[1.81, -1.69)	[-1.69, -2.44)	[-2.44, -2.77)	[-2.77, -2.93) [5.61, ∞)
BB	(∞, 4.25)	[4.25, 3.4)	[3.4, 2.96)	[2.96, 1.63)	[1.63, -1.35)	[-1.35, -2.23)	[-2.23, -2.46) [4.25, ∞)
B	(∞, 5.61)	[5.61, 3.51)	[3.51, 3.08)	[3.08, 2.78)	[2.78, 1.6)	[1.6, -1.31)	[-1.31, -1.77) [5.61, ∞)
CCC	(∞, 5.61)	[5.61, 5.61)	[5.61, 3.04)	[3.04, 2.73)	[2.73, 2.34)	[2.34, 1.01)	[1.01, -0.43) [5.61, ∞)

Calibrated ρ

Fitted transition matrices

1981
1982
1983
…
…
2018
2019

Calibrated Z values

Years	Z values
2019	0.29
2018	0.1
2017	-0.88
2016	-0.24
2015	0.12
2014	-0.19
2013	0.3
2012	-0.36
2011	-1.16
2010	-0.41
2009	-1.44
2008	-1.26
2007	0.93
2006	0.99
2005	0.78
2002	-1.01
2001	0.38
2000	-0.29
1999	0.2
1998	0.7
1997	0.7
1996	1.9
1995	0.49
1994	0.33
1993	1.17
1992	0.37
1991	-0.09
1990	3.18
1989	0.41
1988	0.32
1987	0.15
1986	-0.48
1985	-0.07
1984	-0.13
1983	1.88
1982	1.48
1981	-0.03
1932	-2.12

Figure 5. Vašíček's model – application and calibration.

• The k-means clustering method is an unsupervised machine learning technique used to identify clusters of data objects in a dataset.

• Selection of initial number of k centroids is one of key expert judgments under this approach. Centroids are data points representing the center of a cluster.

Total Between-Clusters Sum of Squares

Total Within-Clusters Sum of Squares

Between SSQ / Within SSQ vs Number of clusters, K

• We have used 1932 matrix, 1931-1935 average matrix and S&P transition matrices from 1981 to 2020.

• Both within and between sum of square charts suggests that around 8 to 9 groups are sufficient to explain the variation in the data.

• Improvements in sum of squares does not improve after 9 clusters.

We given equal weights to all ratings in this analysis. We have considered the following transitions for K-Means analysis:
• AA→A,AA→BBB,AA→BB,AA→B,AA→CCC,AA→D and A→BBB,A→BB,A→B,A→CCC,A→D,
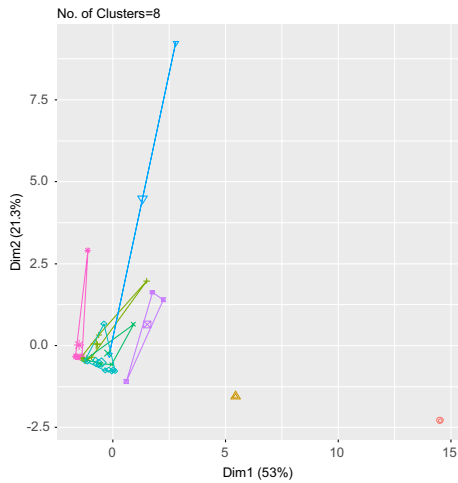• BBB→BB,BBB→B,BBB→CCC,BBB→D and BB→B, BB→CCC, BB→D.

Figure 6. Transition risk – K-means approach.

The results are on the right-hand side of Figure 5. These are the **Z** values that we obtained, applied to our historical data. The value for ρ that we derived was about 8%. There is a relevant paper dealing with Basel II which discusses the appropriate value of the ρ parameter. This gives a range for ρ, but for the banking industry, which has a different exposure to corporate bonds and loans. They have an interval for ρ between 12% and 24%, so the 8% we derive is good enough in our view.

Mr G. J. Mehta, F.I.A.: One of the key non-parametric models we have explored is the K-means model. The idea is very simple. We take the entire historical data for each of the years available and apply the K-means clustering algorithm to see how many total clusters there are in the data. As you can see in the two charts in Figure 6, we have analysed that based on between and within sum of squares errors.

The charts show the relationships between the number of clusters and the sum of the squares. You can see on the right-hand side chart that after eight groups there is not a material improvement.

No. of Clusters=8

- The cluster chart is a pictorial representation of these 8 clusters. These are covered in detail in table below.
- It shows 1932 (Grp 1) and 1935 (Grp 2) are separate groups and the farthest from the rest of the groups as they are the extreme matrices that we have observed in the history.
- The next group of extreme transitions is the group which contains 1986, 2002 and 1982 transition matrices (Grp 7).
- Grp 3, Grp 5 and Grp 8 are more within the body.

Group1	Group2	Group3	Group4	Group5	Group6	Group7	Group8
1932	1935	2018	1993	2015	1987	2002	2019
		2016	1988	2012	1985	1986	2017
		2008	1984	2011		1982	2014
		2006	1983	2009			2013
		2005		2003			2010
		1997		2001			2007
		1995		2000			2004
		1992		1999			1996
				1998			1994
				1991			
				1990			
				1989			
				1981			

Figure 7. Transition risk – K-means approach cluster chart.

We have done the analysis for around nine groups applying equal weighting to each of the ratings. Depending on a firm's exposure to different ratings and the basis risk reduction areas, they might want to increase the weight of some of the transitions. For example, if a firm is holding fewer AAA bonds, they might want to reduce the weight on the transitions from AAA bonds, but increase it, say, in AA or A bonds. Depending on these type of adjustments you will get a slightly different number of clusters and different groupings.

Based on the K-means algorithm, we identify the eight groups shown in Figure 7. For each of the centroids of the cluster we identified how many historical transition matrices fit into this.

In Figure 7 you can see that cluster 1, the red point on the rightmost side, is coming out as the 1932 matrix, which is one of the most extreme observations we have. Cluster 2, the yellow point, is the 1931-1935 average matrix, which was again one of the most severe during that crisis period. The other ones are more or less benign periods. In a few years, 2002 and 1981 for example, some of the transitions were higher. This is how the grouping comes out when you apply the clustering algorithm to the data.

Once the clustering is complete, the next step in calibration is applying expert judgement to decide which percentile is represented by which matrix. For example, 1932 represents a 1-in-200 year event and 1935 represents a 1-in-100 year event. Therefore, the matrices are placed on the real line based on these percentiles. The average of these matrices is a centroid, and the remaining ones are interpolated or extrapolated depending on what we need. So, for this purpose we have put the square of the 1932 matrix (i.e. the 1932 matrix multiplied by itself) as the 0th percentile, or the most extreme one. 1932 has a 1-in-200 percentile, and so on.

That is the essence of the K-means approach. You can see that it is non-parametric and requires expert judgement. It has an advantage as a lot of stakeholders are interested in making sure that the most extreme 1-in-200 year event does represent at least the 1932 matrix. With this method, we can handpick the outputs to ensure these sorts of requirements are met.

Figure 8 shows the model comparison. We have compared each model against different criteria.

In terms of replicating or representing the historical data, the bootstrapping and K-means approaches better represent the data than the Vašíček model, which is based on a single parameter, and therefore likely to produce a poor representation of the data. The two-parameter model is somewhere in between, depending on how you calibrate the two parameters.

Comparison	Bootstrapping	K-means model	Vašíček	Two parameter
Replication of historic movements	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

Figure 8. Model comparison.

The 1932 backtest is meant to ensure that extreme events are represented. For example, the 1932 event is a 1-in-200 year event. The K-means and bootstrapping approaches will be able to reproduce that. With the Vašíček model and the two-parameter model, you might have to increase the volatility parameter or some other parameter to come up with that type of extreme percentile in your distribution.

In terms of objectivity, the K-means approach requires expert judgement whereas the Vašíček or bootstrapping approaches require less expert judgement and can be objectively calibrated. The two-parameter model requires a limited amount of expert judgement compared to the K-means approach, but more than the Vašíček or bootstrapping approaches.

Bootstrapping is the simplest approach, the others introduce varying degrees of complexity. How much complexity you want depends on your portfolio, the objectives of the model, how you calibrate it, how many years of data you are using, and how many sectors you are using. A lot of firms have separate sectors for financial and non-financial, and they have sourced that level of detailed data, and with that information the calibration complexity will change.

Considering its breadth of use, the bootstrapping approach is less appropriate, particularly for Solvency II and the type of work we are trying to do, because you cannot produce a future view – you can only represent what has happened historically. Similarly, in the K-means approach, your best guess is as good as what your expectation of each of the real line events is. The other two models can be tweaked by increasing the parameters or applying a loading for future events.

Now we move to the question-and-answer session.

Question: In the post-pandemic world, with inflation back on the agenda, it feels like there is going to be a lot of volatility and certainly shorter economic cycles. How are these different methods able to adapt to different forward-looking views of the economic cycle?

Mr A. Smith: We have implemented all the models we have described here as through the cycle models. That is not the only way they could be implemented, and indeed Florin (Ginghina) mentioned the Belkin implementation, which tries to capture more about a point in time. Implementing models through the cycle has the disadvantage that, if you know you are about to go into an economic recession, it does not increase your capital requirements.

Question: Whatever modelling approach we use is going to be reliant on some kind of expert judgement, and for the solvency capital requirement (SCR) in particular the judgement about the 1-in-200 year event is probably going to be one of the most sensitive judgements. I think

the industry has converged on the unadjusted 1932 matrix being representative of that 1-in-200-year event. If we look at pandemic modelling, we use a figure of 1.5 per million in Solvency II. We arrived at that figure by looking at the 1918 pandemic and adjusting it to allow for how things have changed since then, meaning more transmission between people but better healthcare. What do you think about this judgement, that 1932 is representative of a 1-in-200-year event, given that the corporate environment is completely different now from what it was in 1932?

Mr Mehta: The 1932 matrix was mainly based on railways and other non-financial industries and in today's world, where most of the bonds or financial assets are more or less supported by property, relevance is a question. However, I think there is a regulatory expectation that you need to represent, and that is what everyone tries to achieve. We did consider applying different scalars to capture that change, but then that scalar would require a calibration in its own right. What sort of cash flows were there for those assets before? What was the risk represented by those assets and what are the risks represented now? That requires a lot of expert judgement and use of many regression techniques.

Mr Ginghina: 1932 is definitely an interesting year. When you see the long duration of the annual transition matrices, it stands out. We can discuss at length what caused that. It was partly the exposure to particular industries at the time. There is a lot of literature discussing the particulars of the Great Recession. If we think about anything coming from the Prudential Regulation Authority (PRA), they do mention 1932 and 1933. In that context I would say that it is a nice idea to anchor to 1932, but a wider consideration is the whole recessionary period. If you think about the Great Recession, it goes from 1930 all the way to 1933, and it is about the cumulative impact of those four years. On top of that, I think every company is different. In today's world we are better prepared to battle with recessions. For example, we have people who specialise in recoveries. I do not think that was the case back then. So yes, these are additional considerations.

Question: Should we be using 1932 without any kind of adjustment? As you mentioned, there are lots of reasons why we are different, and the graphs show that 1932 really stands out. If you look at the Vašíček model and the two-parameter model, they do need judgement to strengthen them to reach the 1932 situation. Should this cause us to think again about whether we should use 1932 unadjusted as our anchoring point?

Mr Ginghina: One of the things that was particularly interesting when we did the research on the Vašíček method was that the correlation parameter does reflect a particular behaviour. For example, when the economy is in a bad state, you would expect more firms to default and vice versa. It is one area where we have seen approaches in the market where people look at the correlation parameter in particular, to better reflect a recessionary period. It is one option. Another option is to look at different distributions, not necessarily the standard normal distribution, but maybe Student's t or other fat-tailed distributions.

Mr Mehta: I think this point does not bite too much, because for most of the firms, whether you are a matching adjustment (MA) firm or non-MA firm, the biting scenario is well below a 1-in-200 year event, particularly for this risk. Therefore, even if you are using a 1-in-200 year event, which is on the extreme side, and you are representing it by 1932, your actual biting scenario is well within the body.

Mr Sharpe: If you wanted to adjust the 1932 matrix, that is fraught with a huge range of problems. There is something appealingly objective about saying "We're stronger than the worse thing that's happened in our data set." There might be reason to have something stronger, but something less strong would perhaps seem rather weak.

Mr Smith: Often people ask "What will the Prudential Regulation Authority (PRA) let me get away with? How far can I push it and still get my model approved?" We need to recognise that that

is not a statistical question – it is a political or social question. Perhaps calling it expert judgement is something of an exaggeration. It is really just describing etiquette, and it is a very UK-specific etiquette we are talking about.

Question: Even though the diversified biting scenario might not be the 1-in-200 year credit event, the shape of our credit distribution is going to inform what scenario we end up with. The extreme tail is going to drive our capital requirements to some extent. In practice, a lot of life insurance firms would really like to see that tail weakened if there was a good reason to do so. The way that we set a 1-in-200 year assumption might not be a statistical matter, but I would not necessarily say it is a social convention either. Presumably, there is a true number that could be derived if somebody was to do the work to understand what credit ratings mean nowadays, and how rating analysts would respond to some adverse events in the broader economy. If we took a more bottom-up view of how the broader economy works and how badly things can go in one year, we could get to a more informed view of what that 1-in-200 year scenario looks like. There would still be lots of judgements along the way, but we could work through those. Somebody must put some thought in, and I find, in practice, rating analysts do not want to do it. Maybe it is a role for actuaries to do that sort of thinking about extreme events.

Question: I wonder whether another criterion to consider might be the modelling of dependencies between risks. A K-means model, for example, might not be as susceptible to producing a series of risk drivers that you could correlate with other risk drivers.

Mr Mehta: You have to calculate the dependency calibration for transitions and spreads separately, but that is a separate assumption. If you analyse historical data for, say, one or two transitions against the spread movement, then the correlation is very weak. That is because, generally, credit events have been either liquidity crises or default and downgrade crises. 1932 was a default and downgrade crisis. The period from 2008 to 2009 was more of liquidity crisis. So, spreads blew up before credit was downgraded significantly. Hence, historical data analysis will be very weak. But, in any case, whether you use any of the models, that assumption will be a separate assumption anyway.

Mr Smith: The paper from Belkin produces a single factor for which you have a time series and you can attempt to find, for example, the correlation between that and equity returns or interest rates. The two-factor model produces two factors that are correlated, not only with each other, but with other things such as spreads and interest rates. As you increase the complexity, you increase the number of dependencies that you have to either estimate or judge in relation to other risks. The difficulty with the Vašíček model is the single factor, the optimism factor. There is no variability in inertia. However, the principal component analysis we did elsewhere in the model showed that inertia was actually a bigger component than the optimism. In other words, the principal components are in a different order, which really makes the correlation tricky. You might imagine that the optimism would be correlated with the equity market. When the equity market goes down then you would expect more downgrades; when the equity market goes up you would expect more upgrades; and that is exactly what the Merton model predicts. In reality, the first factor is actually changes in inertia. You get years where there are lots of upgrades and downgrades, and the rating agencies seem to be very busy, and there are years without many upgrades or downgrades. That is the more significant component and that is very hard to capture. You cannot capture it with the simplest Vašíček model, you need to include other distributions. The t-Copula would give you that flexibility, for example.

Mr Ginghina: When we started work in this working group, I think we were very tempted to look into the insurance context. For example, how you calibrate transition risk on both the asset side and the liability side. However, different factors are relevant in insurance. It is long term, say 30 years. You have the glide path and you have trading considerations, and so on. Hence, we

decided to step back and to look at the question from a statistical point of view only. We intentionally limited our work to statistical considerations because otherwise we would have ended up with hundreds of pages of analysis and considerations.

Question: The two-factor model provides richer dynamics across the whole distribution. How do you trade off the extra flexibility with the extra complexity in the modelling?

Mr Sharpe: You have doubled the number of factors, but it is still very intuitive. When we say complex, what does that mean? When we considered how easy it is to explain the two factors to people, we found that the two-factor model was relatively intuitive. The Vašíček model is also intuitive in that it gives a single factor to each transition matrix to show how strong it is, either up or down. I do not think there is much extra complexity from having an additional factor. The K-means model is more complex, but again, you have this nice, smooth transition between all the historic data. It is a very good non-parametric model.

Question: Is there danger of becoming too focused on the analysis that purports to explain what happened in the past, and too comforted by the supposed understanding of what drove the extreme outcomes and the belief it will not happen again? That is relevant to 1932. Are we trying to read too much into very specific data points?

Mr Smith: There is always the risk that you get too much comfort from having fitted a model and you lose sight of the possibility that your model might be wrong. I think that before questions started being raised about the 1930s, it would have been quite common for some organisations to look only at much more recent data, certainly in the banking sector. When the PRA started to say that people ought to look at the 1930s, I think that was a very uncomfortable issue for people to consider, because the 1930s were so much worse in terms of downgrades than anything any of us can remember within our working lifetimes.

Often data from the 1930s would previously not have been actively disregarded but might not have met data quality criteria. There is a nice, consistent data set from S&P starting in 1981. It has some issues, as with any data set, but it has being compiled by the same organisation on a more or less consistent basis, and the data is in the public domain, so it will be subject to some controls. People would perhaps take some comfort from that. Writing a data quality statement about the data from the 1930s is much harder. It is a lot longer ago, is mostly US data, and there are questions about the different sectors. It would have been easy to think of reasons to discard the 1930s, but the regulator chose not to, because the period was so awful.

Mr Mehta: If we put ourselves in the shoes of a regulator, it is very hard to justify excluding the data or believing that a future event will not be as bad as 1932. There are more observations, and there is more reporting nowadays, but, at the same time, boom and bust economies and shorter economic cycles are very much a possibility. This is especially true given artificial intelligence and the unknowns in the area of blockchain technology.

One point I wanted to add on your question is that when you are analysing the correlation, there is a lag. It takes six to eight months for rating agencies to complete their analysis and publish their reports, so when you are analysing correlation, that impacts the assessment as well.

Question: How differently does the two-factor model behave in the tail, when compared to the Vašíček model?

Mr Sharpe: We did a chi-squared test and the two-factor model seemed to fit better in the tail. However, we used a fairly simple weighting, so there could be more sophisticated weightings for both Vašíček and the two-factor model.

Mr Ginghina: The Belkin approach, which is the implementation of the Vašíček method applied to transition matrices, uses a more sophisticated approach. The goodness-of-fit looks at the number of issues by credit rating. It means that you require quite a rich data set, and it is rarely available.

Question: Should we consider modelling financials and non-financials differently on the basis of what happened in the great financial crisis where they did behave very differently, and if so, what are the pros and cons of doing so?

Mr Mehta: One of the key advantages of modelling them differently is that you are reducing the chances of possible basis risk in your portfolio. If you were, say, strong on financials and weaker on non-financials, then you would be basing your calibration on that. Another advantage is that, historically at least, during the spread crisis, financial firms were more exposed than non-financial firms. Financial firms may be more risky because there is a lack of backing assets behind them, compared to non-financial firms where there are sometimes physical assets. The advantages are that it offers flexibility, it reduces basis risk, and there is a more theoretical backing based on what history has taught us. Conversely, as we split the data into financial and non-financial, the number of data points is smaller for each of the splits than for one overall non-split data set, and the credibility of the data analysis also reduces. That is one of the key challenges. Another is that, in the past, we have seen financial crises. We have not seen the future. The next crisis could be financial or non-financial, and it could be more or less severe than previous crises. What we have seen in the past is not necessarily likely to happen in the future. This creates biases that are present in the calibration.

Question: If I understand correctly, you used a principal component analysis to compare their historical behaviour. Would a principal component analysis itself be a viable model?

Mr Smith: The reason we did not use principal component analysis to construct a model is because of the properties of the transition rates. The feasible range of the transition rates in a row is in what we call a simplex, a set of non-negative numbers that add up to one. It is a kind of tetrahedron, or some high dimension analogue of that. The usual way to apply principal component analysis is to decompose these things into separate components that are statistically independent of one another. Each of those components would have to have a bounded range, and so the feasible region is a cuboid. The problem is that there is no way of transforming a cuboid so that it doesn't either sit entirely inside the tetrahedron and you are missing out the corners, or it pokes out at the edges and you get lots of negative probabilities. We did not find a way to solve this, but we did try quite hard. The best approach that we found, for the historical methodology, was to use principal component analysis on a model that we built in a different way. We then summarised the output of what could be quite a high dimensional model, so that it could be compared to similar output from the historical data. We did quite a lot of thinking about what we call granulation with this binomial variability on top of changes in transition probabilities, which complicates things a lot. I am not going to go into the detail, but it means you need to be quite careful in how you do the principal components analysis. I am not saying there is not a possible solution, but it's much harder than it looks.

Question: Would it be possible to perform a principal component analysis that excludes the diagonal with a logit transformation to ensure that the rules add up to 100%?

Mr Smith: It is possible to do that. The problem is granulation, meaning it is difficult to look through that sort of transformation. The problem with granulation impact is that you get rates of transition equal to 0. So, in most years, for example, there will be not a single AAA bond that defaults and you then get a log of 0 appearing in the output, which is undefined.

Mr Mehta: I think the bigger problem with the principal component analysis approach on this data set is that there are 56 transitions, or a minimum of 56 dimensions that we are trying to analyse. You are trying to reduce 56 dimensions to 3 or 4 components. You can see in the historical data that only 8 or 9 transitions have real movement that you can model. So, from 56 dimensions you are reducing down to 8 or 9, and then generalising based on the 8 that you have

analysed, reducing that down to 3 and then extrapolating it back to 56. So, the shape of the transition is also sometimes strange.

Question: In the two-parameter model, you mentioned that you weighted it by default rate, which means that it is very strongly weighted towards sub-investment grade securities. It strikes me that you might want to weight it with reference to your actual portfolio mix, or what your portfolio mix would be after a 1-in-200 year stress, or maybe after a more normal range of downgrades. Have you experimented with that and did you find that it worked better or worse? Did you get a more realistic result for optimism?

Mr Sharpe: We did not experiment with different weightings. We just applied the method from the Roesch and Scheule book, which is weighting with default probability. You are absolutely right – it might be advantageous if you used the weighting in your matching adjustment portfolio, for example, perhaps just looking at investment grade bonds.

Mr Smith: We have worked from public data, and we have described it in a sufficient level of detail for somebody to replicate our calculations. So, that is a slightly different contribution from saying "We have used potentially proprietary data from behind a paywall, and we have analysed lots of different combinations of that data." It is giving you a blueprint for these models, which have not been very completely described previously.

Moderator: The discussion around the pros and cons shows the nuances and complexity of the problem that the working party is trying to solve, and also the different approaches available. Hopefully, this gives the users enough insight into the different methodologies to start having a view as to what approaches are best for their needs, and why. It remains for me to express my thanks to the presenters and all the other members of the Extreme Event Working Party who have worked on this project.