The Hazard Rate Matrix Approach to Credit Rating Transitions

By J. M. Pimbley[1]

Debt securities are financial instruments that obligate the issuer of the debt to make principal and interest payments at specified times to the investor. Failure of the issuer to make such payments is "default." One may characterize all outstanding debt instruments as being "in default" or "not in default."

The great majority of debt instruments — to which we now refer as "debt" or "bonds" — will not default prior to maturity. Let's consider the designations of "in default" or "not in default" as a primitive rating system for bond performance. Every outstanding bond falls into one of these two categories.

Credit rating agencies (CRAs) are incorporated financial firms that analyze bonds and provide opinions on bond default risk in the form of "ratings" that extend the "not in default" category of the primitive rating system.[2] These CRAs generally give a symbol D to the "in default" category and give symbols such as AAA, AA, A, BBB, BB, B, and CCC to seven "not in default"categories that represent increasing risk of future default. The precise symbols for these "letter grades" vary from one CRA to another. Further, each CRA has an expanded scale that boosts the number of "not in default" ratings from seven to roughly 20. In this article, we will generally use these seven "major" rating levels merely to simplify the discussion.

Meaning of the CRA Credit Ratings

The bond rating that carries the least risk is AAA. The "lower" rating levels of AA down to CCC denote progressively increasing default risk. The four largest CRAs have published "default tables" that show either targets or historical assessments of default probability over time for each rating level.[3] These tables would seem to provide a definition, or at least a benchmark, for the meaning of the CRA ratings. The CRAs, however, have generally avoided making the explicit claim that their default tables define their ratings.[4]

The CRAs also publish "rating transition matrices" that show the frequency of rating changes from one level (such as AA) to another (such as A) over periods of one year or longer.[5] These transition matrices are consistent with and complementary to the default tables since "default" is synonymous with rating transition to the default rating D. Thus, for example, the default probabilities for each rating for a tenor of one year form the (bottom) row of the one-year transition matrix for the final rating of D.[6]

Time-Dependence of the Credit Rating

Contemplation of the CRA rating transition matrix (RTM) leads us to consider the process for time evolution of credit ratings. Previous work begins with the one-year transition matrix and makes the assumption that rating changes are Markovian.[7] In this context, the description "Markovian" means simply that the likelihood of a rating change depends only on a bond's current rating and the RTM.

In practice, rating changes are not Markovian because the CRAs — inadvertently or otherwise — behave in a manner that depresses rating volatility. For example, if a currently AA-rated bond deserves a downgrade to BB, CRAs will typically downgrade the bond to A or BBB rather than impose the full rating change in one rating action.[8] Moody's Investors Service states: "Moody's ratings management practice, by avoiding reversals, necessarily produces positive serial momentum in ratings changes. A downgrade is much more likely to be followed by another downgrade than it is an upgrade."[9]

Before continuing, let us emphasize this point. Existing treatments of rating transitions employ a Markov paradigm. This treatment is most tractable and most sensible: ratings should be Markovian to maintain accuracy of default risk estimation. Yet CRA rating transitions are clearly not Markovian in practice. Hence, we will not try to fit empirical rating transition data with our model results. Rather, we view the relevance of this study to be the development of idealized (Markov) rating transition behavior to assist banks, portfolio managers and other practitioners in the development of accurate default rating scales.

Consequently, we develop in this article an alternative Markovian process for rating changes that we find to be more intuitive than existing methods that begin with the one-year RTM. The new treatment, which ultimately merges with the conventional method, has the advantage of both an easier derivation and better examination of the bond rating behavior for large times. We find, for example, that a certain matrix eigenvector forms a "natural rating distribution" to which all initial rating states evolve.

The nature of this finding — that we identify an eigenvector with a credit rating distribution — implies that the mathematical development and the finance implication are strongly linked. Much of the remaining exposition of this article is unavoidably mathematical. The following sections (up to "Practical Result – Examples") constitute this analysis, and the ensuing sections provide the finance interpretation.

Mathematics of Rating Transitions: Simple Case

To begin, let us consider the "primitive rating system" in which a bond is either "in default" or "not in default." At time $t=0$, we specify that the bond is not in default and that the probability of the bond not being in default at time $t>0$ is $S(t)$. The initial condition is $S(0)=1$. We specify the probability of the bond defaulting (which is identical to the probability of the bond being downgraded to the "in default" rating level) during the time period $(t,t+dt)$ as $\lambda S(t)dt$. Since a rating transition to the "in default" rating is a rating transition out of the "not in default" rating, we write

$$\frac{dS}{dt} = S'(t) = -\lambda S \tag{1}$$

Equation (1) shows the familiar hazard rate process that is well known for bond defaults and other applications.[10] The solution of equation (1) for the survival time $S(t)$ with a time-dependent hazard rate $\lambda(t)$ is

$$S(t) = exp\left[-\int_0^t du\, \lambda(u)\right] \tag{2}$$

We specify $\lambda(t)>0$ for all $t>0$ to indicate the presence of default risk at all times. While the probability $S(t)$ of the bond being in the "not in default" rating level declines from 1 at time $t=0$ to zero as t goes to infinity, the probability of the bond being in default - $1-S(t)$- moves in the opposite manner from zero at $t=0$ to 1 as t goes to infinity. The default state is "absorbing" in the sense that a bond remains in default once it reaches default.

Mathematics of Rating Transitions: Full Rating Scale

The purpose of our explaining the well-known hazard rate model of equations (1) and (2) for the bond default process is to show motivation for a similar "hazard rate matrix" approach for credit rating transitions. Instead of beginning with a bond that is simply "in default" or "not in default," we specify that the bond has a designated probability of being in one of the eight rating categories (AAA, AA, A, BBB, BB, B, CCC, or D). Let these eight probabilities sum to one and be written as the column vector $\vec{r}$. This vector has components r_i with $i=1,\cdots,8$ designating the eight rating levels. If one wishes to specify, for example, that the current bond rating is definitely BBB, then one would set $r_4=1$ and all other r_i to zero. More generally, let us say there are N rating categories with category N being "default."

Analogous to equation (1), the rating probability vector $\vec{r}(t)$ evolves in time as

$$\frac{d\vec{r}}{dt} = \vec{r}'(t) = -Q\,\vec{r} \tag{3}$$

where $Q(t)$ is the time-dependent hazard rate matrix (HRM) for rating transitions. In the limit as $\Delta t \rightarrow 0$, the HRM Q has the meaning that $I-Q\Delta t$ is the rating transition matrix for the time period Δt where I is the identity matrix. More specifically, the ij element of $I-Q\Delta t$ is the probability that a bond in rating level j will transition to rating level i in the time period Δt. We specify a negative sign in equation (3) to maintain consonance with equation (1). As a result, the components Q_{ij} of Q satisfy the following:

$$Q_{iN} = 0 \ , \ \ i = 1, \cdots, N \quad \text{(4a)}$$

$$Q_{ii} > 0 \ , \ \ i = 1, \cdots, N-1 \quad \text{(4b)}$$

$$Q_{ij} < 0 \ , \ \ i = 1, \cdots, N; \ \ j = 1, \cdots, N-1; \ \ \text{and } i \neq j \quad \text{(4c)}$$

$$\sum_{i=1}^{N} Q_{ij} = 0 \ , \ \ j = 1, \cdots, N \quad . \quad \text{(4d)}$$

The special designation in (4a) for rating category N reflects the condition that a bond in default cannot experience a rating transition to leave this default state. Equation (4d) imposes the requirement that the sum of probabilities of all rating levels must be constant (at the value 1). Thus, equations (4a-d) show us that Q is an N x N matrix in which column N is filled with zeroes, all other diagonal elements are positive, all other off-diagonal elements are negative and the sum of the elements in each column is zero.

Just as we can solve the scalar hazard rate equation (1) to get the survival time $S(t)$ of equation (2), the solution to equation (3) is

$$\vec{r}(t) \ = \ exp(-Qt)\,\vec{r_o} \quad (5)$$

for the special case in which Q has no time dependence. In equation (5), the argument of the exponential is an N x N matrix, with "$-t$" multiplying every component of Q.[11] The initial rating probability vector is $\vec{r_o}$. For a time-dependent HRM $Q(t)$, it is not possible to determine a solution similar to equation (5), unless the time dependence is of the form of a constant matrix A multiplied by an arbitrary scalar function $g(t)$. Thus, with $Q(t)=Ag(t)$, the time-dependent rating probability vector $\vec{r}(t)$ is

$$\vec{r}(t) \ = \ exp\left[-A\int_0^t du\, g(u)\right]\vec{r_o} \quad (6)$$

Unless stated otherwise for the remainder of this study, we take the HRM Q to be constant, so that equation (5) is the relevant expression for the rating probability vector. It would be worthwhile to employ a time-dependent $Q(t)$ in modeling exercises for which one desires a rating volatility that increases or decreases with a prescribed $g(t)$.

Standard Rating Transition Matrix

Equation (5) permits us to write the RTM for any future time t as $exp(-Qt)$. The most typical period is one year, so the one-year RTM would be simply $exp(-Q)$. Many previous studies of rating transition dynamics have treated this one-year RTM as the analytical starting point and then created a "generator matrix" similar to the HRM Q of this article.[12]

The primary difference between these earlier generator matrices and our HRM is our interpretation of Q as the agent of time evolution of the rating probability vector in equation (3).[13] We consider the HRM to be the guiding concept and relegate the RTM to an ancillary role.

Our HRM differs from the generator matrices of past studies in less significant terms as well. The HRM has a preceding negative algebraic sign for consistency with the scalar default hazard rate. To accommodate rating probability column vectors in equations (3), (5) and (6), our hazard rate matrix indices are transposed from the convention of past studies. That is, the component Q_{ij} carries information on the probability of rating migration from rating level j to rating level i rather than from rating level i to rating level j.

Past research projects have applied a hazard rate matrix approach to problems other than the time evolution of credit ratings. Takada and Sumita (2011) studied financial default risk based on obligor industry and macro-economic considerations. Singer and Spilerman (1976) investigated Markov processes relevant to sociology topics such as immigration and juvenile delinquency recidivism. Keilson and Kester (1974) provided mathematical background and properties for hazard rate matrices in Markov processes.

Necessary Properties of the Hazard Rate Matrix

We list here and discuss briefly a collection of properties of the HRM Q in addition to those of equations (4a) – (4d).

Q has a zero eigenvalue

Since column N of Q has all zero entries, this matrix is singular. All singular matrices have at least one zero eigenvalue. One determines the corresponding eigenvector, with little effort, to be the column vector with all zero entries, save for the N^{th} element, which we set to 1. Writing λ_N and $\vec{v}_N$ as this eigenvalue and eigenvector, respectively, these statements become

$$Q\,\vec{v}_N \ = \ \lambda_N\vec{v}_N \ \text{ with } \lambda_N = 0 \ \text{ and } \ \vec{v}_N \ = \ \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \quad (7)$$

All other eigenvalues of Q are real and positive

For all other eigenvalues λ_i and eigenvectors $\vec{v_i}$ with $i=1,\cdots,N-1$, it would be highly desirable if the eigenvalues were real and positive. While the matrix Q is real, it is not symmetric. It is in principle possible for some eigenvalues to be complex even with the properties of equations (4a) – (4d).[14] Therefore, we impose the constraint on the elements Q_{ij} of Q that all eigenvalues be real. Our motivation for this requirement is the behavior of solutions of equation (5), which we discuss later.

Considering the sub-matrix of Q that excludes the N^{th} row and column, Gerschgorin's Circle Theorem (GCT) and equations (4b) - (4d) are sufficient to show that all eigenvalues of the sub-matrix are greater than zero (i.e., $\lambda_i > 0$).[15] Note that these "eigenvalues of the sub-matrix" are simply the remaining eigenvalues λ_i with $i=1,\cdots,N\text{-}1$ of Q. Equations (4b) – (4d) are sufficient support for the assertion that the sub-matrix is strictly diagonally dominant. Since all diagonal elements are positive, the radii of the Gerschgorin circles of the GCT do not permit zero or negative eigenvalues.

For convenience, we number the eigenvalues from largest to smallest. Incorporating the possibility that two or more eigenvalues may have the same value, we write

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_{N-1} > \lambda_N = 0 \tag{8}$$

The matrix Q has a set of N linearly independent eigenvectors

We've already shown in equation (7) the eigenvector $\vec{\upsilon}_N$ corresponding to the eigenvalue $\lambda_N=0$. While Q has N–1 additional eigenvalues that are real and positive, as assumed or proven earlier, it is possible that Q does not have a full set of N linearly independent eigenvectors.

If the N–1 positive eigenvalues of equation (8) are distinct, then Q would have this complete set of eigenvectors. But there is no assurance of distinct eigenvalues. As with the earlier concern for complex eigenvalues, we add a constraint on the permitted elements Q_{ij} of Q that there exists a set of N linearly independent eigenvectors $\vec{\upsilon}_i$ with $i=1,\cdots,N$. With this condition, the matrix Q is diagonalizable.

The elements of each of the first N-1 eigenvectors sum to zero

For all eigenvalues and eigenvectors, we can write $Q\vec{\upsilon}_i = \lambda_i\vec{\upsilon}_i$ with $i=1,\cdots,N$. Excluding the last eigenvalue/eigenvector pair, we know $\lambda_i \neq 0$. We define the row vector $\vec{y}^T$ to have rank N with all elements equal to 1. Pre-multiplying the HRM Q by $\vec{y}^T$ gives another rank N row vector in which all elements are zero (because the sum of each column of Q is zero). This observation means that $\lambda_i \vec{y}^T \vec{\upsilon}_i = 0$. The pre-multiplication of $\vec{\upsilon}_i$ by $\vec{y}^T$ simply sums the elements of $\vec{\upsilon}_i$. Since $\lambda_i \neq 0$, the sum of the elements of each eigenvector $\vec{\upsilon}_i$ must be zero. Note that this zero-sum condition does not apply to the last eigenvector $\vec{\upsilon}_N$, since $\lambda_N=0$.

The eigenvalues and eigenvectors of the rating transition matrix are readily determined

We observed earlier that the one-year RTM is $exp(-Q)$ and the more general RTM for time t is $exp(-Qt)$. Working with the latter, this RTM is a matrix with the infinite series representation

$$exp(-Qt) = I - tQ + \frac{t^2}{2}Q^2 - \frac{t^3}{6}Q^3 + \cdots \tag{9}$$

We've written the scalar t to the left of the terms with powers of the HRM Q. The notation Q^k simply means that the matrix Q is multiplied by itself k times. Since $Q\vec{\upsilon}_i = \lambda_i\vec{\upsilon}_i$ implies $Q^k\vec{\upsilon}_i = \lambda_i^k\vec{\upsilon}_i$, we see that $\vec{\upsilon}_i$ is also an eigenvector of $exp(-Qt)$, with

$$\begin{aligned} exp(-Qt)\,\vec{v}_i &= \left(I - tQ + \frac{t^2}{2}Q^2 - \frac{t^3}{6}Q^3 + \cdots\right)\vec{v}_i \\ &= \left[1 - \lambda_i t + \frac{(\lambda_i t)^2}{2} - \frac{(\lambda_i t)^3}{6} + \cdots\right]\vec{v}_i \\ &= e^{-\lambda_i t}\,\vec{v}_i \end{aligned} \tag{10}$$

Thus, the eigenvectors of the RTM $exp(-Qt)$ are simply the eigenvectors of the HRM Q. In terms of the eigenvalues of the HRM, the RTM eigenvalues are $e^{-\lambda_i t}$.

The rating probability vector $\vec{r}(t)$ is a linear combination of eigenvectors

Since the matrix Q has a set of N linearly independent eigenvectors $\vec{\upsilon}_i$ with $i=1,\cdots,N$, we can write the rank N rating probability vector $\vec{r}(t)$ as a time-dependent, weighted sum of the eigenvectors:

$$\vec{r}(t) = \sum_{i=1}^{N} \alpha_i(t)\,\vec{v}_i \tag{11}$$

In equation (11), the coefficients α_i are time-dependent while the eigenvectors $\vec{\upsilon}_i$ are constant. Applying equation (3),

$$\begin{aligned} &\vec{r}' = -Q\vec{r} \Rightarrow \sum_{i=1}^{N}\alpha_i'\,\vec{v}_i = -\sum_{i=1}^{N}\alpha_i\,Q\vec{v}_i = -\sum_{i=1}^{N}\lambda_i\,\alpha_i\,\vec{v}_i \\ &\Rightarrow \sum_{i=1}^{N}(\alpha_i' + \lambda_i\alpha_i)\,\vec{v}_i = 0 \Rightarrow \alpha_i(t) = \alpha_i^o\,e^{-\lambda_i t} \end{aligned} \tag{12}$$

We use the notation α_i^o to denote the (time-independent) initial value of α_i. With equations (11) and (12), the probability vector $\vec{r}(t)$ becomes

$$\vec{r}(t) = \sum_{i=1}^{N} \alpha_i{}^o \, e^{-\lambda_i t} \, \vec{v}_i \qquad (13a)$$

The initial rating probability vector $r_o \equiv r(0)$ is

$$\vec{r}_o = \sum_{i=1}^{N} \alpha_i{}^o \, \vec{v}_i \qquad (13b)$$

Since the eigenvectors υ_i are known, we use equation (13b) to determine the $\alpha_i{}^o$ given the choice of $\vec{r}_o$.

The probability character of the vector $\vec{r}(t)$ is preserved

To interpret the vector $\vec{r}$(t) as a "rating probability" vector, all its elements must be non-negative and sum to one. Clearly, one would choose an initial probability vector $\vec{r}_o$ that satisfies these conditions. From equation (13b), we see requiring the sum of the elements of $\vec{r}_o$ to be one implies that $\alpha_N{}^o=1$.[16] With this assignment, equation (13a) then shows that the sum of elements of $\vec{r}$(t) for all $t>0$ is also one (since $\lambda_N=0$).

To show that all elements of $\vec{r}$(t) remain non-negative provided they are non-negative at time t, consider $\vec{r}(t+\Delta t)$. For sufficiently small Δt, $\vec{r}(t+\Delta t) \approx [I-Q\Delta t]\vec{r}(t)$. Given that Q has the properties (4a)-(4d) and that the elements of $\vec{r}$(t) are non-negative, the observation that all elements of the matrix $I-Q\Delta t$ are positive implies that all elements of $\vec{r}(t+\Delta t)$ are non-negative. Applying these time steps Δt repeatedly shows that all elements of $\vec{r}$(t) remain non-negative as time t increases.

It is worth noting that this property of persistent non-negativity for the elements of $\vec{r}$(t) does not arise from a non-negativity property of the elements of the eigenvectors $\vec{\upsilon}_i$ $(i=1,\cdots,N)$ of the HRM Q. Other than the eigenvector $\vec{\upsilon}_N$ corresponding to the zero eigenvalue, all the $\vec{\upsilon}_i$ will have at least one negative element (given that the sum of vector elements is zero and the elements are not all zero).

Asymptotic Behavior of Credit Ratings for Large Time

Just as we know that the survival probability $S(t)$ for the simple hazard rate model of equation (2) must approach zero as time t goes to infinity, we have a similar expectation for the rating probability vector $\vec{r}$(t) as $t \to \infty$.

Regardless of the initial rating state $\vec{r}_o$, all bonds eventually default. Hence, all probability will accumulate in the default state. The probability vector will approach the column vector with one as the N^{th} component and all other elements zero. From equation (7), this is also the N^{th} eigenvector $\vec{\upsilon}_N$. Symbolically, this is depicted as follows:

$$\vec{r}(t) \rightarrow \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} = \vec{v}_N \text{ as } t \rightarrow \infty \qquad (14)$$

By inspection of equation (13a), we can improve the equation (14) asymptotic expression by retaining the next largest term:

$$\vec{r}(t) \sim \vec{v}_N + \alpha_{N-1}{}^o \, e^{-\lambda_{N-1} t} \, \vec{v}_{N-1} \text{ as } t \rightarrow \infty \qquad (15)$$

The ordering of eigenvalues of equation (8) motivates this expression. The smallest positive eigenvalue is λ_{N-1} and we designate this λ_{N-1} as the "penultimate eigenvalue." We consider λ_N (=0) to be the "last eigenvalue" and make the assumption that the antepenultimate eigenvalue, λ_{N-2}, is strictly greater than the penultimate eigenvalue. All other terms of equation (13a) are exponentially smaller than the two terms of equation (15) as time t goes to infinity.

A striking feature of equation (15) is that the "penultimate eigenvector" $\vec{\upsilon}_{N-1}$ provides a fixed distribution of rating probabilities above the default state (N). That is, imagine that the probability of each rating level above D is specified as being proportional to each corresponding element of $\vec{\upsilon}_{N-1}$. Then, as time increases, the relative probability "amount" of each rating above D remains the same even though the absolute probabilities of each rating are diminishing with the $e^{-\lambda_{N-1}t}$ exponential of equation (15).

If the financial world deliberately added new bonds to replace bonds that default, then one could interpret the penultimate eigenvector as providing a steady-state rating distribution above D. In reality, there is no replacement mechanism of this type, so we will think of $\vec{\upsilon}_{N-1}$ as providing merely a "natural rating distribution." Regardless of the initial rating state, the rating probability vector will evolve over time to the shape of this "natural rating distribution."

Given this meaning of $\vec{\upsilon}_{N-1}$, we conclude that each of the first N-1 elements of this penultimate eigenvector must either have the same algebraic sign or be zero. To add clarity, these N-1 elements cannot consist of both positive and negative entries. This statement is not an additional requirement to place on the HRM. Rather, the properties we developed earlier are sufficient to prove this behavior.

Since $\vec{r}$(t) maintains its character as a probability vector, then it must do so in the asymptotic limit expressed in equation (15). If the first N-1 elements of $\vec{\upsilon}_{N-1}$ are non-negative, (non-positive), then the constant $\alpha_{N\text{-}1}{}^{o}$ is positive (negative). One can also show that the N^{th} element of the penultimate eigenvector must have the opposite algebraic sign to that of the first N-1 elements.

Equations (13a) and (15) also show why complex eigenvalues are undesirable. Earlier, we announced a restriction on the HRM to ensure that all eigenvalues are real. In the absence of this restriction, complex eigenvalues would be possible and they would occur in complex conjugate pairs.

If the penultimate eigenvalue $\lambda_{N\text{-}1}$ were one of a complex conjugate pair, then we would need to include its conjugate, the antepenultimate eigenvalue $\lambda_{N\text{-}2}$, in the asymptotic expression of equation (15), since both $\lambda_{N\text{-}1}$ and $\lambda_{N\text{-}2}$ would have the same real part. This inclusion is certainly feasible mathematically, but it produces an asymptotic rating probability vector $\vec{r}(t)$ that oscillates like a cosine function. Such oscillation is not permitted within the context of rating transition to default, so this is the basis for our forbidding complex eigenvalues. We should add that our "anti-complexity argument" strictly applies only to the smallest non-zero eigenvalues $\lambda_{N\text{-}1}$ and $\lambda_{N\text{-}2}$, but intuition suggests a ban on all complex eigenvalues is most reasonable.

Practical Result - Examples

As we related earlier, we do not gather empirical data from the CRAs to test this hazard rate matrix model for rating transitions since rating agency downgrades and upgrades are markedly non-Markovian. Rather, we create and study an idealized rating transition model. As a first example, consider the HRM of Figure 1 (below) with all values rounded to the nearest 0.0001.

Figure 1: Example of Hazard Rate Matrix with *N* = 8

0.1500	(0.0665)	(0.0334)	(0.0181)	(0.0099)	(0.0057)	(0.0022)	0.0000
(0.0761)	0.2100	(0.0668)	(0.0361)	(0.0198)	(0.0115)	(0.0044)	0.0000
(0.0381)	(0.0739)	0.2400	(0.0723)	(0.0395)	(0.0230)	(0.0089)	0.0000
(0.0190)	(0.0370)	(0.0743)	0.2700	(0.0791)	(0.0459)	(0.0178)	0.0000
(0.0095)	(0.0185)	(0.0371)	(0.0803)	0.3000	(0.0918)	(0.0356)	0.0000
(0.0048)	(0.0092)	(0.0186)	(0.0402)	(0.0878)	0.3300	(0.0711)	0.0000
(0.0024)	(0.0046)	(0.0093)	(0.0201)	(0.0439)	(0.1021)	0.3600	0.0000
(0.0001)	(0.0002)	(0.0005)	(0.0030)	(0.0200)	(0.0500)	(0.2200)	0.0000

With $N = 8$, the rating categories are {AAA, AA, A, BBB, BB, B, CCC, D}. Each off-diagonal entry is related to a probability of transition from the column rating state to the row rating state.[17] Hence, since it is not possible to transition out of the default state D, the last column is filled with zeroes. The bottom row consists of default hazard rates, because the values connote transition into the default state D.

To create the example of Figure 1, we first conjured the diagonal entries in a manner that is loosely consistent with values of Moody's (2011). We imposed the behavior of the diagonal value monotonically increasing from the highest (AAA) category to the lowest (CCC) category above default, because the CRAs often describe their ratings as being less volatile in the higher categories. For off-diagonal entries, we imposed plausible default hazard values in the last row, geometric decreases in absolute values moving up and down from the diagonal, and the zero-sum condition for each column (equation (4d)).

To compute eigenvalues and eigenvectors of the Figure 1 and other matrices, we apply the methods and algorithms of Press, Flannery, Teukolsky and Vetterling (1992).[18] For this first example, we list the eigenvalues and associated eigenvectors in Figure 2 (below) with all values rounded to the nearest 0.001.

Figure 2: Eigenvalues and Eigenvectors for the HRM of Figure 1

		1	2	3	4	5	6	7	8
Eigenvalues		0.440	0.384	0.335	0.293	0.238	0.148	0.021	0.000
Eigenvector Components	1	(0.000)	(0.000)	0.001	(0.057)	0.175	(0.210)	0.092	0.000
	2	(0.000)	(0.001)	(0.048)	0.237	(0.169)	(0.086)	0.095	0.000
	3	(0.001)	(0.026)	0.219	(0.185)	(0.169)	0.031	0.092	0.000
	4	(0.009)	0.184	(0.256)	(0.142)	(0.025)	0.111	0.080	0.000
	5	0.137	(0.326)	(0.072)	0.025	0.089	0.139	0.064	0.000
	6	(0.341)	0.042	0.105	0.110	0.121	0.123	0.046	0.000
	7	0.363	0.274	0.174	0.128	0.114	0.096	0.031	0.000
	8	(0.149)	(0.147)	(0.124)	(0.115)	(0.138)	(0.204)	(0.500)	1.000

Since an eigenvector multiplied by any non-zero constant remains an eigenvector of the same eigenvalue, we are free to choose a normalizing condition. In Figure 2 and other representations of eigenvectors, our normalization is that the sum of the absolute values of the elements is one and that the last element of each eigenvector is negative (with the exception of the last eigenvector).

Consistent with equation (8), we've ordered the eigenvalues in Figure 2 from largest to smallest. The last and smallest eigenvalue λ_8=0. The eigenvector for this last eigenvalue, $\vec{v}_8$, is shown in the last column of Figure 2 to consist of all zeroes for the first seven elements and 1 for the last element, as we expect.

The penultimate eigenvalue is λ_7=0.021 and the penultimate eigenvector lies in the column beneath this eigenvalue. The components of the penultimate eigenvector above the D rating represent the "natural rating distribution" for the HRM of Figure 1. These seven components, which begin with 0.092 and 0.095 and end with 0.031, show this natural distribution is primarily (and somewhat uniformly) in the AAA, AA and A ratings, and then trails off at lower ratings.

Based on equations (13a) and (13b), a group of bonds that all have initial AA ratings will have a rating probability vector that we write as a linear combination of the eigenvectors of the HRM. Assuming the HRM of Figure 1, the eigenvalues of the top row of Figure 2 give the rate of exponential decay of each eigenvector component. Since the penultimate eigenvalue λ_7=0.021, the "time constant" for decay of the penultimate eigenvector component (i.e., the "natural rating distribution component") is roughly 48 years (the inverse of 0.021). Thus, this component is long-lived.

The next largest (antepenultimate) eigenvalue is λ_6=0.148, which implies a much shorter time constant of just about seven years. Hence, after roughly seven years, the rating distribution of the remaining, undefaulted bonds that were initially rated AA will be approximately proportional to the natural rating distribution (components of the penultimate eigenvector of Figure 2).

As a second example, consider the alternative HRM of Figure 3 (below), with all values rounded to the nearest 0.0001.

Figure 3: Example of Hazard Rate Matrix with *N* = 8

0.0400	(0.0316)	(0.0212)	(0.0115)	(0.0087)	(0.0048)	(0.0040)	0.0000
(0.0203)	0.0900	(0.0425)	(0.0230)	(0.0175)	(0.0095)	(0.0079)	0.0000
(0.0101)	(0.0301)	0.1400	(0.0460)	(0.0350)	(0.0190)	(0.0159)	0.0000
(0.0051)	(0.0150)	(0.0404)	0.1600	(0.0699)	(0.0381)	(0.0317)	0.0000
(0.0025)	(0.0075)	(0.0202)	(0.0438)	0.2400	(0.0761)	(0.0635)	0.0000
(0.0013)	(0.0038)	(0.0101)	(0.0219)	(0.0666)	0.2600	(0.1270)	0.0000
(0.0006)	(0.0019)	(0.0051)	(0.0109)	(0.0333)	(0.0725)	0.4000	0.0000
(0.0001)	(0.0002)	(0.0005)	(0.0030)	(0.0090)	(0.0400)	(0.1500)	0.0000

Whereas the HRM of Figure 1 is an approximate, idealized version of data from Moody's (2011), Figure 3 is an idealized and approximate rendering of data in Standard & Poor's (2008). The diagonal values differ markedly between the two HRM examples.[19]

For this second example, we list the eigenvalues and associated eigenvectors in Figure 4, with all values rounded to the nearest 0.001.

Figure 4: Eigenvalues and Eigenvectors for the HRM of Figure 2

		8	7	6	5	4	3	2	1
Eigenvalues		0.449	0.314	0.214	0.173	0.113	0.061	0.006	0.000
Eigenvectors	1	(0.000)	0.001	(0.001)	0.013	0.134	(0.373)	0.185	0.000
Components	2	(0.000)	0.004	(0.007)	0.159	(0.412)	0.086	0.110	0.000
	3	(0.001)	0.020	0.148	(0.422)	0.019	0.115	0.074	0.000
	4	(0.004)	0.103	(0.412)	0.043	0.132	0.130	0.061	0.000
	5	(0.033)	(0.434)	0.106	0.110	0.095	0.079	0.035	0.000
	6	(0.323)	0.285	0.178	0.126	0.085	0.063	0.025	0.000
	7	0.500	0.087	0.068	0.050	0.035	0.027	0.011	0.000
	8	(0.137)	(0.066)	(0.080)	(0.078)	(0.088)	(0.127)	(0.500)	1.000

Our qualitative observations for the eigenvalues and eigenvectors of Figure 2 (the first example) also hold for this second example. The last and penultimate eigenvectors are of the form we expect. Here, the penultimate and antepenultimate eigenvalues are smaller than those of the first example, which implies the time constants for decay of eigenvector components to the initial rating probability vector will be longer. The natural rating distribution (proportional to penultimate eigenvector) is more skewed to AAA in this second example which may be due to the small AAA diagonal element (0.04) of the HRM of Figure 3.

Finally, we present a third example in which the HRM, like that of Figure 1, is motivated by the Moody's (2011) rating transition data. Now we add in the customary sub-categories in which all letter ratings from AA down to CCC have three distinct rating levels (described by "+" and "-" signs or by numerical modifiers or by "high" and "low" designations).

Instead of N=8, this example has N=20 rating levels. Rather than showing all 400 elements of this HRM, we describe it as Figure 1, with interpolation along the bottom row (transition to default) and along the diagonals with one modification: the diagonal elements need to increase by at least 50% relative to the Figure 1 diagonal values. This increase in the diagonal elements is sensible, because increasing the number of rating levels should necessarily increase the number of rating transitions. If we do not increase the diagonal values in this manner, then the HRM will have complex pairs of eigenvalues.

This HRM for 20 rating levels has 20 eigenvalues and 20 corresponding eigenvectors. The eigenvalues are all real and distinct, and range from a high of 0.662 for λ_1 down to 0.013 and 0.000 (zero) for λ_{19} and λ_{20}, respectively. Figure 5 (below) shows the largest two eigenvalues with associated eigenvectors and the smallest three eigenvalues with eigenvectors (with all values rounded to the nearest 0.001).

Figure 5: Five of the Eigenvalues and Eigenvectors for the HRM with *N* = 20

		20	19	3	2	1
Eigenvalues		0.662	0.632	0.080	0.013	0.000
Eigenvectors	1	(0.000)	0.000	(0.050)	0.035	0.000
Components	2	(0.000)	0.000	(0.056)	0.044	0.000
	3	(0.000)	0.000	(0.045)	0.045	0.000
	4	(0.000)	0.000	(0.032)	0.045	0.000
	5	(0.000)	0.000	(0.015)	0.044	0.000
	6	(0.000)	0.000	0.003	0.042	0.000
	7	(0.000)	0.000	0.020	0.039	0.000
	8	(0.000)	0.000	0.034	0.036	0.000
	9	(0.000)	0.000	0.045	0.032	0.000
	10	(0.000)	0.000	0.052	0.028	0.000
	11	(0.000)	(0.001)	0.056	0.025	0.000
	12	(0.000)	0.007	0.057	0.021	0.000
	13	0.002	(0.038)	0.054	0.017	0.000
	14	(0.017)	0.116	0.049	0.014	0.000
	15	0.075	(0.191)	0.042	0.011	0.000
	16	(0.196)	0.122	0.034	0.009	0.000
	17	0.310	0.103	0.026	0.006	0.000
	18	(0.222)	(0.170)	0.017	0.004	0.000
	19	0.113	0.152	0.012	0.003	0.000
	20	(0.065)	(0.101)	(0.303)	(0.500)	1.000

The two columns furthest to the right of Figure 5 show that the last and penultimate eigenvalues and eigenvectors have the properties we derived and described earlier. In particular, the penultimate eigenvector gives the shape of the natural rating distribution.

Parting Thoughts

This article considers the well-known credit rating transition dynamic and re-formulates the problem in continuous time with a hazard rate matrix. As an extension and analogy with the hazard rate model for bond default, we create the rating probability vector and show how the hazard rate matrix governs its evolution in time. We explain the mathematical properties that this matrix must satisfy (both by imposition and by proof) and also deduce analytical insights of the hazard rate matrix (HRM).

The eigenvalues and eigenvectors of this HRM provide the best view of the behavior of bond credit ratings at large time from initial rating. In particular, the penultimate eigenvector gives the shape of the "natural rating distribution." We believe this study is the first to identify and derive this rating distribution that prevails long after inception.

Our analysis applies only to credit rating migration that is Markovian. We do not attempt to fit or explain actual credit rating agency data, since rating changes of the large public rating agencies are decidedly not Markovian. Rating accuracy, however, requires Markovian evolution. As investors, banks or asset managers create their own (internal) rating scales and procedures, it is our hope that the framework, analysis and results of this article will be helpful.

Footnotes

1. The author is Principal of Maxwell Consulting (http://www.maxwell-consulting.com/) and welcomes followers on Twitter (http://www.twitter.com/JoePimbley).

2. There exist 10 to 20 CRAs worldwide that have some significant form of governmental, regulatory or market recognition. The largest four are Moody's Investors Service, Standard & Poor's, Fitch Ratings and DBRS.

3. See, for example, Moody's (2011), Standard & Poor's (2008), Fitch (2012) and DBRS (2012).

4. The published default tables are often associated with the structured finance ratings of the CRAs, with no clear statement that these tables also apply to the ratings of non-structured finance debt such as sovereigns or banks. Further, Moody's Investors Service describes its ratings in terms of "expected loss" rather than "default probability," but the conversion between these two risk measures is straightforward.

5. See, for example, Moody's (2011), Standard & Poor's (2008), Fitch (2012) and DBRS (2012).

6. Alternatively, the CRA may define its transition matrix such that the default probabilities occupy the last column rather than the last row.

7. See, for example: Israel, Rosenthal and Wei (2001), and Gupton, Finger and Bhatia (2007).

8. See, for example, Lando and Skødeberg (2002), Altman and Kao (1992), and Carty and Fons (1993).

9. See the Moody's Investors Service "Special Comment" of Mann and Metz (2011).

10. See, for example, Löffler and Posch (2011).

11. One interprets the exponential function of a matrix argument as the MacLaurin series expansion of the exponential with scalar product powers of the matrix argument.

12. See, for example: Israel, Rosenthal and Wei (2001); Lando and Skødeberg (2002); and Jarrow, Lando and Turnbull (1997).

13. Jarrow, Lando and Turnbull (1997) also wrote a "Kolmogorov equation" similar to equation (3) to relate the "generator matrix" to the time evolution of the RTM. Our work differs from this study in our focus on the HRM and the rating probability vector.

14. As an example, imagine there are only four rating states (including the default state), so that N=4. Let the HRM be $\begin{pmatrix} 0.15 & -0.17 & -0.01 & 0 \\ -0.01 & 0.20 & -0.02 & 0 \\ -0.13 & -0.01 & 0.25 & 0 \\ -0.01 & -0.02 & -0.22 & 0 \end{pmatrix}$ Of the four eigenvalues of this matrix, two are real (0 and 0.097) and two are complex (0.252 + i 0.048 and 0.252 – i 0.048). Relatively small changes to some off-diagonal matrix elements produce four real eigenvalues.

15. See, for example, Weisstein (2012). In our application of this theorem, we consider the sum of column (rather than row) element absolute values for the Gerschgorin radius. This substitution is permissible since a matrix and its transpose have the same eigenvalues.

16. Recall that the sum of all elements of each eigenvector $\vec{v}_i$ is zero for $i \neq N$ and that the sum of the elements of $\vec{v}_N$ is one from equation (7).

17. As stated previously, the HRM Q has the meaning that I – $Q\Delta t$ is the rating transition matrix for the time period Δt,where I is the identity matrix. Thus, the ij element of I – $Q\Delta t$ is the probability that a bond in rating level j will transition to rating level i in the time period Δt.

18. More specifically, we reduced the HRM to upper Hessenberg form by a series of similarity transformations. We then applied an iterative "QR algorithm" to determine the eigenvalues of the upper Hessenberg matrix (which are also the eigenvalues of the original HRM). With these eigenvalues, we next calculated all eigenvectors by inverse iteration.

19. We emphasize again that we are not attempting here to infer hazard rate matrices from rating agency data. This is inappropriate because the amount of data is insufficient to generate the precision we ascribe to our matrices and because credit ratings in practice are not Markovian. Our intent in referencing past data from Moody's Investors Service and Standard & Poor's is simply to show that the numerical values we use are similar to real-world credit rating transition data.

Joe Pimbley (PhD) is Principal of Maxwell Consulting, a consulting firm he founded in 2010, and a member of Risk Professional's editorial board. His recent and current engagements include financial risk management advisory, underwriting for structured and other financial instruments, and litigation testimony and consultation. In a prominent engagement from 2009 to 2010, Joe served as a lead investigator for the Examiner appointed by the Lehman bankruptcy court to resolve numerous issues pertaining to history's largest bankruptcy.

References

E. Altman and D. Kao. "The implications of corporate bond rating drift," *Financial Analysts Journal*, 64-75, 1992.

B. Belkin, S. Suchower and L. Forest Jr. "A one-parameter representation of credit risk and transition matrices," *Credit Metrics Monitor*, 1998.

L. Carty and J. Fons. "Measuring changes in corporate credit quality," Moody's Special Report, 1993.

J. Christensen, E. Hansen and D. Lando. "Confidence sets for continuous-time rating transition probabilities," *J. Banking & Finance* 28, 2575-2602, 2004.

DBRS Industry Study. "2011 DBRS Corporate Rating Transition and Default Study," March 2012.

Fitch Ratings Special Report. "Fitch Ratings Global Corporate Finance 2011 Transition and Default Study," March 2012.

G. M. Gupton, C. C. Finger and M. Bhatia. CreditMetrics – Technical Document, RiskMetrics Group, Inc., 2007.

R. B. Israel, J. S. Rosenthal and J. Z. Wei. "Finding generators for Markov chains via empirical transition matrices with applications to credit ratings," *Mathematical Finance* 11(2), 245-265, 2001.

R. A. Jarrow, D. Lando, and S. M. Turnbull. "A Markov model for the term structure of credit risk spreads," *Review of Financial Studies* 10, 481-523, 1997.

J. Keilson and A. Kester. "Monotone matrices and monotone Markov processes," Technical Report No. 79 for National Science Foundation Grant GP-32326X, Stanford University, 1974.

D. Lando and T. M. Skødeberg. "Analyzing rating transitions and rating drift with continuous observations," *J. Banking & Finance* 26, 423-444, 2002.

G. Löffler and P. N. Posch. *Credit risk modeling using Excel and VBA*, Second Edition, John Wiley & Sons, Ltd., 2011.

J. Mann and A. Metz. "Measuring the Performance of Credit Ratings," Moody's Investors Service Special Comment, November 2011.

Moody's Investors Service Special Comment. "Corporate Default and Recovery Rates, 1920-2010," February 2011.

W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing*, 2nd Edition, Cambridge University Press, 1992.

B. Singer and S. Spilerman. "The representation of social processes by Markov models," *American Journal of Sociology* 82, 1-54, 1976.

Standard & Poor's Ratings Direct. "2007 Annual Global Corporate Default Study and Rating Transitions," February 2008.

H. Takada and U. Sumita. "Credit risk model with contagious default dependencies affected by macro-economic condition," *European Journal of Operational Research*, 2011.

E. W. Weisstein. "Gerschgorin Circle Theorem," from MathWorld – a Wolfram Web Resource: http://mathworld.wolfram.com/GershgorinCircleTheorem.html, 2012.