Exploring the Stochastic Migration Model

Matt Reimherr

1 Introduction

How does one define risk? Risk is a very relative term which can mean something to one person and something entirely different to someone else. Therefore, to analyze risk mathematically one has to concretely define what is meant by "Risk." This paper will be dealing with the risk associated with businesses and corporations from the perspective of a bank or any potential lender. Now risk can take on a more concrete meaning. A lender only sees risk in a company based on how likely it is that the lender will not receive the return on their investment. Therefore if we look at a specific loan we can define risk as the probability that that the lender will not receive a full return on that loan to some company. This is event is equivalent to the company defaulting while the loan is still outstanding. As recommended by the Basle committee, we can express risk in terms of the probability of time till default such that the probability that company i will default before time h given that they are currently at time t can be expressed as:

$$P_i[\tau > h|\text{not at default at time } t] = \exp[e^{(\mathbf{x}_{i,t}^T b)} a(h;\theta)],$$

where $\mathbf{x}_{i,t}^T$ is the transpose of observed covariates, or the factors specific to company i that influence its risk and $a(h,\theta)$ is a baseline hazard function. This function can be thought of as accounting for the risk associated with even the perfect company. b and θ are parameters. This model is called the Proportional Hazard Model. In this case we could define a quantitative credit score, $s_{i,t}$, for company i at time t such that $s_{i,t} = x_{i,t}^T \hat{b}$, where $\hat{b}$ is the estimated value of the parameter b. Notice that this is all we would need to define a credit score for company i at time t since only $\mathbf{x}_{i,t}^T$ depends on the company i and time t.

This is a basic model for understanding the risk of a specific company, but current rating systems do no use continuous credit scores. Scores are often discretized so that individuals that have a similar amount of risk can be group together.

2 Basic Framework

Let us examine a generic credit score with K total scores with possible values $1, 2, \ldots, K$, where 1 indicates the minimum amount of risk and K is default. Now let us introduce a variable $Y_{i,t}$ to denote the credit score of the individual i at time t and $\vec{Y}_{i,t}$ to denote the entire credit history up to and including time t. Notice that credit scores are not typically updated at a continuous rate. We are examining periods of some specified length. Therefore we can assume that t only takes on integer values. $\vec{Y}_{i,t}$ is also known as the lagged credit history for company i . Therefore we can now characterize the risk associated with a company i from t to $t+1$ as:

$$P(Y_{i,t+1} = K \mid Y_{i,t} \neq K)$$

We can also examine the scores of a whole industry or group of individuals. Let us denote $\mathbf{Y}_t$ as the vector $(Y_{1,t}, \ldots, Y_{n,t})$ and $\vec{Y}_t = (\mathbf{Y}_1, \ldots, \mathbf{Y}_t)$ as the lagged ratings for the whole industry. Notice that each company has K possible scores and thus $\mathbf{Y}_t$ can have K^n different values if there are n different companies. In other words, the process $\mathbf{Y}_t$ has a state space with K^n elements.

3 Stochastic Migration Model

Let π_t be $K \times K$ random matrices which form a Markov process. We also assume that for every fixed t all elements of π_t are between 0 and 1 and the sum of the elements in each row must be 1. This means that each realization of π_t is a transition matrix. If the sequence $\boldsymbol{\pi} = (\pi_t, 1 \leq t < \infty)$ is known we use $P_{\boldsymbol{\pi}}$ and $E_{\boldsymbol{\pi}}$ to denote the corresponding probability and expected value. This means that $P_{\boldsymbol{\pi}}$ and $E_{\boldsymbol{\pi}}$ are the conditional probability and expected value given $\pi_t, 1 \leq t < \infty$. Let $\pi_t(j,k)$ denote the (j,k) element of π_t.

Definition 3.1. *The individual rating histories will satisfy a Stochastic Migration Model if $Y_{1,t}, \ldots, Y_{n,t}$ are stochastic processes satisfying:*
(i) $P_{\boldsymbol{\pi}}(Y_{i,t} = k | Y_{i,t-1} = j) = \pi_t(j,k)$ for all $1 \leq i \leq n, 1 \leq j,k \leq K$ and $1 \leq t < \infty$
and
(ii) for all sets $C_1, \ldots, C_n$ we have

$$\begin{aligned} &P_{\boldsymbol{\pi}}(\{Y_{1,t}, 1 \leq t < \infty\} \in C_1, \ldots, \{Y_{n,t}, 1 \leq t < \infty\} \in C_n) \\ &\quad = P_{\boldsymbol{\pi}}(\{Y_{1,t}, 1 \leq t < \infty\} \in C_1) \cdots P_{\boldsymbol{\pi}}(\{Y_{n,t}, 1 \leq t < \infty\} \in C_n). \end{aligned}$$

Sets $C_1, \ldots, C_n$ are vectors of infinite length whose elements take on values in $1, \ldots, K$. Conditions (i) and (ii) mean that if the transition matrices are given then $Y_{i,t}, 1 \leq i \leq n$ are independent identically distributed markov chains.

We should take note that the individuals are independent if the sequence $\pi_t, 1 \leq t < \infty$ is known and dependant otherwise. Let us denote the transition matrix for $\mathbf{Y}_t$ as Π_t. If we know the whole lagged history and the whole sequence π_t for all t then $Y_{1,t}, \ldots, Y_{n,t}$ independant and identically distributed markov chains. Therefore we can easily develop Π_t for any t. The transition matrix for Y_t will be characterized by:

$$\begin{aligned} P(\mathbf{Y}_{t+1} = \mathbf{k} \mid \vec{Y}_t, \boldsymbol{\pi}) &= P_{\boldsymbol{\pi}}(\mathbf{Y}_{t+1} = \mathbf{k} \mid \vec{Y}_t) \\ &= P_{\boldsymbol{\pi}}(Y_{1,t+1} = k_1, \ldots, Y_{n,t+1} = k_n \mid \vec{Y}_t) \\ &= P_{\boldsymbol{\pi}}(Y_{1,t+1} = k_1 \mid \vec{Y}_t) \cdots P_{\boldsymbol{\pi}}(Y_{n,t+1} = k_n \mid \vec{Y}_t), \end{aligned}$$

where the last equality holds because of (ii). Here $\mathbf{k} = (k_1, \ldots, k_n)$ where $k_1, \ldots, k_n$ are taking values in $1, \ldots, K$. Π_t can be written as:

$$\Pi_t = \begin{pmatrix} [\pi_t(1,1)]^n & \ldots & [\pi_t(1,n)]^n \\ \vdots & \ddots & \vdots \\ [\pi_t(n,1)]^n & \ldots & [\pi_t(n,n)]^n \end{pmatrix}$$

Thus with appropriately ordered states Π_t can factored as:

$$\begin{aligned}
\Pi_t &= \begin{pmatrix} [\pi_t(1,1)]^{n-1}\pi_t & \ldots & [\pi_t(1,n)]^{n-1}\pi_t \\ \vdots & \ddots & \vdots \\ [\pi_t(n,1)]^{n-1}\pi_t & \ldots & [\pi_t(n,n)]^{n-1}\pi_t \end{pmatrix} \\
&= \begin{pmatrix} [\pi_t(1,1)]^{n-1} & \ldots & [\pi_t(1,n)]^{n-1} \\ \vdots & \ddots & \vdots \\ [\pi_t(n,1)]^{n-1} & \ldots & [\pi_t(n,n)]^{n-1} \end{pmatrix} \otimes \pi_t \\
&= \cdots = \pi_t \otimes \cdots \otimes \pi_t \\
&= \otimes^n \pi_t,
\end{aligned}$$

$\otimes$ denotes the Kronecker product and $\otimes^n$ denotes the n-fold Kronecker product. The h–step probabilities for $\mathbf{Y}_t$ can be expressed in terms of the h–step transtion matrices of $Y_{1,t}, \ldots, Y_{n,t}$.

$$\begin{aligned}
&P_{\boldsymbol{\pi}}(\mathbf{Y}_{t+h} = \mathbf{k} \mid \vec{Y}_t) \\
&\quad = P_{\boldsymbol{\pi}}(Y_{1,t+h} = k_1, \ldots, Y_{n,t+h} = k_n \mid \vec{Y}_t) \\
&\quad = P_{\boldsymbol{\pi}}(Y_{1,t+h} = k_1 \mid \vec{Y}_t) \ldots P(Y_{n,t+h} = k_n \mid \vec{Y}_t).
\end{aligned}$$

Therefore Π_t^h can be expressed as:

$$\Pi_t^h = \otimes^n \pi_t^h,$$

where Π_t^h is the h–step transtion matrix from t to $t+h$ for $\mathbf{Y}_t$ and π_t^h is the h-step transition matrix from t to $t+h$ for $Y_{i,t}, i = 1, \ldots, n$. Both the single and h-step transition matrices are derived using the independence and known transition probabilities of the $Y_{i,t}$ chains.

Now suppose the sequence π_t is unknown for all t. Now π_t is a Markov process and $Y_{1,t}, \ldots, Y_{n,t}$ are dependent processes. Assembling Π_t becomes more complicated at this point. Now we must know something about the underlying distribution driving the matrices π_t. If we know something about the underlying distribution of π_t then we can integrate them out of our probabilities for $\mathbf{Y}_t$ by taking conditional expected values. Let us start by proving a lemma that will be used throughout the rest of this paper.

Lemma 3.1. *Suppose that* $\mathbf{A}$ *is an event dependent upon a vector* $\mathbf{B}$ *and a matrix* $\mathbf{C}$*. Then* $E[\mathbf{A} \mid \mathbf{B}] = E(E[\mathbf{A} \mid \mathbf{B}, \mathbf{C}] \mid \mathbf{B})$.

Proof. Here we will offer a proof for when $\mathbf{A}, \mathbf{B}, \mathbf{C}$ take on discrete values. Let us start by simplifying $E[\mathbf{A} \mid \mathbf{B}, \mathbf{C}]$.

$$E[\mathbf{A} \mid \mathbf{B}, \mathbf{C}] = \sum_{\mathbf{A}} A \cdot P(\mathbf{A} = A \mid \mathbf{B}, \mathbf{C}).$$

Now $E(E[\mathbf{A} \mid \mathbf{B}, \mathbf{C}] \mid \mathbf{B})$ can be written as:

$$\begin{aligned}
&E(E[\mathbf{A} \mid \mathbf{B}, \mathbf{C}] \mid \mathbf{B}) = E[\sum_{\mathbf{A}} A \cdot P(\mathbf{A} = A \mid \mathbf{B}, \mathbf{C}) \mid \mathbf{B}] \\
&= \sum_{\mathbf{C}} \sum_{\mathbf{A}} A \cdot P(\mathbf{A} = A \mid \mathbf{B}, \mathbf{C}) P(\mathbf{C} = C \mid \mathbf{B}) \\
&= \sum_{\mathbf{C}} \sum_{\mathbf{A}} A \cdot \frac{P(\mathbf{A} = A, \mathbf{B} = B, \mathbf{C} = C)}{P(\mathbf{B} = B, \mathbf{C} = C)} \frac{P(\mathbf{C} = C, \mathbf{B} = B)}{P(\mathbf{B} = B)} \\
&= \sum_{\mathbf{C}} \sum_{\mathbf{A}} A \cdot \frac{P(\mathbf{A} = A, \mathbf{B} = B, \mathbf{C} = C)}{P(\mathbf{B} = B)} \\
&= \sum_{\mathbf{A}} \frac{A}{P(\mathbf{B} = B)} \sum_{\mathbf{C}} P(\mathbf{A} = A, \mathbf{B} = B, \mathbf{C} = C) \\
&= \sum_{\mathbf{A}} \frac{A}{P(\mathbf{B} = B)} P(\mathbf{A} = A, \mathbf{B} = B) \\
&= \sum_{\mathbf{A}} A \cdot P(\mathbf{A} = A \mid \mathbf{B} = B) \\
&= E[A \mid B]
\end{aligned}$$

□

Now using Lemma 3.1, Π_t will be characterized by:

$$\begin{aligned}
&P(\mathbf{Y}_{t+1} = \mathbf{k} \mid \vec{Y}_t) \\
=&E[P(\mathbf{Y}_{t+1} = \mathbf{k} \mid \vec{Y}_t, \boldsymbol{\pi}) \mid \vec{Y}_t] \\
=&E[P_{\boldsymbol{\pi}}(\mathbf{Y}_{t+1} = \mathbf{k} \mid \vec{Y}_t) \mid \vec{Y}_t] \\
=&E[P_{\boldsymbol{\pi}}(Y_{1,t+1} = k_1 \mid \vec{Y}_t) \dots P_{\boldsymbol{\pi}}(Y_{n,t+1} = k_n \mid \vec{Y}_t) \mid \vec{Y}_t]. \qquad (3.1)
\end{aligned}$$

We can also derive Π_t^h using Lemma 3.1

$$\begin{aligned}
&P(\mathbf{Y}_{t+h} = \mathbf{k} \mid \vec{Y}_t) \\
=&E[P(\mathbf{Y}_{t+h} = \mathbf{k} \mid \vec{Y}_t, \boldsymbol{\pi}) \mid \vec{Y}_t] \\
=&E[P_{\boldsymbol{\pi}}(\mathbf{Y}_{t+h} = \mathbf{k} \mid \vec{Y}_t) \mid \vec{Y}_t] \\
=&E[P_{\boldsymbol{\pi}}(Y_{1,t+h} = k_1 \mid \vec{Y}_t) \dots P_{\boldsymbol{\pi}}(Y_{n,t+h} = k_n \mid \vec{Y}_t) \mid \vec{Y}_t]
\end{aligned}$$

We cannot develop the distribution of $\mathbf{Y}_t$ much further without more information about the underlying distribution driving the sequence π_t, but we can see that the resulting transition matrix for $\mathbf{Y}_t$ would still be symmetric with regards to the i indices of the $Y_{i,t}$ chains that make up $\mathbf{Y}_t$. Therefore it is now clear that under a stochastic migration model the distribution of $\mathbf{Y}_t$ is symmetric with respect to the the i indices, in other words, changing which company is i and which company is j has no effect on the distribution of $\mathbf{Y}_t$.

3.1 Population Behavior

$\mathbf{Y}_t$ has a state space of K^n elements which makes Π_t a rather large matrix. However, since we assumed identical distributions for $Y_{1,t}, \ldots, Y_{n,t}$ then the distribution of $\mathbf{Y}_t$ is independant of which company takes which index. Therefore, when examining the behavior of a whole population it may be interesting to ignore which individual takes which value and only to examine how may indviduals take on each value. Indeed, we can see that when two individuals are in the same state at time $t-1$ then their next steps to time t will have identical distributions, that is:

$$\begin{aligned} &P(Y_{i,t} = k_1, Y_{j,t} = k_2 \mid Y_{i,t-1} = k_0, Y_{j,t-1} = k_0) \\ &\quad = P(Y_{i,t} = k_2, Y_{j,t} = k_1 \mid Y_{i,t-1} = k_0, Y_{j,t-1} = k_0) \end{aligned}$$

Therefore, we shall analyze how many total individuals take on a specific value. Let us introduce $\mathbf{Y}_t^*$ that is a vector of length K where:

$$\mathbf{Y}_t^* = \begin{pmatrix} \sum_{i=1}^{n} I_{Y_{i,t}=1} \\ \sum_{i=1}^{n} I_{Y_{i,t}=2} \\ \vdots \\ \sum_{i=1}^{n} I_{Y_{i,t}=K} \end{pmatrix}.$$

Here I_X is an indicator function equalling 1 if the event X holds and zero otherwise. Now $\mathbf{Y}_t^*$ has a state space of $\binom{n+K-1}{n}$ elements which is significantly less than the state space of $\mathbf{Y}_t$ (K^n elements) and therefore maybe more useful for describing the behavior of a whole industry.

Theorem 3.1. *Assume $Y_{1,t}, \ldots, Y_{n,t}$ satisfy a stochastic migration model, then the number of elements in the state space of $\boldsymbol{Y}_t^*$ will be strictly less than the number of elements in the state space of $\boldsymbol{Y}_t$ when the population size n is strictly greater than 1 and the number of possible credit scores K is strictly greater than 1.*

Proof. Here an inductive proof with respect to n or K can be used. Here is the proof with respect to K. First let us prove $\binom{n+K-1}{n} < K^n$ when $K = 2$ and $n \geq 2$.

$$\binom{n+1}{n} < 2^n$$
$$\text{if and only if } n+1 < 2^n,$$

which holds for $n \geq 2$. Now let us assume $\binom{n+K-1}{n} < K^n$ is true. Now we will show that this implies $\binom{n+K}{n} < (K+1)^n$, where we have now replaced K with $K+1$. Let us start by simplifing $\binom{n+K}{n}$.

$$\begin{aligned}\binom{n+K}{n} &= \frac{(n+K)!}{n!K!} \\ &= \frac{n+K}{K}\frac{(n+K-1)!}{n!(K-1)!} \\ &< \frac{n+K}{K}K^n = nK^{n-1} + K^n. \end{aligned} \tag{3.2}$$

Now we can simplify $(K+1)^n$ using the binomial theorem:

$$\begin{aligned}(K+1)^n &= \sum_{i=0}^{n}\binom{n}{i}K^{n-i} \\ &= K^n + nK^{n-1} + \sum_{i=2}^{n}\binom{n}{i}K^{n-i}. \end{aligned} \tag{3.3}$$

Now combining 3.2 and 3.3 we have:

$$nK^{n-1} + K^n < K^n + nK^{n-1} + \sum_{i=2}^{n}\binom{n}{i}K^{n-i}$$
$$\text{if and only if } 0 < \sum_{i=2}^{n}\binom{n}{i}K^{n-i},$$

where the last inequality holds for $n \geq 2$. Therefore the inductive argument holds. □

If we denote π_t^* as the transtion matrix for $\mathbf{Y}_t^*$ and the set $\Omega_{\mathbf{k}^*} := \{\mathbf{k} : \mathbf{Y}_t = \mathbf{k} \Rightarrow \mathbf{Y}_t^* = \mathbf{k}^*\}$ then the entries of π_t^* will be characterized by:

$$\begin{aligned}&P(\mathbf{Y}_{t+1}^* = \mathbf{k}^* \mid \vec{Y}_t) \\ &= \sum_{\mathbf{k}\in\Omega_{\mathbf{k}^*}} P(\mathbf{Y}_{t+1} = \mathbf{k} \mid \vec{Y}_t).\end{aligned}$$

Now using equation 3.1 we have:

$$\begin{aligned}&P(\mathbf{Y}_{t+1}^* = \mathbf{k}^* \mid \vec{Y}_t) \\ &= \sum_{\mathbf{k}\in\Omega_{\mathbf{k}^*}} E[P_{\boldsymbol{\pi}}(Y_{1,t+1} = k_1 \mid \vec{Y}_t)\dots P_{\boldsymbol{\pi}}(Y_{n,t+1} = k_n \mid \vec{Y}_t) \mid \vec{Y}_t],\end{aligned}$$

where $\mathbf{k} = (k_1, \dots, k_n)$ and $k_1, \dots, k_n$ take values in $1, \dots, K$.

3.2 Transition Estimation for Large Populations

Let us now examine the case where the number of individuals included in Y_t is tending to infinity. With such a large number of individuals, any unbiased estimators for our transition probabilities will converge in probability to the theoretical values. That is, if we define $N_{k;t}$ to be the number of individuals in state k at time t, $N_{k,l;t}$ to be the number of individuals stepping from k to l from time t to time $t+1$, and n to be the total number of individuals in the population then $\frac{N_{k,l;t}}{N_{k;t}}$ is an unbiased estimator for $P(Y_{i,t+1} = l \mid Y_{i,t} = k)$. Therefore we have:

$$\frac{N_{k,l;t}}{N_{k;t}} \to P(Y_{i,t+1} = l \mid Y_{i,t} = k),$$

in probability as $n \to \infty$. Thus we can now see that for large populations π_t is observable for any passed t.

4 Correlation

Now that the stochastic migration model has been defined we can compute the correlation of some specified movement between two companies. Notice that correlation is only relevant in the case where the sequence of π_t for future t is assumed to be unknown, otherwise the companies will be independent and therefore uncorrelated. Let us start by first defining the correlation between the movement of a company i from state k to k^* and a company j moving from state l to l^*, denoted as $\rho_t(k, k^*; l, l^*)$:

$$\rho_t(k, k^*; l, l^*) = \frac{cov(I_{Y_{i,t+1}=k^*}, I_{Y_{j,t+1}=l^*} \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l)}{Var(I_{Y_{i,t+1}=k^*} \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l)^{1/2} Var(I_{Y_{j,t+1}=l^*} \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l)^{1/2}}.$$

Notice that the indices of the companies do not matter therefore the above equation depends only on the current credit score of the two companies not which company is i and j. Also note that the correlation is conditioned upon $\vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l$ in order to emphasize the present state of the companies but also to acknowledge their dependence on the whole lagged history of the population. Let us start by simplifying the value of the covariance in the numerator.

$$\begin{aligned}
&cov(I_{Y_{i,t+1}=k^*}, I_{Y_{j,t+1}=l^*} \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l) \\
&\quad= E[I_{Y_{i,t+1}=k^*}, I_{Y_{j,t+1}=l^*} \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l] \\
&\quad- E[I_{Y_{i,t+1}=k^*} \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l] E[I_{Y_{j,t+1}=l^*} \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l].
\end{aligned}$$

Now using Lemma 3.1:

$$\begin{aligned}
&cov(I_{Y_{i,t+1}=k^*}, I_{Y_{j,t+1}=l^*} \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l) \\
&\quad= E[\pi_t(k, k^*)\pi_t(l, l^*) \mid \vec{Y}_t] - E[\pi_t(k, k^*) \mid \vec{Y}_t] E[\pi_t(l, l^*) \mid \vec{Y}_t].
\end{aligned} \tag{4.4}$$

Now the denominator can be simplified in a similar fashion.

$$\begin{aligned}
&Var(I_{Y_{i,t+1}=k^*} \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l) \\
&\quad= E[I_{Y_{i,t+1}=k^*} \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l] - E[I_{Y_{i,t+1}=k^*} \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l]^2 \\
&\quad= E[I_{Y_{i,t+1}=k^*} \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l](1 - E[I_{Y_{i,t+1}=k^*} \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l]).
\end{aligned}$$

Again, using Lemma 3.1:

$$\begin{aligned}
&Var(I_{Y_{i,t+1}=k^*} \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l) \\
&\quad= E[\pi_t(k, k^*) \mid \vec{Y}_t](1 - E[\pi_t(k, k^*) \mid \vec{Y}_t).
\end{aligned} \tag{4.5}$$

Now, combining equations 4.4 and 4.5 we have:

$$\rho_t(k, k^*; l, l^*) = \frac{E[\pi_t(k, k^*)\pi_t(l, l^*) \mid \vec{Y}_t] - E[\pi_t(k, k^*) \mid \vec{Y}_t]E[\pi_t(l, l^*) \mid \vec{Y}_t]}{\sqrt{E[\pi_t(k, k^*) \mid \vec{Y}_t](1 - E[\pi_t(k, k^*) \mid \vec{Y}_t)}\sqrt{E[\pi_t(l, l^*) \mid \vec{Y}_t](1 - E[\pi_t(l, l^*) \mid \vec{Y}_t)}}$$

One correlation that is of particular interest is that of two companies with the same credit rating moving simultaneously into default, also called default correlation. Therefore, if two companies are in a state l where $l \neq K$ then their default correlation is defined as $\rho_t(l, K; l, K)$.