Analysis of the Stochastic Migration Model

Matthew Logan Reimherr
Advisor: Professor Lajos Horváth

1 Abstract

Stochastic Migration Models are a type of probability model that can be used to model different aspects of a particular population. Gagliardini and Gourieroux (2005) showed how this general model could be used to model credit risk. The purpose of this paper is to provide concrete analytic grounding for the model proposed by Gagliardini and Gourieroux (2005). We start by describing the basic framework for credit scoring and then move into defining what is meant by Stochastic Migration Model. We derive several formulas for different probabilities associated with the model. After the model is defined, we describe the state space of the model. It is assumed, given enough time, that any company will go bankrupt. We provide proofs verifying such an assumption. Next we discuss an extension of the model to describe a collective group. We conclude by describing the correlation between the credit scores of two different companies. Also, we provide simulations emphasizing different ideas discussed in the paper.

Contents

2   Introduction

How does one define risk? Risk is a relative term which can mean something to one person and something entirely different to someone else. To analyze risk mathematically, one has to concretely define what is meant by risk. This paper will address 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 a return on their investment. If we look at a specific loan, we can define risk as the probability that the lender will not receive a full return on the loan. This event is equivalent to the company defaulting while the loan is still outstanding. As recommended by the Basel Committee [1], we can express risk in terms of the probability of time until 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. Symbols 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.

This is a basic model for understanding the risk of a specific company, but current rating systems do not use continuous credit scores. Scores are often discretized so that individuals that have a similar amount of risk can be grouped together.

---

[1]The Basel Comittee is an international group that sets banking standards used in many countries including the United States. A country is typically represented by its central bank.

3   Basic Framework

Let us examine a generic credit rating with K total scores taking 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. 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. 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. Each company has K possible scores, thus $\mathbf{Y}_t$ can take 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.

4 Stochastic Migration Model

The Stochastic Migration Model was in introduced by Gagliardini and Gourieroux (2005) as a generalized model for analyzing credit risk. Let $(\pi_t, 1 \leq t < \infty)$ be $K \times K$ random matrices which form a stochastic 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)th element of π_t.

Definition 4.1. *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

$$P_{\boldsymbol{\pi}}(\{Y_{1,t}, 1 \leq t < \infty\} \in C_1, \ldots, \{Y_{n,t}, 1 \leq t < \infty\} \in C_n)$$

$$= P_{\boldsymbol{\pi}}(\{Y_{1,t}, 1 \leq t < \infty\} \in C_1) \cdots P_{\boldsymbol{\pi}}(\{Y_{n,t}, 1 \leq t < \infty\} \in C_n).$$

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 and identically distributed Markov chains.

Note 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 $Y_{1,t}, \ldots, Y_{n,t}$ are independent and identically distributed Markov chains. Therefore, we can easily develop Π_t for any t.

The transition matrix for $\mathbf{Y}_t$ will be characterized by

$$P(\mathbf{Y}_{t+1} = \mathbf{k} \mid \vec{Y}_t, \boldsymbol{\pi}) = P_{\boldsymbol{\pi}}(\mathbf{Y}_{t+1} = \mathbf{k} \mid \vec{Y}_t)$$

$$= P_{\boldsymbol{\pi}}(Y_{1,t+1} = k_1, \dots, Y_{n,t+1} = k_n \mid \vec{Y}_t)$$

$$= P_{\boldsymbol{\pi}}(Y_{1,t+1} = k_1 \mid \vec{Y}_t) \cdots P_{\boldsymbol{\pi}}(Y_{n,t+1} = k_n \mid \vec{Y}_t),$$

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

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

With appropriately ordered states Π_t can be factored in the following way:

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

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

$$= \cdots = \pi_t \otimes \cdots \otimes \pi_t$$

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

$\otimes$ 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 $Y_{1,t}, \ldots, Y_{n,t}$. Using Definition 4.1(ii) we conclude

$$P_{\boldsymbol{\pi}}(\mathbf{Y}_{t+h} = \mathbf{k} \mid \vec{Y_t})$$
$$= P_{\boldsymbol{\pi}}(Y_{1,t+h} = k_1, \ldots, Y_{n,t+h} = k_n \mid \vec{Y_t})$$
$$= P_{\boldsymbol{\pi}}(Y_{1,t+h} = k_1 \mid \vec{Y_t}) \ldots P_{\boldsymbol{\pi}}(Y_{n,t+h} = k_n \mid \vec{Y_t}).$$

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 independent and known transition probabilities of the $Y_{i,t}$ chains.

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 4.1. *Suppose that ξ, γ, and ζ are random variables with finite expected values. Then*

$$E[\xi \mid \gamma = y] = E(E[\xi \mid \gamma = y, \zeta] \mid \gamma = y),$$

where y is an arbitrary outcome of γ.

Proof. Suppose that ξ takes values in an arbitrary set X and ζ takes values in an arbitrary set Z. We shall now provide a proof when ξ, ζ, and γ are discrete and a seperate proof for the continuous case. We start with the discrete case. Notice that

$$\mathrm{E}[\xi \mid \gamma = y, \zeta]$$

is a function of of γ and ζ. However, $\gamma = y$ and is therefore nonrandom. Now we have that

$$
\begin{aligned}
\mathrm{E}(\mathrm{E}[\xi \mid \gamma = y, \zeta] \mid \gamma = y) \\
&= \sum_{z \in Z} \mathrm{E}[\xi \mid \gamma = y, \zeta = z] \mathrm{P}(\zeta = z \mid \gamma = y) \\
&= \sum_{z \in Z} \sum_{x \in X} x \mathrm{P}(\xi = x \mid \gamma = y, \zeta = z) \mathrm{P}(\zeta = z \mid \gamma = y) \\
&= \sum_{z \in Z} \sum_{x \in X} x \frac{\mathrm{P}(\xi = x, \gamma = y, \zeta = z)}{\mathrm{P}(\gamma = y, \zeta = z)} \frac{\mathrm{P}(\zeta = z, \gamma = y)}{\mathrm{P}(\gamma = y)} \\
&= \sum_{x \in X} \frac{x}{\mathrm{P}(\gamma = y)} \sum_{z \in Z} \mathrm{P}(\xi = x, \gamma = y, \zeta = z) \\
&= \sum_{x \in X} x \frac{\mathrm{P}(\xi = x, \gamma = y)}{\mathrm{P}(\gamma = y)} \\
&= \sum_{x \in X} x \mathrm{P}(\xi = x \mid \gamma = y) \\
&= \mathrm{E}[\xi \mid \gamma = y].
\end{aligned}
$$

Exchanging the above summations is justified since all three variables have finite expected values, which implies the above sums are absolutely convergent. Now assume that ξ, ζ, and γ are continuous random variables and let f denote a density function. Then following the arguments used in the discrete case we obtain that

$$\mathrm{E}(\mathrm{E}[\xi \mid \gamma = y, \zeta] \mid \gamma = y)$$

$$= \int_Z \mathrm{E}[\xi \mid \gamma = y, \zeta = z] f_\zeta(z) \mathrm{d}z$$

$$= \int_Z \int_X x f_{\xi \mid \zeta, \gamma}(x \mid z, y) f_{\zeta \mid \gamma}(z \mid y) \mathrm{d}x \mathrm{d}z$$

$$= \int_Z \int_X x \frac{f_{\xi, \zeta, \gamma}(x, z, y)}{f_{\zeta, \gamma}(z, y)} \frac{f_{\zeta, \gamma}(z, y)}{f_\gamma(y)} \mathrm{d}x \mathrm{d}z.$$

We can now justify switching the integrals because they must be absolutely convergent if the random variables have a finite mean. Now we have

$$= \int_X \frac{x}{f_\gamma(y)} \int_Z f_{\xi, \zeta, \gamma}(x, z, y) \mathrm{d}z \mathrm{d}x$$

$$= \int_X x \frac{f_{\xi, \gamma}(x, y)}{f_\gamma(y)} \mathrm{d}x$$

$$= \int_X x f_{\xi \mid \gamma}(x \mid y) \mathrm{d}x$$

$$= \mathrm{E}[\xi \mid \gamma = y].$$

$\square$

Now using Lemma 4.1, Π_t will be characterized by:

$$P(\mathbf{Y}_{t+1} = \mathbf{k} \mid \vec{Y_t})$$

$$= \mathrm{E}[P(\mathbf{Y}_{t+1} = \mathbf{k} \mid \vec{Y_t}, \boldsymbol{\pi}) \mid \vec{Y_t}]$$

$$= \mathrm{E}[P_{\boldsymbol{\pi}}(\mathbf{Y}_{t+1} = \mathbf{k} \mid \vec{Y_t}) \mid \vec{Y_t}]$$

$$(4.1) \qquad = \mathrm{E}[P_{\boldsymbol{\pi}}(Y_{1,t+1} = k_1 \mid \vec{Y_t}) \dots P_{\boldsymbol{\pi}}(Y_{n,t+1} = k_n \mid \vec{Y_t}) \mid \vec{Y_t}].$$

We can also derive a formula for Π_t^h using Lemma 4.1 when we condition on $\vec{Y_t}$:

$$P(\mathbf{Y}_{t+h} = \mathbf{k} \mid \vec{Y_t})$$

$$= \mathrm{E}[P(\mathbf{Y}_{t+h} = \mathbf{k} \mid \vec{Y_t}, \boldsymbol{\pi}) \mid \vec{Y_t}]$$

$$= \mathrm{E}[P_{\boldsymbol{\pi}}(\mathbf{Y}_{t+h} = \mathbf{k} \mid \vec{Y_t}) \mid \vec{Y_t}]$$

$$= \mathrm{E}[P_{\boldsymbol{\pi}}(Y_{1,t+h} = k_1 \mid \vec{Y_t}) \dots P_{\boldsymbol{\pi}}(Y_{n,t+h} = k_n \mid \vec{Y_t}) \mid \vec{Y_t}].$$

The resulting transition matrix for $\mathbf{Y}_t$ will be symmetric with respect to the i indices of the $Y_{i,t}$ chains that make up $\mathbf{Y}_t$. In other words, changing which company is i and which company is j has no effect on the distribution of $\mathbf{Y}_t$.

4.1   Classifying the State Space

It is assumed once the state of default, K, is entered the process can never leave. This type of state is referred to as an absorbing barrier. Classifying which states are transient and which are recurrent can reveal what the long term distribution of any $Y_{i,t}$ chain may look like. Our simulations (see Section 6) suggest that the only recurrent state for any $Y_{i,t}$ chain is the default state and all other states are transient. This should not be surprising since the default state is assumed to be an absorbing barrier. Now we provide a proof that the product of $\pi_1 \dots \pi_n$ converges almost surely when $\pi_1, \dots, \pi_n$ are independent and identically distributed transition matrices with an absorbing barrier in state K.

Theorem 4.1. *Assume* $\boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_n$ *are independent and identically distributed random transition matrices where* $0 < \boldsymbol{\pi}_r(i,j) < 1$ *with probability 1 for* $r = 1, \dots, n$, $i = 1, \dots, K-1$ *and* $j = 1, \dots, K$. *Assume* $\boldsymbol{\pi}_r(K,j) = 0$ *for* $j = 1, \dots, K-1$ *and*

$\pi_r(K, K) = 1$. *Finally, assume that* $E[\pi_r] = \pi$ *for* $r = 1, \ldots, n$, *where*

$$\lim_{n \to \infty} \pi^n = \begin{pmatrix} 0 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix}.$$

Then we have

$$\lim_{n \to \infty} \pi_1 \cdot \ldots \cdot \pi_n = \begin{pmatrix} 0 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix}$$

with probability 1.

Proof. This proof will require using several smaller steps. First, let us show that the product of any two $K \times K$ transition matrices will again be a transition matrix. Let us denote $\mathbf{A}$ and $\mathbf{B}$ as two arbitrary $K \times K$ transition matrices. Then $\mathbf{A}$ and $\mathbf{B}$ will have the following property:

$$\sum_{j=1}^{K} \mathbf{A}(i, j) = \sum_{j=1}^{K} \mathbf{B}(i, j) = 1,$$

for $i = 1, \ldots, K$, by the definition of a transition matrix. The sum over a row of the product of $\mathbf{A}$ and $\mathbf{B}$ can be expressed as

$$(\mathbf{A} \cdot \mathbf{B})(i, j) = \sum_{s=1}^{K} \mathbf{A}(i, s) \mathbf{B}(s, j),$$

by the definition of a matrix product. Therefore, the sum over a row of $\mathbf{A} \cdot \mathbf{B}$ can be

expressed as:

$$\sum_{j=1}^{K} (\mathbf{A} \cdot \mathbf{B})(i,j) = \sum_{j=1}^{K} \sum_{s=1}^{K} \mathbf{A}(i,s)\mathbf{B}(s,j)$$

$$= \sum_{s=1}^{K} [\mathbf{A}(i,s) \sum_{j=1}^{K} \mathbf{B}(s,j)]$$

$$= \sum_{s=1}^{K} \mathbf{A}(i,s) \cdot 1$$

$$= 1.$$

The summations are interchangeable because K is assumed to be finite. Therefore, all entries of $\mathbf{A} \cdot \mathbf{B}$ are positive and the rows all add up to 1, which implies $\mathbf{A} \cdot \mathbf{B}$ is a transition matrix. It is a trivial inductive argument to show that the product of n arbitrary transition matrices is again a transtion matrix.

Let us now examine

$$E\big[\lim_{n\to\infty} \boldsymbol{\pi}_1 \cdot \ldots \cdot \boldsymbol{\pi}_n\big].$$

The terms in the sequence $\{(\boldsymbol{\pi}_1 \cdot \ldots \cdot \boldsymbol{\pi}_n)(i,j)\}_{n=1}^{\infty}$ are bounded below by 0 and above by 1 for any $i = 1, \ldots, K$ and $j = 1, \ldots, K$, since the product of n transition matrices is again a transition matrix. Therefore, by the Lebesgue Dominated Convergence Theorem

$$E[\lim_{n\to\infty} \boldsymbol{\pi}_1 \cdot \ldots \cdot \boldsymbol{\pi}_n]$$

$$= \lim_{n\to\infty} E[\boldsymbol{\pi}_1 \cdot \ldots \cdot \boldsymbol{\pi}_n]$$

$$= \lim_{n\to\infty} E[\boldsymbol{\pi}_1] \cdot \ldots \cdot E[\boldsymbol{\pi}_n]$$

$$= \lim_{n\to\infty} \boldsymbol{\pi}^n$$

which impiles

$$E[\lim_{n \to \infty} \boldsymbol{\pi}_1 \cdot \ldots \cdot \boldsymbol{\pi}_n] = \begin{pmatrix} 0 & \ldots & 1 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & 1 \end{pmatrix}.$$

Now let us examine the product $\boldsymbol{\pi}_1 \cdot \ldots \cdot \boldsymbol{\pi}_n \cdot \boldsymbol{\pi}_{n+1}$

$$(\boldsymbol{\pi}_1 \cdot \ldots \cdot \boldsymbol{\pi}_n \cdot \boldsymbol{\pi}_{n+1})(i, j) = \sum_{s=1}^{K} (\boldsymbol{\pi}_1 \cdot \ldots \cdot \boldsymbol{\pi}_n)(i, s) \cdot \boldsymbol{\pi}_{n+1}(s, j).$$

Therefore, the (i, K) entry can be written as:

$$
\begin{aligned}
(\boldsymbol{\pi}_1 \cdot \ldots \cdot \boldsymbol{\pi}_n \cdot \boldsymbol{\pi}_{n+1})(i, K) &= \sum_{s=1}^{K} (\boldsymbol{\pi}_1 \cdot \ldots \cdot \boldsymbol{\pi}_n)(i, s) \cdot \boldsymbol{\pi}_{n+1}(s, K) \\
&= (\boldsymbol{\pi}_1 \cdot \ldots \cdot \boldsymbol{\pi}_n)(i, K) \cdot \boldsymbol{\pi}_{n+1}(K, K) + \sum_{s=1}^{K-1} (\boldsymbol{\pi}_1 \cdot \ldots \cdot \boldsymbol{\pi}_n)(i, s) \cdot \boldsymbol{\pi}_{n+1}(s, K) \\
&> (\boldsymbol{\pi}_1 \cdot \ldots \cdot \boldsymbol{\pi}_n)(i, K),
\end{aligned}
$$

since $\boldsymbol{\pi}_{n+1}(K, K) = 1$. Therefore the sequence $\{(\boldsymbol{\pi}_1 \cdot \ldots \cdot \boldsymbol{\pi}_n)(i, K)\}_{n=1}^{\infty}$ is monotonically increasing and bounded, which implies

$$\boldsymbol{\pi}_1 \cdot \ldots \cdot \boldsymbol{\pi}_n(i, K) \to \xi,$$

almost surely, where ξ is a random variable taking values in $[0, 1]$ with probability 1 and $E[\xi] = 1$. We shall now show that $P(\xi = 1) = 1$. Suppose that there exists an $\epsilon > 0$ such that $P(\xi < 1 - \epsilon) = q > 0$. Therefore $P(\xi \geq 1 - \epsilon) = 1 - q$ and

$$E[\xi] \leq (1 - \epsilon)q + (1)(1 - q) = 1 - q\epsilon.$$

If $q > 0$ and $\epsilon > 0$, we have that

$$1 - q \cdot \epsilon < 1,$$

implying that $\mathrm{E}[\xi] < 1$, which is a contradiction. Therefore $\mathrm{P}(\xi = 1) = 1$. We now see that

$$\lim_{n \to \infty} \boldsymbol{\pi}_1 \cdot \ldots \cdot \boldsymbol{\pi}_n(i, K) = 1$$

with probability 1.

The last step is to show all other values go to zero, i.e.,

$$\lim_{n \to \infty} \boldsymbol{\pi}_1 \cdot \ldots \cdot \boldsymbol{\pi}_n(i, j) = 0$$

with probability 1 for $i = 1, \ldots, K$ and $j = 1, \ldots, K - 1$. Let us examine

$$\limsup_{n \to \infty} \boldsymbol{\pi}_1 \cdot \ldots \cdot \boldsymbol{\pi}_n(i, j).$$

Since the lim sup must still be an entry of a transition matrix we have:

$$\limsup_{n \to \infty} \boldsymbol{\pi}_1 \cdot \ldots \cdot \boldsymbol{\pi}_n(i, j)$$

$$\leq 1 - \liminf_{n \to \infty} \boldsymbol{\pi}_1 \cdot \ldots \cdot \boldsymbol{\pi}_n(i, K)$$

We have shown that $\lim_{n \to \infty} \boldsymbol{\pi}_1 \cdot \ldots \cdot \boldsymbol{\pi}_n(i, K)$ exists, therefore

$$\liminf_{n \to \infty} \boldsymbol{\pi}_1 \cdot \ldots \cdot \boldsymbol{\pi}_n(i, K) = \lim_{n \to \infty} \boldsymbol{\pi}_1 \cdot \ldots \cdot \boldsymbol{\pi}_n(i, K).$$

This implies

$$1 - \lim_{n \to \infty} \inf\{\boldsymbol{\pi}_1 \cdot \ldots \cdot \boldsymbol{\pi}_n(i, K)\} = 1 - \lim_{n \to \infty} \boldsymbol{\pi}_1 \cdot \ldots \cdot \boldsymbol{\pi}_n(i, K) = 0.$$

Therefore we have

$$0 \leq \lim_{n \to \infty} \boldsymbol{\pi}_1 \cdot \ldots \cdot \boldsymbol{\pi}_n(i, j)$$

$$\leq \limsup_{n \to \infty} \boldsymbol{\pi}_1 \cdot \ldots \cdot \boldsymbol{\pi}_n(i, j) \leq 0,$$

which implies

$$\lim_{n \to \infty} \boldsymbol{\pi}_1 \cdot \ldots \cdot \boldsymbol{\pi}_n(i, j) = 0$$

for $i = 1, \ldots, K$ and $j = 1, \ldots, K - 1$. Now combining the previous steps we have

$$\lim_{n \to \infty} \boldsymbol{\pi}_1 \cdot \ldots \cdot \boldsymbol{\pi}_n = \begin{pmatrix} 0 & \ldots & 1 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & 1 \end{pmatrix}.$$

$\square$

4.2   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 independent 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 examine how many 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 step to

time t will have identical distributions, i.e.,

$$P(Y_{i,t} = k_1, Y_{j,t} = k_2 \mid Y_{i,t-1} = k_0, Y_{j,t-1} = k_0)$$

$$= P(Y_{i,t} = k_2, Y_{j,t} = k_1 \mid Y_{i,t-1} = k_0, Y_{j,t-1} = k_0)$$

for any two values k_1 and k_2 in the state space. Therefore, we shall analyze how many 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 arbitrary event X holds, and zero otherwise. $\mathbf{Y}_t^*$ 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, may be more useful for describing the behavior of a whole industry.

Theorem 4.2. *Assume $Y_{1,t}, \ldots, Y_{n,t}$ satisfy a Stochastic Migration Model; then the number of elements in the state space of $\mathbf{Y}_t^*$ will be strictly less than the number of elements in the state space of $\mathbf{Y}_t$ when the population size n is strictly greater than 1, and the number of possible credit scores K is strictly greater than 1.*

Proof. The following is an inductive proof with respect to K. First let us prove $\binom{n+K-1}{n} <$

K^n when $K = 2$ and $n \geq 2$. We observe that

$$\binom{n+1}{n} < 2^n$$

holds if and only if $n + 1 < 2^n$, which holds for $n \geq 2$. Assume $\binom{n+K-1}{n} < K^n$ is true. We will show this implies $\binom{n+K}{n} < (K+1)^n$, where we have now replaced K with $K + 1$. Start by simplifing $\binom{n+K}{n}$.

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

(4.2)

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

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

(4.3)

Combining 4.2 and 4.3 we have

$$nK^{n-1} + K^n < K^n + nK^{n-1} + \sum_{i=2}^{n} \binom{n}{i} K^{n-i},$$

if and only if $0 < \sum_{i=2}^{n} \binom{n}{i} K^{n-i}$, which holds for $n \geq 2$. Therefore the inductive argument holds. $\qquad \square$

Let π_t^* denote the transtion matrix for $\mathbf{Y}_t^*$ and define $\Omega_{\mathbf{k}^*}$ as the set of all $\mathbf{k}$ in the

sample space such that $\mathbf{Y}_t = \mathbf{k}$ implies that $\mathbf{Y}_t^* = \mathbf{k}^*$, i.e.,

$$\Omega_{\mathbf{k}^*} = \{\mathbf{k} : \mathbf{Y}_t = \mathbf{k} \Rightarrow \mathbf{Y}_t^* = \mathbf{k}^*\}.$$

Then the entries of π_t^* will be given by

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

Using equation 4.1 we have:

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

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

4.3 Limiting Behavior

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 the transition probabilities of π_t 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 $N_{k,l;t}/N_{k;t}$ is an unbiased estimator for $\pi_t(k,l)$. Therefore we have:

$$\frac{N_{k,l;t}}{N_{k;t}} \to \pi_t(k,l),$$

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

5 Correlation

Now that the Stochastic Migration Model has been defined, we can compute the correlation of some specified movement between two companies. 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. 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 $\rho_t(k, k^*; l, l^*)$, as

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

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, and to acknowledge their dependence on the entire lagged history of the population. Start with simplifying the value of the covariance in the numerator. Using the definition of the conditional covariance we write

$$\mathrm{Cov}(I_{Y_{i,t+1}=k^*}, I_{Y_{j,t+1}=l^*} \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l)$$

$$= \mathrm{E}[I_{Y_{i,t+1}=k^*}, I_{Y_{j,t+1}=l^*} \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l]$$

$$- \mathrm{E}[I_{Y_{i,t+1}=k^*} \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l] \mathrm{E}[I_{Y_{j,t+1}=l^*} \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l].$$

Now using Lemma 4.1 we get

$$\mathrm{Cov}(I_{Y_{i,t+1}=k^*}, I_{Y_{j,t+1}=l^*} \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l)$$

(5.4)
$$= \mathrm{E}[\pi_t(k, k^*)\pi_t(l, l^*) \mid \vec{Y}_t] - \mathrm{E}[\pi_t(k, k^*) \mid \vec{Y}_t]\mathrm{E}[\pi_t(l, l^*) \mid \vec{Y}_t].$$

Now the denominator can be simplified in a similar fashion resulting in

$$\mathrm{Var}(I_{Y_{i,t+1}=k^*} \mid \vec{Y_t}, Y_{i,t} = k, Y_{j,t} = l)$$

$$=\mathrm{E}[I_{Y_{i,t+1}=k^*} \mid \vec{Y_t}, Y_{i,t} = k, Y_{j,t} = l] - \mathrm{E}[I_{Y_{i,t+1}=k^*} \mid \vec{Y_t}, Y_{i,t} = k, Y_{j,t} = l]^2$$

$$=\mathrm{E}[I_{Y_{i,t+1}=k^*} \mid \vec{Y_t}, Y_{i,t} = k, Y_{j,t} = l](1 - \mathrm{E}[I_{Y_{i,t+1}=k^*} \mid \vec{Y_t}, Y_{i,t} = k, Y_{j,t} = l]).$$

Again, using Lemma 4.1 we conclude

$$\mathrm{Var}(I_{Y_{i,t+1}=k^*} \mid \vec{Y_t}, Y_{i,t} = k, Y_{j,t} = l)$$

(5.5)
$$=\mathrm{E}[\pi_t(k, k^*) \mid \vec{Y_t}](1 - \mathrm{E}[\pi_t(k, k^*) \mid \vec{Y_t}]).$$

We shall denote equation 5.5 as $\sigma(k, k^*; \vec{Y_t})^2$. Now, combining equations (5.4) and (5.5) we have

$$\rho_t(k, k^*; l, l^*) = \frac{\mathrm{E}[\pi_t(k, k^*)\pi_t(l, l^*) \mid \vec{Y_t}] - \mathrm{E}[\pi_t(k, k^*) \mid \vec{Y_t}]\mathrm{E}[\pi_t(l, l^*) \mid \vec{Y_t}]}{\sigma(k, k^*; \vec{Y_t})\sigma(l, l^*; \vec{Y_t})}$$

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

The terms in the sequence $\{\rho_t(k, k^*; l, l^*)\}_{k*=1}^K$ are dependent random variables. We investigate the sum

$$\sum_{k*=1}^K \rho_t(k, k^*; l, l^*)\mathrm{Var}(I_{Y_{i,t+1}=k^*} \mid \vec{Y_t}, Y_{i,t} = k, Y_{j,t} = l).$$

This sum is equivalent to

$$\sum_{k*=1}^{K} \frac{\text{Cov}(I_{Y_{i,t+1}=k^*}, I_{Y_{j,t+1}=l^*} \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l)}{\text{Var}(I_{Y_{i,t+1}=l^*} \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l)^{1/2}}$$

$$\sum_{k*=1}^{K} \frac{\text{P}(Y_{i,t+1} = k^*, Y_{j,t+1} = l^* \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l)}{\text{Var}(I_{Y_{i,t+1}=l^*} \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l)^{1/2}}$$

$$- \sum_{k*=1}^{K} \frac{\text{P}(Y_{j,t+1} = l^* \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l)\text{P}(Y_{i,t+1} = k^* \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l)}{\text{Var}(I_{Y_{i,t+1}=l^*} \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l)^{1/2}}.$$

Notice we are summing over all possible values of $Y_{i,t+1}$, therefore

$$\sum_{k*=1}^{K} \frac{\text{P}(Y_{i,t+1} = k^*, Y_{j,t+1} = l^* \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l)}{\text{Var}(I_{Y_{i,t+1}=l^*} \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l)^{1/2}}$$

$$- \sum_{k*=1}^{K} \frac{\text{P}(Y_{j,t+1} = l^* \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l)\text{P}(Y_{i,t+1} = k^* \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l)}{\text{Var}(I_{Y_{i,t+1}=l^*} \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l)^{1/2}}$$

$$= \frac{\text{P}(Y_{j,t+1} = l^* \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l)}{\text{Var}(I_{Y_{i,t+1}=l^*} \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l)^{1/2}} - \frac{\text{P}(Y_{j,t+1} = l^* \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l)}{\text{Var}(I_{Y_{i,t+1}=l^*} \mid \vec{Y}_t, Y_{i,t} = k, Y_{j,t} = l)^{1/2}}$$

$$= 0.$$

6 Simulations

The following code simulates a sequence of independent and identically distributed 5 by 5 transtion matrices. The fifth row of any matrix is assumed to be an absorbing barrier and represents the default state. We used the following R code:

```
n<-5
p<-.5

TM<-function(n,p){
 y<-rnorm(n-1,mean=.25,sd=.08)
 pi<-diag(c(1-y,1))
 for(i in 1:n-1){
  nrm=0
  nrm<-pgeom(i-1,p)+pgeom(n-i,p)-2*dgeom(0,p)
  for(j in 1:n){
   if(i != j) pi[i,j]<-((1-pi[i,i])*dgeom(abs(i-j),p))/nrm
  }
 }
 return(x)
}

TRL10<-TM(n,p)
for(i in 1:10) TRL10=TRL10%*%TM(n,p)

TRL100<-TM(n,p)
for(i in 1:100) TRL100=TRL100%*%TM(n,p)

TRL1000<-TM(n,p)
for(i in 1:1000) TRL1000=TRL1000%*%TM(n,p)

TRL10000<-TM(n,p)
for(i in 1:10000) TRL10000=TRL10000%*%TM(n,p)

TRL100000<-TM(n,p)
for(i in 1:100000) TRL10000=TRL100000%*%TM(n,p)
```

This program generates an independent and identically distributed sequence of transition matrices by modifying the identity matrix. First a normal random variable with mean $= .25$ and variance $= .08$ is generated for each state except the default state ($K - 1$

independent and identically distributed normal random variables). This sequence is de-

noted $\{y_i\}_{i=1}^{K-1}$. Then $\pi(i,i) = 1 - y_i$ for $i = 1, \ldots, K - 1$ and $\pi(K,K) = 1$. Next,

y_i is distributed into the off–diagonal entries according to a geometric scheme. Each

off–diagonal entry gets the value

$$(6.6) \qquad \pi(i,j) = \frac{((1 - x(i,i)) * P(geom(p) = |i - j|)}{P(geom(p) \leq i - 1) + P(geom(p) \leq n - i) - 2 * P(geom(p) = 0)}$$

for $i \neq j$, $geom(p)$ is used to indicate an arbitrary geometric random variable with mean $=$

$1/p$. The denominator standardizes all the entries to ensure the sum over any row is 1,

and thus we have generated a transition matrix. This program produced the following

output:

```
> TRL1
            [,1]       [,2]       [,3]       [,4]       [,5]
[1,]  0.74036827 0.13847026 0.06923513 0.03461756 0.01730878
[2,]  0.08270657 0.77255695 0.08270657 0.04135328 0.02067664
[3,]  0.03599434 0.07198868 0.78403396 0.07198868 0.03599434
[4,]  0.02972865 0.05945730 0.11891459 0.67298488 0.11891459
[5,]  0.00000000 0.00000000 0.00000000 0.00000000 1.00000000

> TRL10
            [,1]       [,2]       [,3]       [,4]       [,5]
[1,]  0.13968539 0.2753065 0.1804053 0.1269599 0.2776429
[2,]  0.10841753 0.2623285 0.1795247 0.1299426 0.3197867
[3,]  0.09628662 0.2241265 0.1660348 0.1238358 0.3897162
[4,]  0.08522965 0.2026122 0.1584807 0.1272861 0.4263913
[5,]  0.00000000 0.0000000 0.0000000 0.0000000 1.0000000

> TRL100
            [,1]        [,2]        [,3]        [,4]       [,5]
[1,]  0.003311207 0.006706110 0.008580879 0.005259308 0.9761425
[2,]  0.003026417 0.006129330 0.007842855 0.004806966 0.9781944
[3,]  0.002800917 0.005672630 0.007258478 0.004448795 0.9798192
[4,]  0.002569497 0.005203941 0.006658762 0.004081223 0.9814866
[5,]  0.000000000 0.000000000 0.000000000 0.000000000 1.0000000
```

```
> TRL1000
             [,1]          [,2]          [,3]          [,4] [,5]
[1,] 2.129608e-18 3.782552e-18 4.682232e-18 2.951377e-18    1
[2,] 2.072351e-18 3.680853e-18 4.556344e-18 2.872025e-18    1
[3,] 1.925095e-18 3.419301e-18 4.232582e-18 2.667946e-18    1
[4,] 1.809864e-18 3.214630e-18 3.979230e-18 2.508250e-18    1
[5,] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00    1

> TRL10000
              [,1]           [,2]           [,3]           [,4] [,5]
[1,] 1.509437e-168 3.313584e-168 4.174466e-168 2.163808e-168    1
[2,] 1.497106e-168 3.286515e-168 4.140364e-168 2.146131e-168    1
[3,] 1.358130e-168 2.981428e-168 3.756014e-168 1.946906e-168    1
[4,] 1.280975e-168 2.812054e-168 3.542637e-168 1.836303e-168    1
[5,]  0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00    1

> TRL100000
     [,1] [,2] [,3] [,4] [,5]
[1,]    0    0    0    0    1
[2,]    0    0    0    0    1
[3,]    0    0    0    0    1
[4,]    0    0    0    0    1
[5,]    0    0    0    0    1
```

These simulations suggest that the product $\pi_1 \ldots \pi_n$ converges almost surely as $n \to \infty$ when the sequence is independent and identically distributed and state K is an absorbing barrier. Next we shall simulate the time until convergence of the product of these transition matrices. The necessary code is an extension of the code used above. We shall assume that the product has converged when all entries are either less than .01 or greater than .99. We used the following R code:

```
K=5000
p=.5
n=5
mconv<-function(K,n,p)
{
 num<-numeric(K)
 for(j in 1:K)
 {
```

```
   num[j]<-0
   conv<-diag(n)
   while(sum(conv[,n]<.99)>0){
    num[j]=num[j]+1
    conv=conv%*%TM(n,p)
   }
 }
 return(num)
}
x=mconv(K,n,p)
hist(x,  xlab="Time", ylab="Frequency", freq = FALSE,main=paste("Time Until Convergence"
```

which generated Figure 7.1 (see Section 7). The histogram in Figure 7.1 indicates that the time until reaching the limit is normally distributed.

Our final group of simulations will simulate the times until default when starting in the different states. We have been examining a credit score with five possible values therefore we will simulate the time until default from states one through four. The following R code simulates the four times until default:

```
Trials<-function(nn,start){
 n<-5
 p<-.5
 xx<-numeric(nn)
 for(i in 1:nn){
  run=start
  while(run != n){
   xx[i]=xx[i]+1
   check=0
   j=1
   pi<-TM(n,p)
   u=runif(1)
   while(check<u){
    check=check+pi[run,j]
    temp=j
    j=j+1
   }
   run=temp
  }
 }
```

```
 return(xx)
}
T1=Trials(5000,1)
T2=Trials(5000,2)
T3=Trials(5000,3)
T4=Trials(5000,4).
```

Figure 7.2 is the density of the time until default from state one; Figure 7.3 is from state

two; Figure 7.4 is from state three; and Figure 7.5 is from state four.
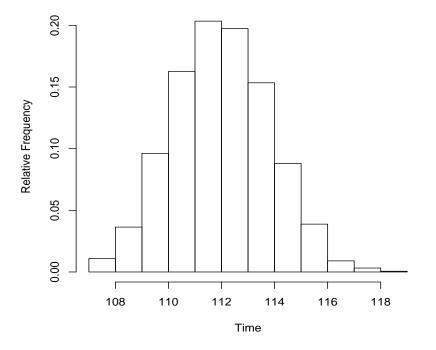
7   Appendix

FIGURE   7.1. Histogram of 5000 simulated times needed to reach the limit for the product of random transition matrices
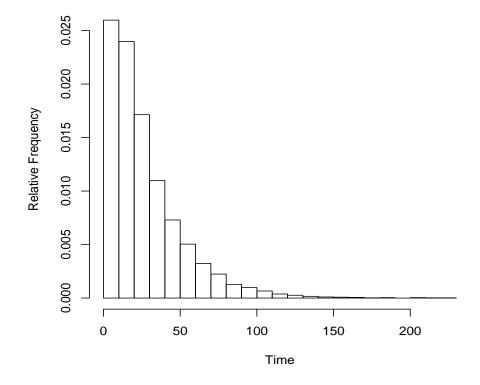
FIGURE 7.2. Histogram of 5000 simulated times until default from state one
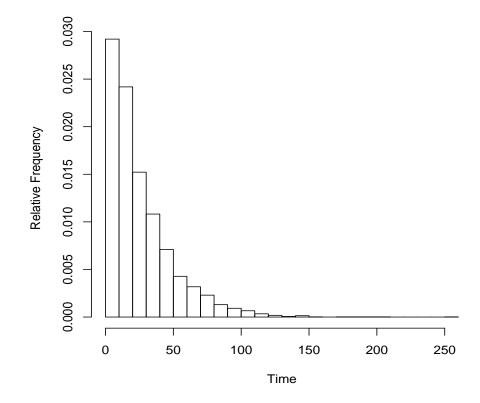
FIGURE 7.3. Histogram of 5000 simulated times until default from state two
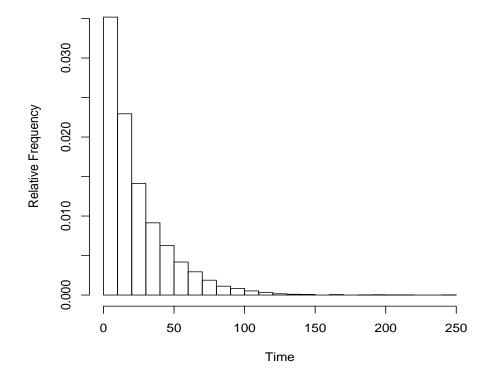
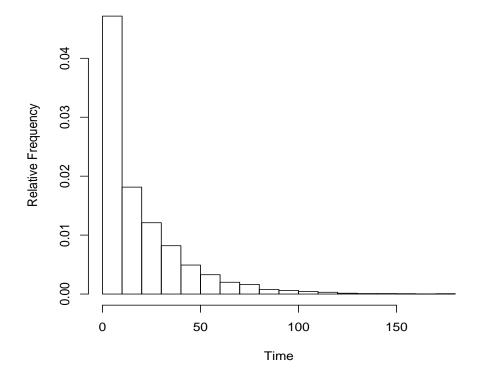FIGURE 7.4. Histogram of 5000 simulated times until default from state three

FIGURE   7.5. Histogram of 5000 simulated times until default from state four

8 Bibliography

1. Bain, Lee J. and Engelhardt, Max. *Introduction to Probability and Mathematical Statistics.* Duxbury Press; Second Edition 2000.

2. Gourieroux, Christian and Gagliardini, Patrick. "Stochastic Migration Models with Application to Corporate Risk" . Journal of Financial Econometrics, Vol. 3, No. 2, pp. 188–226, 2005 Available at http://ssrn.com/abstract=821748.

3. Ross, Sheldon M. *Introduction to Probability Models.* Academic Press; Eighth Edition 2002.