



Diplomarbeit

The potential approach to the term structure of interest rates- Theory and Application

ausgeführt am
Institut für Wirtschaftsmathematik
der Technischen Universität Wien

unter Anleitung von Ao. Univ.-Prof. Dr. Josef Teichmann

durch
Andreas Magenschab
Bräuhäusgasse 45/21
1050 Wien

Datum

Unterschrift

Danksagung

Mein besonderer Dank gilt Hr. Prof. Josef Teichmann, vom Institut für Wirtschaftsmathematik an der TU Wien, der mein Interesse an diesem Thema erweckt hat und mich während der Entstehung dieser Arbeit unterstützt hat. Weiterer Dank gilt natürlich meiner Familie und dabei ganz besonders meinen Eltern, die mir mein Studium ermöglicht haben.

Abstract

I wrote this diploma thesis under the supervision of Prof. Teichmann in the winter term 2007/2008. The main aim of this text is the study of a new approach to the term structure of interest rates called the potential approach.

The first chapter recalls some basic facts about Stochastic Analysis. First some definitions of stochastic processes are introduced, then Markov processes and Brownian motion are discussed. For the Markov processes the main interest is in finite state-space Markov chains and the framework of potential operators.

In the second chapter we give some basic definitions and ideas of interest rate theory. First technical terms like the spot-rate, forward-rate or the yield to maturity are introduced, then some economic theories are presented which try to explain the different shapes of the yield curves.

Chapter three goes along with [Rogers 1] and [Rogers 2] and shows how we can use potential theory to build interest rate models. Some examples are given using a Markov process with continuous state-space.

In the fourth chapter we try to simulate the stochastic processes defined in chapter three. First we do this by using a simple Markov chain with just five states, then we use a process with continuous state-space. Also a fit to real yield curve data is presented.

Some additional background knowledge and the code of the programmed functions in Scilab can be viewed in the Appendix.

Contents

1	Introduction	1
1.1	Preliminaries of stochastic analysis	1
1.1.1	Basic measure theory	1
1.1.2	Stochastic processes	3
1.2	Markov processes	5
1.2.1	General theory of Markov processes	5
1.2.2	Transition functions	6
1.2.3	The strong Markov property and potential operators	6
1.2.4	Finite state space Markov chains	7
1.2.5	Time reversible Markov chains	13
1.2.6	Calculation of transition probability functions	14
1.3	Brownian motion	15
1.3.1	Brownian motion as a martingale	16
1.3.2	Brownian motion as a Markov process	17
1.3.3	Brownian motion as a diffusion	18
2	Introduction to interest rate theory	19
2.1	Bond prices and interest rates	19
2.1.1	Short-term interest rate	19
2.1.2	Zero-coupon bonds	20
2.1.3	Forward interest rates	20
2.1.4	Arbitrage-free family of bond prices	21
2.2	The term structure of interest rates	22
2.2.1	The typical shape of the yield curve	22
2.2.2	Theoretical Models	23
3	The potential approach	27
3.1	Introduction	27
3.2	Pricing	29
3.2.1	Pricing with the spot rate	30
3.3	The potential approach	31
3.4	Markov processes and potentials	31
3.4.1	Generic approach 1	32
3.4.2	Generic approach 2	33
3.5	Application of Markov chain potential models	33
3.6	Hedging in the potential model	34
3.6.1	Mini-example	34
3.7	Examples	35
3.7.1	Exponential-linear example	35
3.7.2	Exponential-quadratic example	36
3.7.3	Quadratic example	37
3.7.4	cosh example	37
3.7.5	Exponential-linear example 2	38
3.7.6	Multi-type branching diffusion example	38

3.7.7	Rational Log-normal Model	39
3.8	Foreign Exchange in the Potential approach	39
3.8.1	Foreign Exchange in complete markets	39
3.8.2	Example	40
4	Simulation	43
4.1	Simulation of a finite Markov chain	43
4.2	Calculation of bond-prices and yields	46
4.2.1	Generic approach 1	46
4.2.2	Generic approach 2	50
4.3	Simulation of spot-rates and state-price densities	54
4.4	Yields in calendar time	56
4.4.1	Generic Approach 1	57
4.4.2	Generic Approach 2	57
4.4.3	Results	58
4.5	A continuous example	59
4.6	A trial fit to real data	62
4.6.1	Fit with a Markov chain	63
4.6.2	Continuous case	63
A	Appendix	66
A.1	Year fraction, Day-count convention	66
A.2	The L^p -spaces	67
A.3	Generator of an Itô-diffusion	68
A.4	Code	70

1 Introduction

1.1 Preliminaries of stochastic analysis

This chapter is meant to recall the basic ideas and definitions of stochastic processes (in particular Markov processes and Brownian motion) which are needed for better understanding of the following chapters. The reader with mathematical background can directly jump to chapter 2. Most proofs of the theorems are left out but referred to literature. The most is taken from [Bauer 1] and [Wertz].

1.1.1 Basic measure theory

Before we mention some stochastic, probabilistic framework, we introduce in short some basics from measure theory.

Let Ω be a general set and let $\mathcal{P}(\Omega)$ be the potential set of Ω . This is the system of all subsets of Ω . Then for all sets $(A_i)_{i \in I}$ from $\mathcal{P}(\Omega)$ also $\bigcup_{i \in I} A_i$ and $\bigcap_{i \in I} A_i$ are subsets of $\mathcal{P}(\Omega)$. Further $\mathcal{P}(\Omega)$ contains the complement $\mathbb{C}A$ to every set $A \in \mathcal{P}(\Omega)$.

In the following we are just interested in at least countable index sets I .

Definition 1.1.1 (σ -algebra) A system \mathcal{A} of subsets of a set Ω is called a σ -algebra, if it fulfills the following properties:

- (1) $\Omega \in \mathcal{A}$
- (2) $A \in \mathcal{A} \Rightarrow \mathbb{C}A \in \mathcal{A}$
- (3) For every sequence $(A_n)_{n \in \mathbb{N}}$ of sets from \mathcal{A} : $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

Clearly $\mathcal{P}(\Omega)$ is a σ -algebra. Now we look at the definition of another system of sets quite similar to that of a σ -algebra.

Definition 1.1.2 (Ring) A system \mathcal{R} of subsets of a set Ω is called ring (in Ω) if it has the following properties:

- (1) $\emptyset \in \mathcal{R}$
- (2) $A, B \in \mathcal{R} \Rightarrow A \setminus B \in \mathcal{R}$
- (3) $A, B \in \mathcal{R} \Rightarrow A \cup B \in \mathcal{R}$

If a ring \mathcal{R} additionally fulfills

- (4) $\Omega \in \mathcal{R}$ then we call it an algebra (in Ω).

Every ring does not only contain the union, but also the intersection of two (and therefore $n < \infty$) subsets. Because for $A, B \in \mathcal{R}$ we can write $A \cap B = A \setminus (A \setminus B)$.

Now we use these definitions to define measures on them.

Definition 1.1.3 We call the pair (Ω, \mathcal{A}) a measurable space if \mathcal{A} is a σ -algebra on the set Ω . The sets $A \in \mathcal{A}$ are called measurable sets.

For maps we have:

Definition 1.1.4 Let (Ω, \mathcal{A}) and (Ω', \mathcal{A}') be two measurable spaces and consider the map $X : \Omega \rightarrow \Omega'$. Then X is called measurable if $X^{-1}(A') \in \mathcal{A}$ for all $A' \in \mathcal{A}'$.

Definition 1.1.5 Let \mathcal{R} be a ring in Ω and μ be a function on \mathcal{R} with values in $[0, \infty)$. Then we call μ a pre-measure if

- (1) $\mu(\emptyset) = 0$.
- (2) For every sequence $(A_n)_{n \in \mathbb{N}}$ of pairwise disjuncted subsets of \mathcal{R} with union in \mathcal{R} :

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

(σ -additivity)

The property (2) clearly also holds for finite union.

Every pre-measure μ on \mathcal{R} fulfills the following properties:

- ((A, B, A_1, A_2, \dots) are all subsets of \mathcal{R})
- (3) $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$
- (4) $A \subset B \Rightarrow \mu(A) \leq \mu(B)$ (isotony)
- (5) $A \subset B, \mu(A) \leq +\infty \Rightarrow \mu(B \setminus A) = \mu(B) - \mu(A)$ (subtractivity)
- (6) $\mu(\bigcup_{i=1}^n A_i) \leq \sum_{i=1}^n \mu(A_i)$ (subadditivity)
- (7) For every sequence $(A_n)_{n \in \mathbb{N}}$ of pairwise disjuncted subsets of \mathcal{R} with union in \mathcal{R} :

$$\sum_{n=1}^{\infty} \mu(A_n) \leq \mu\left(\bigcup_{n=1}^{\infty} A_n\right)$$

If the (A_n) are not disjuncted in (7) we clearly have

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n)$$

if the $\bigcup_{n=1}^{\infty} A_n \in \mathcal{R}$. For the proofs of this properties lookup [Bauer 1].

Now we use our definition of a pre-measure to define what a measure is.

Definition 1.1.6 (Measure) Every pre-measure μ defined on a σ -algebra \mathcal{A} in a set Ω is called a measure on \mathcal{A} . The value $\mu(A)$ for $A \in \mathcal{A}$ is called the μ -measure of the set A . If $\mu(\Omega) < \infty$ (and therefore $\mu(A) < \infty \quad \forall A \in \mathcal{A}$) we say μ is finite.

Definition 1.1.7 (Probability measure) A measure μ on a σ -algebra \mathcal{A} in Ω is called a probability measure if $\mu(\Omega) = 1$. From the isotony property we can follow that for every probability measure

$$0 \leq \mu(A) \leq 1 = \mu(\Omega) \quad (A \in \mathcal{A}).$$

Definition 1.1.8 A measure λ on \mathcal{A} is absolute continuous relative to μ if $\forall A \in \mathcal{A}$: $\mu(A) = 0 \Rightarrow \lambda(A) = 0$ and we write $\lambda \ll \mu$.

Lemma 1.1.1 Suppose that λ and μ are finite measures on a measurable space. Then $\lambda \ll \mu \Leftrightarrow$ if for $\epsilon > 0$ we can find a $\delta > 0$ such that

$$\forall A \in \mathcal{A} : \mu(A) < \delta \Rightarrow \lambda(A) < \epsilon.$$

Definition 1.1.9 We again assume λ and μ to be finite. Further $\lambda \ll \mu$ and $\mu \ll \lambda$. Then we say that λ and μ are equivalent. They have the same null-sets.

Now we use this to derive the following very important result.

Theorem 1.1.1 (Radon-Nikodym) *Let (Ω, \mathcal{A}) be a measurable space and let λ and μ be measures on (Ω, \mathcal{A}) . If μ is finite, the following statements are equivalent:*

- (1) $\lambda \ll \mu$
- (2) λ has a density relative to μ .

From this we can follow:

Theorem 1.1.2 *Let $\lambda = f\mu$ be a measure of the density f relative to a σ -finite measure μ on \mathcal{A} . Then f is almost everywhere μ , λ is σ -finite $\Leftrightarrow f$ is μ a.e. real-valued. We write*

$$f = \frac{d\lambda}{d\mu} \quad \text{a.e.}(\mu)$$

for the Radon-Nikodym density.

Now we recall the definition of a measurable space (Ω, \mathcal{A}) and continue to the following definition.

Definition 1.1.10 *If additionally a measure μ is defined on the σ -algebra \mathcal{A} we call the triple $(\Omega, \mathcal{A}, \mu)$ a measure space. If μ is a probability space, we call the triple a probability measure.*

1.1.2 Stochastic processes

After this basic introduction we can define the basic properties of a stochastic process. First we will define the process itself.

Definition 1.1.11 *Consider a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ a measure space (S, Σ) and an index set $T \neq \emptyset$. We define the map $X_t : (\Omega, \mathcal{A}, \mu) \rightarrow (S, \Sigma)$ for all $t \in T$. Then the family $X := (X_t)_{t \in T}$ is called a stochastic process on $(\Omega, \mathcal{A}, \mu)$ with state space (S, Σ) .*

Sometimes a stochastic process is also referred to as a sequence (family) of random variables. Mostly we have $T \subseteq \mathbb{R}$ with T interpreted as index of time.

Definition 1.1.12 *Let X be a stochastic process. Then every map $\pi : t \rightarrow X_t(\omega)$ for fixed $\omega \in \Omega$ is called path (trajectory, realisation) of X .*

$$\pi : t \rightarrow X_t(\omega) : t \in T, \omega \in \Omega$$

is called the set of paths.

Often we have assumptions about π : For example paths can be continuous (left- or right-continuous).

Some authors use the notation $X(t)$ instead of X_t to denote that X is a function of time.

Definition 1.1.13 (Filtration) *We say that $(\mathcal{F}_t)_{t \in T}$ is a filtration of a measurable space (Ω, \mathcal{F}) if*

- (1) $\mathcal{F}_t \subseteq \mathcal{F}$ is a σ -algebra $\forall t \in T$.
- (2) $s \leq t \Rightarrow \mathcal{F}_s \subseteq \mathcal{F}_t \quad \forall s, t \in T$.

In other words: $(\mathcal{F}_t)_{t \in T}$ is an increasing family of σ -algebras. It carries all the information about the process up to time t .

Remark 1.1.1 *If X is a stochastic process and \mathcal{F}_t is the σ -algebra produced by the process itself we write $\mathcal{F}_t = \sigma(X_s, s \leq t)$ for the natural filtration of X .*

In the following we will often write $(\Omega, \mathcal{F}_t, \mathbb{P})$ to state that the σ -algebra we usually had is a filtration. If not explicitly given otherwise we take \mathcal{F}_t to be the natural filtration.

Definition 1.1.14 (Adapted process) Let X be a stochastic process on $(\Omega, \mathcal{F}_t, \mathbb{P})$, then X is \mathcal{F}_t adapted $\Leftrightarrow X_t$ is \mathcal{F}_t measurable for all $t \in T$.

Definition 1.1.15 (Previsible process) The previsible σ -algebra \mathcal{F} on $(0, \infty) \times \Omega$ is defined to be the smallest σ -algebra on $(0, \infty) \times \Omega$ such that every adapted process is \mathcal{F} -measurable. A process with time-parameter set $(0, \infty)$ is called previsible if it is a \mathcal{F} -measurable map from $(0, \infty) \times \Omega \rightarrow \mathbb{R}$.

Another useful definition is the concept of a stopping time. This is a random variable that depends on the underlying process.

Definition 1.1.16 (Stopping Time) We say T is a stopping time (in continuous time) with respect to the filtration $(\mathcal{F}_t)_{t \geq 0}$ if T is a random variable in $[0, \infty)$, such that for any $t \geq 0$

$$[T \leq t] \in \mathcal{F}_t.$$

Proposition 1.1.1 If T and S are stopping times on the same filtration $(\mathcal{F}_t)_{t \geq 0}$ then also

(1) $T + S, T \wedge S, T \vee S^1$ are stopping times.

If $(T_n)_{n \geq 1}$ is a sequence of stopping times on the same filtration $(\mathcal{F}_t)_{t \geq 0}$ then

(2) $\sup_{n \geq 1} \{T_n\}$ is a stopping time.

Most stochastic processes are in general sequences of identically independent distributed (iid.) random variables, or sums of them. Such sequences have the fundamental property that the distribution of the sequence not change if we remove $n < \infty$ elements at the beginning. This leads to following definition:

Definition 1.1.17 A stochastic process X is called stationary if for all $t_1, \dots, t_n \in T$ and all h the vectors $(X_{t_1}, \dots, X_{t_n})$ and $(X_{t_1+h}, \dots, X_{t_n+h})$ have the same distribution.

Another class of stochastic processes:

Definition 1.1.18 A stochastic process X is called process with independent increments if for $t_1 < t_2 < \dots < t_n$ the random variables

$$(X_{t_1}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}})$$

are independent.

Now we introduce in short some basics of martingale theory. This concept has a lot of applications not only in mathematical finance. Also many classical theorems can be proved with martingale theory.

Definition 1.1.19 (S-Martingales) Let X be a stochastic process on $(\Omega, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$. Then X is a martingale if it fulfills the following properties:

- (1) X_t is \mathcal{F}_t adapted $\quad \forall t \in T$
- (2) $\mathbb{E}[|X_t|] < \infty \quad \forall t \in T$ (X is integrable)
- (3) $\mathbb{E}[X_t | \mathcal{F}_s] = X_s \quad \forall s \leq t, s, t \in T$

If we have the relation " \leq " in property (3) we say the process X is a supermartingale, if the relation is " \geq " X is a submartingale.

¹ $T \wedge S = \min(T, S), T \vee S = \max(T, S)$

When the stochastic process X is a martingale we can think of a fair game. In expectation we can neither win nor lose under the condition that we use the available information. If we translate this idea to financial markets we have that all "fair" prices have to be martingales under the assumption that there is no arbitrage allowed.

Definition 1.1.20 (Local Martingale) A process $(X_t)_{t \geq 0}$ is a local martingale if there is a sequence of $(\mathcal{F})_{t \geq 0}$ stopping times $(T_n)_{n \geq 1}$ such that $X_{t \wedge T_n}$ is a martingale for each n and

$$\mathbb{P}\left[\lim_{n \rightarrow \infty} T_n = \infty\right] = 1.$$

Definition 1.1.21 (Orthogonal Martingales) Two martingales M_1, M_2 are called orthogonal if $\langle M_1, M_1 \rangle = 0$.

Proposition 1.1.2 A stopped martingale is again a martingale. So if we follow a strategy in a fair game, which tells us to stop after the process hits a certain barrier, we can again neither win nor lose in expectation.

1.2 Markov processes

1.2.1 General theory of Markov processes

In this section we will give the general definitions of a Markov process. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, (E, \mathcal{E}) a measurable space and T an arbitrary index set.

Definition 1.2.1 Let $X = (X_t)_{t \in T}$ be a stochastic process with values in (E, \mathcal{E}) . One says that X is a Markov process with respect to a filtration $(\mathcal{F}_t)_{t \in T}$ of \mathcal{F} provided that

- (1) X is adapted to \mathcal{F}_t
- (2) $\forall t \in T$ the σ -algebras \mathcal{F}_t and $\sigma(X_s : s \geq t)$ are conditionally independent given X_t :

$$\mathbb{P}(A \cap B | X_t) = \mathbb{P}(A | X_t) \mathbb{P}(B | X_t) \quad (1.1)$$

whenever $A \in \mathcal{F}_t, B \in \sigma(X_s : s \geq t)$.

The intuitive meaning of this condition should be clear. Namely given the present, X_t , the past $\sigma(X_s : s \leq t)$ and the future $\sigma(X_s : s > t)$ are independent.

Now we give some equivalent formulations of Definition 1.2.1:

Theorem 1.2.1 (a) Let $X = (X_t)_{t \in T}$ be a stochastic process adapted to the family \mathcal{F}_t . Then the following statements are equivalent:

- (1) X is a Markov process with respect to \mathcal{F}_t
- (2) for each $t \in T$ and $Y \in \sigma(X_s : s \geq t)$ we have

$$\mathbb{E}[Y | \mathcal{F}_t] = \mathbb{E}[Y | X_t]; \quad (1.2)$$

- (3) if t and s are in T and $t \leq s$ then

$$\mathbb{E}[f(X_s) | \mathcal{F}_t] = \mathbb{E}[f(X_s) | X_t] \quad (1.3)$$

for all f in $b\mathcal{E}$.

- (b) X is a Markov process (with respect to $\sigma(X_s : s \geq t)$) if for each finite collection $t_1 \leq \dots \leq t_n \leq t$ from T and $f \in b\mathcal{E}$ we have

$$\mathbb{E}[f(X_t) | X_{t_1}, \dots, X_{t_n}] = \mathbb{E}[f(X_t) | X_{t_n}]. \quad (1.4)$$

Proof 1.2.1 Lookup [Blumenthal/Gettoor].

1.2.2 Transition functions

In this section we will assume that the parameter set T is $\mathbb{R}_+ = [0, \infty)$ which is the case of most interest to us.

Definition 1.2.2 Let (E, \mathcal{E}) be a measurable space, then a function $P_{t,s}(x, A)$ defined for $0 \leq t < s < \infty$, x in E , and A in \mathcal{E} , is called a Markov transition function on (E, \mathcal{E}) provided

- (1) $A \rightarrow P_{t,s}(x, A)$ is a probability measure on \mathcal{E} for each t, s and x .
- (2) $x \rightarrow P_{t,s}(x, A)$ is in \mathcal{E} for each t, s and A .
- (3) if $0 \leq t < s < u$, then

$$P_{t,u}(x, A) = \int P_{t,s}(x, dy) P_{s,u}(y, A) \quad (1.5)$$

for all x and A .

The relationship (1.5) is called the Chapman-Kolmogorov equation.

Definition 1.2.3 A Markov transition function $P_{t,s}(x, A)$ on (E, \mathcal{E}) is said to be temporally homogeneous provided there exists a function $P_t(x, A)$ defined for $t > 0, x \in E$, and $A \in \mathcal{E}$ such that $P_{t,s}(x, A) = P_{s-t}(x, A)$, $0 \leq t < s < \infty$ for all x and A .

In this case $P_t(x, A)$ is called a temporally homogeneous Markov transition function over (E, \mathcal{E}) and the Chapman-Kolmogorov equation becomes

$$P_{t+s}(x, A) = \int P_t(x, dy) P_s(y, A) \quad (1.6)$$

for all t, s, x and A .

Definition 1.2.4 Let X be a stochastic process with values in (E, \mathcal{E}) that is adapted to \mathcal{F}_t and let $P_{t,s}(x, A)$ be a Markov transition function on (E, \mathcal{E}) . We say that X is a Markov process with respect to \mathcal{F}_t having $P_{t,s}(x, A)$ as transition function provided

$$\mathbb{E}[f(X_s) | \mathcal{F}_t] = P_{t,s}(X_t, f) \quad (1.7)$$

for all $0 \leq t < s$ and f in $b\mathcal{E}$, where

$$P_{t,s}(x, f) = \int P_{t,s}(x, dy) f(y).$$

Taking conditional expectations with respect to $\sigma(X_t)$ in (1.7) we obtain

$$\mathbb{E}[f(X_s) | X_t] = P_{t,s}(X_t, f). \quad (1.8)$$

The fact that a Markov process has a transition function means that it is possible to define "nice" conditional probability distributions for the conditional probabilities $\mathbb{P}(X_s \in A | \mathcal{F}_t)$, $0 \leq t < s, A \in \mathcal{E}$. $P_{t,s}(x, A)$ is the conditional probability that $X_s \in A$, given that $X_t = x$ when $0 \leq t < s$.

1.2.3 The strong Markov property and potential operators

Now we will check if the Markov property will hold for stopping times as well as for fixed times t . We develop a sufficient condition for this "extended" or "strong" Markov property.

Definition 1.2.5 (Strong Markov) Let X be a stochastic process, then X is strong Markov if and only if for each stopping time T and f in $b\mathcal{E}$ one has

$$\mathbb{E}^x[f(X_{t+T}) | \mathcal{F}_t] = \mathbb{E}^{X_T}[f(X_t)] \quad \forall \quad t, x. \quad (1.9)$$

If f in $b\mathcal{E}$ we may now write

$$P_t f(x) = \int P_t(x, dy) f(y) = \mathbb{E}^x[f(X_t)]. \quad (1.10)$$

Clearly P_t is a positive linear operator from $b\mathcal{E}$ to $b\mathcal{E}$. It follows from the Chapman-Kolmogorow equations that $(P_t)_{t \geq 0}$ is a semigroup of linear operators on $b\mathcal{E}$.

Definition 1.2.6 (Potential Operator, Resolvent) *If $\alpha > 0$ and $f \in b\mathcal{E}$ we can define*

$$R_\alpha f(x) = \int_0^\infty e^{-\alpha t} P_t f(x) dt = \mathbb{E}^x \left[\int_0^\infty e^{-\alpha t} f(X_t) dt \right]. \quad (1.11)$$

The operator R_α is called the α -potential operator of X and the family $(R_\alpha, \alpha > 0)$ is called the resolvent of the semi-group $(P_t)_{t \geq 0}$.

This is just the (componentwise) Laplace transform of the semigroup. When X is a finite state-space Markov chain², with $P_t = \exp(tQ)$, it follows that

$$R_\alpha = (\alpha - Q)^{-1}. \quad (1.12)$$

Proof 1.2.2

$$\begin{aligned} R_\alpha &= \int_0^\infty e^{-\alpha t} P_t dt = \int_0^\infty e^{-\alpha t} \exp(tQ) dt \\ &= \int_0^\infty e^{-t(\alpha - Q)} dt = -(\alpha - Q)^{-1} e^{-t(\alpha - Q)} \Big|_0^\infty \\ &= (\alpha - Q)^{-1}. \quad \square \end{aligned}$$

Remark 1.2.1 *It is easy to check the following simple algebraic properties of the resolvent for $\alpha, \beta > 0$:*

- (1) $\|R_\alpha\| \leq \alpha^{-1}$
- (2) $R_\alpha - R_\beta = (\beta - \alpha)R_\alpha R_\beta$
- (3) $R_\alpha R_\beta = R_\beta R_\alpha$

1.2.4 Finite state space Markov chains

In this section we will have a look at the case when the state-space \mathcal{N} of the Markov process is finite, $\mathcal{N} = \{0, 1, \dots, N\}$. This is a Markov chain³.

For this section let $T = [0, \infty)$ and consider a stochastic process $X = (X_t)_{t \geq 0}$ taking values on $\mathcal{N} = \{0, 1, \dots, N\}$. The most in this subsection is taken from [Kijima].

We recall the general definitions of the general form for this special case.

Definition 1.2.7 (Markov property) *If for each $s \geq 0, t > 0$, each $i, j \in \mathcal{N}$ and every history $X_u, 0 \leq u < s$*

$$\mathbb{P}[X_{t+s} = j | X_s = i, X_u = x(u) : 0 \leq u < s] = \mathbb{P}[X_{t+s} = j | X_s = i], \quad (1.13)$$

we call X a continuous time Markov chain. In other words the conditional distribution of the future state of the process, given the present and the past, does only depend on the present state and is independent of the past.

²The exact definition of this term will be given in the next section.

³Some authors use the term chain to signify that the time parameter is discrete.

For this special type of process we define:

Definition 1.2.8 (Transition function) Let X be a continuous time Markov chain and define

$$p_{ij}(s, t) = \mathbb{P}[X_t = j | X_s = i], \quad 0 \leq s < t. \quad (1.14)$$

The conditional probability $p_{ij}(s, t)$ is called the transition probability function from state i to state j and the Matrix $P(s, t) = (p_{ij}(s, t))_{i, j \in \mathcal{N}}$ is called the transition matrix function.

Definition 1.2.9 We call the matrix $P(s, t)$ stochastic if $P(s, t) \geq \mathbf{0}^4$ and $P(s, t)\mathbf{1} = \mathbf{1}$.

For any fixed $0 \leq s < t$, we assume that the Matrix $P(s, t)$ is stochastic.

For $0 \leq s < u < t$, taking a stop at u and using the Markov property we must have

$$p_{ij}(s, t) = \sum_k p_{ik}(s, u) p_{kj}(u, t), \quad 0 \leq s < u < t. \quad (1.15)$$

This equation is called the *Chapman-Kolmogorov equation*. In Matrix notation this is written as

$$P(s, t) = P(s, u)P(u, t), \quad 0 \leq s < u < t.$$

When the transition probability functions $p_{ij}(s, t)$ depends only on the difference $t - s$, we have

$$p_{ij}(t - s) = \mathbb{P}[X_{t-s} = j | X_s = i], \quad 0 \leq s < t.$$

for all $i, j \in \mathcal{N}$, the continuous-time Markov chain X_t is said to be *homogeneous* (non-homogeneous otherwise). For any homogeneous Markov chain the Chapman-Kolmogorov equation is expressed as

$$p_{ij}(s + t) = \sum_k p_{ik}(s) p_{kj}(t), \quad s, t > 0, \quad (1.16)$$

or in matrix form

$$P(s + t) = P(s)P(t), \quad s, t > 0, \quad (1.17)$$

where $P(t) = (p_{ij}(t))$ which satisfies $P(t - s) = P(s, t)$ for $t > s \geq 0$.

In the following we consider the homogeneous case only and assume that the transition matrix function $P(t)$ is stochastic for all $t \geq 0$. We also assume that every transition probability function $p_{ij}(t)$ is continuous in $t > 0$.

Definition 1.2.10 A transition matrix function $P(t) = (p_{ij}(t))$ is called *standard* if, for every i and j , $\lim_{t \rightarrow 0+} p_{ij}(t) = \delta_{i,j}$ or, equivalently,

$$\lim_{t \rightarrow 0+} P(t) = I.$$

In the following we assume that the transition matrix function under consideration is standard unless stated otherwise.

Lemma 1.2.1 For every i , the transition probability function $p_{ij}(t)$ is positive for all $t \geq 0$. If $p_{ij}(s) > 0$ then $p_{ij}(t) > 0$ for all $t \geq s$.

This lemma can be sharpened into the following fundamental result, called Levy's theorem.

Theorem 1.2.2 For every i and j , the transition probability function $p_{ij}(t)$ is either identically zero or positive for all $t > 0$.

⁴The bold $\mathbf{0}, \mathbf{1}$ should signify that this is a vector or matrix with all entries 0 or 1.

Theorem 1.2.3 For every i and j , we have

$$|p_{ij}(t+h) - p_{ij}(t)| \leq 1 - p_{ii}(h), \quad h > 0, t \geq 0.$$

Proof 1.2.3 Lookup [Kijima].

A consequence of this theorem is the following:

Corrolary 1.2.1 For every i and j , $p_{ij}(t)$ is uniformly continuous in $t \geq 0$.

Moreover, $p_{ij}(t)$ is differentiable with respect to $t \geq 0$ as we shall see below. (The derivative $p'_{ij}(0+)$ is the right hand derivative at $t = 0$.) We say that the matrix $P(t) = (p_{ij}(t))$ is differentiable with respect to t if each component $p_{ij}(t)$ is differentiable with respect to t . Also $\int P(t)dt$ means the componentwise integral.

Lemma 1.2.2 There exists some $h > 0$ such that $\int_0^h P(t)dt$ is nonsingular.

Proof 1.2.4 Lookup [Kijima].

Theorem 1.2.4 The transition matrix function $P(t)$ is differentiable with respect to $t > 0$.

Proof 1.2.5 From the Chapman-Kolmogorov equation we have

$$\int_t^{t+h} P(u)du = \int_0^h P(t+u)du = P(t) \left(\int_0^h P(u)du \right) = \left(\int_0^h P(u)du \right) P(t).$$

Fix $h > 0$ so that $\int_0^h P(u)du$ is nonsingular. Then

$$P(t) = \int_t^{t+h} P(u)du \left(\int_0^h P(u)du \right)^{-1} = \left(\int_0^h P(u)du \right)^{-1} \int_t^{t+h} P(u)du.$$

Since $P(t)$ is continuous in t , $\int_t^{t+h} P(u)du$ is differentiable with respect to $t > 0$, whence the theorem. \square

From the above proof the derivative of $P(t)$ is given by

$$P'(t) = (P(t+h) - P(t)) \left(\int_0^h P(u)du \right)^{-1} = \left(\int_0^h P(u)du \right)^{-1} (P(t+h) - P(t)), \quad (1.18)$$

which shows that $P(t)$ is infinitely differentiable with respect to $t > 0$. We define

$$Q = P'(0+). \quad (1.19)$$

The matrix $Q = (q_{ij})$ is called the infinitesimal generator, or generator for short. Since $P(0) = I$, we have

$$q_{ij} = \begin{cases} \lim_{h \rightarrow 0+} \frac{p_{ij}(h)}{h} \geq 0 & , i \neq j \\ \lim_{h \rightarrow 0+} \frac{p_{ij}(h)-1}{h} \leq 0 & , i = j \end{cases} \quad (1.20)$$

or in matrix form,

$$Q = \lim_{h \rightarrow 0+} \frac{P(h) - I}{h}. \quad (1.21)$$

Also, since for any $h \geq 0$

$$1 - p_{ii}(h) = \sum_{j \neq i} p_{ij}(h),$$

dividing both sides by h and letting h decrease to zero yields the relation

$$-q_{ii} = \sum_{j \neq i} q_{ij}, \quad i \in \mathcal{N}. \quad (1.22)$$

We define $q_i = \sum_{j \neq i} q_{ij} = -q_{ii}$. From (1.20) and (1.21) the infinitesimal generator Q has the following properties:

Theorem 1.2.5 *Let Q be the infinitesimal generator of a finite Markov chain. Then Q is finite componentwise, the diagonal elements of Q are nonpositive, the off-diagonal elements are nonnegative, and the row sums of Q are all zero, i.e., $Q\mathbf{1} = \mathbf{0}$.*

Referring to the Chapman-Kolmogorov equation we have

$$\frac{P(t+h) - P(t)}{h} = \frac{P(h) - I}{h} P(t) = P(t) \frac{P(h) - I}{h}, \quad h > 0.$$

Since $P(t)$ is differentiable, it follows from (1.20) that

$$P'(t) = QP(t); \quad P'(t) = P(t)Q, \quad t \geq 0, \quad (1.23)$$

which are systems of ordinary linear differential equations, known as the *backward Kolmogorov equation* and the *forward Kolmogorov equation*. The unique solution of the system under the initial condition $P(0) = I$ is given by

$$P(t) = \exp(Qt) = \sum_{n=0}^{\infty} \frac{Q^n t^n}{n!}, \quad t \geq 0. \quad (1.24)$$

There is a one to one correspondence between the infinitesimal generator Q and the transition matrix function $P(t)$. The generator Q plays a similar role as the one step transition matrix in discrete time.

Suppose that we want to evaluate the joint probability

$$\mathbb{P}[X_{t_0} = i_0, \dots, X_{t_n} = i_n],$$

where $0 = t_0 < t_1 < \dots < t_n$. Let $\alpha = (\alpha_i)$ be the initial distribution of X . Then the chain rule of conditional probability together with repeated application of the Markov property and homogeneity yields

$$\mathbb{P}[X_{t_0} = i_0, \dots, X_{t_n} = i_n] = \alpha_{i_0} p_{i_0, i_1}(t_1) \dots p_{i_{n-1}, i_n}(t_n - t_{n-1}). \quad (1.25)$$

We see that the initial distribution α and the transition matrix function $P(t)$, which is determined by the generator Q , together suffice to determine every joint distribution of the continuous time Markov chain X . Because of this property any finite Markov chain in continuous time can be described by giving its generator and its initial distribution.

We consider a continuous time Markov chain with infinitesimal generator Q and recall that $q_i = -q_{ii} = \sum_{j \neq i} q_{ij}$.

Definition 1.2.11 (1) *State i is called absorbing if $q_i = 0$.*

(2) *State i is called stable if $0 < q_i < \infty$.*

Definition 1.2.12 *We define*

$$T_i = \inf\{t \geq 0 : X_t \neq i\}$$

if the set is not empty and $T_i = \infty$ otherwise. When $X_0 = i$, the random variable T_i is called the holding time in state i . If i is an absorbing state then $T_i = \infty$.

We write $\mathbb{P}_i[A]$ for $\mathbb{P}_i[A|X_0 = i]$.

Theorem 1.2.6 *For a finite Markov chain X_t in continuous time with infinitesimal generator $Q = (q_{ij})$, suppose that $q_i > 0$. Then:*

(1) $\mathbb{P}_i[T_i > t] = e^{-q_i t}$ for $t \geq 0$.

(2) $\mathbb{P}_i[X(T_i) = j] = q_{ij}/q_i$ for $i \neq j$.

Proof 1.2.6 *Lookup [Kijima].*

Suppose that $X_0 = i$. If state i is absorbing, then the chain remains there forever. If $q_i > 0$, then the chain stays in i for a finite but strictly positive amount of time T_i which is exponentially distributed with parameter q_i . At the end of this holding time the chain moves to state j according to the transition law q_{ij}/q_i . Now, because of homogeneity and the strong Markov property, the future behavior of the chain after T_i is independent of the past behavior and the chain repeats this procedure as if the initial state were j .

Just for completeness we mention some other classification of states for a continuous time Markov chain, which can be directly followed from the properties of a discrete time Markov chain since for every $a > 0$ the sequence $\eta(n) = \xi(an)$ is a Markov chain in discrete time. (Hence: the notation is discrete! For further definitions and background information lookup [Kijima].)

Let T_j be the first time the chain X_t visits the state $j \in \mathcal{N}$, i.e. $T_j = \inf\{n \geq 1 : X_n = j\}$. With the convention that the infimum of the empty set is ∞ . The random variable T_j is called the first passage time to state j and is a stopping time dependent on the realisation of X_n . That is, given $\omega \in \Omega$, $T_j(\omega) = n$ if and only if $X_k(\omega) \neq j$ for $k = 1, \dots, n-1$ and $X_n(\omega) = j$.

We define

$$f_{ij}(n) = \mathbb{P}_i[T_j = n] \quad n = 1, 2, \dots$$

The distribution $\{f_{ij}(n)\}$ with respect to n is called the first passage-time distribution of X_n from state i to state j . It can be recursively calculated by

$$f_{ij}(n) = \begin{cases} p_{ij}, & n = 1, \\ \sum_{k \neq j} p_{ik} f_{kj}(n-1), & n \geq 2. \end{cases}$$

The transition probabilities and the first-passage-time distributions are connected through the following identity. Hence: δ_{ij} is the Kronecker-delta, meaning $\delta_{ij} = 1$ if $i = j$ and 0 otherwise.

Theorem 1.2.7 *For a Markov chain with transition matrix $P = (p_{ij})$,*

$$p_{ij}(n) = \begin{cases} \delta_{ij}, & n = 0, \\ \sum_{\nu=1}^{\infty} f_{ij}(\nu) p_{jj}(n-\nu), & n \geq 1. \end{cases}$$

Let N_j be the number of visits to state j for the Markov chain X_n , so that

$$N_j = \sum_{n=0}^{\infty} I_{\{X_n=j\}}.$$

For a given $\omega \in \Omega$, if $N_j(\omega)$ is finite, then the Markov chain X_n eventually leaves state j never to return. If $N_j(\omega) = \infty$, X_n visits state j repeatedly. Since

$$p_{ij}(n) = \mathbb{P}_i[X_n = j] = \mathbb{E}_i[I_{\{X_n=j\}}],$$

it follows that

$$\sum_{n=0}^{\infty} p_{ij}(n) = \sum_{n=0}^{\infty} \mathbb{E}_i[I_{\{X_n=j\}}] = \mathbb{E}_i[N_j]. \quad (1.26)$$

The second equality follows from Fubini's theorem. We then have

$$p_{ij}^* = \mathbb{E}_i[N_j], \quad i, j \in \mathcal{N},$$

the expected number of visits to state j starting from state i . Also,

$$f_{ij}^*(n) = \sum_{n=1}^{\infty} f_{ij}(n) = \mathbb{P}_i[T_j < \infty] \quad i, j \in \mathcal{N}.$$

It follows that $p_{ij}^* = f_{ij}^* p_{jj}^*$ or

$$\mathbb{E}_i[N_j] = \mathbb{P}_i[T_j < \infty] \mathbb{E}_j[N_j], \quad i, j \in \mathcal{N},$$

otherwise we have $p_{ij}^* = (1 - f_{jj}^*)^{-1}$ or

$$\mathbb{E}_j[N_j] = \frac{1}{1 - \mathbb{P}_j[T_j < \infty]}, \quad j \in \mathcal{N}, \quad (1.27)$$

where the right side is interpreted as ∞ if $f_{jj}^* = 1$. Therefore we observe that $f_{jj}^* = 1$ if and only if the expected number of returns to state j is infinite.

Now we classify whether return is certain or not, and when return is certain, whether the mean time of return is finite or infinite.

Definition 1.2.13 (1) State j is called recurrent if $f_{jj}^* = 1$. If $f_{jj}^* < 1$ then j is called transient.

(2) A recurrent state j is said to be positive-recurrent if $\mathbb{E}_j[T_j] < \infty$. Otherwise it is called null-recurrent.

State j is recurrent if and only if, starting from j , the probability of returning to j is unity. Because $f_{jj}^* = 1 \Leftrightarrow p_{jj}^* = \infty$, by (1.27) the expected number of returns to j is infinite. Thus state j is recurrent $\Leftrightarrow \mathbb{E}_j[N_j] = \infty \Leftrightarrow \mathbb{P}_j[N_j = \infty] = 1$. If state j is transient then $f_{jj}^* < 1$ and there is a positive probability $1 - f_{jj}^*$ of never returning to j . Therefore state j is transient $\Leftrightarrow \mathbb{E}_j[N_j] < \infty \Leftrightarrow \mathbb{P}_j[N_j < \infty] = 1$.

Recurrent states can be classified further as follows:

Definition 1.2.14 A recurrent state j is said to be periodic with period d , if $d \geq 2$ is the greatest common divisor of all integers $n \geq 1$ for which $p_{jj}(n) > 0$ ⁵. If there is no such $d \geq 2$, then j is called aperiodic.

Remark 1.2.2 There is no periodicity for continuous time Markov chains.

Definition 1.2.15 State j is said to be accessible from state i , in which case we write $i \rightarrow j$, if there is some integer $n \geq 0$ such that $p_{ij}(n) > 0$. Thus, for $i \neq j$, $i \rightarrow j$ if and only if $p_{ij}^* = \sum_{n=0}^{\infty} p_{ij}(n) > 0$ and hence if and only if $f_{ij}^* > 0$.

Accessibility does not concern the actual value of the probability, but only whether or not it is zero.

Definition 1.2.16 (1) A set of states is called closed if no state outside is accessible from any state inside it.

(2) A state forming a closed set by itself is called absorbing.

(3) A closed set is said to be irreducible if no proper subset of it is closed.

(4) A Markov chain is called irreducible if its only closed set is the set of all states.

Some consequences of Definition (1.2.15) are the following: A state j is absorbing if and only if $p_{jj} = 1$. A Markov chain is irreducible if all states are accessible from another. Otherwise the chain can be splitted into disjuncted classes and is reducible.

For an irreducible Markov chain we define

⁵If $p_{jj}(n) = 0$ for all $n \geq 1$, we define $d = 0$.

Definition 1.2.17 An irreducible Markov chain is called *positive recurrent* (null recurrent or transient respectively) if all states are positive recurrent (null recurrent or transient). A recurrent Markov chain is called *aperiodic* or *periodic with period d* if all the states are aperiodic or periodic with period d respectively.

If we combine some of the last definitions we get

Definition 1.2.18 A Markov chain is called *ergodic* if it is irreducible, positive recurrent and aperiodic.

Then the limit of $p_{ij}(n)$ as $n \rightarrow \infty$ exists which is positive and independent of the initial state i . For more informations about the classification of states lookup [Kijima].

1.2.5 Time reversible Markov chains

We begin with the following definition:

Definition 1.2.19 A continuous-time Markov chain is called *uniformizable* if its infinitesimal generator $Q = (q_{ij})$ is stable and conservative and satisfies $\sup_i q_i < \infty$, where $q_i = -q_{ii}$.

Let X_t be a uniformizable Markov chain with infinitesimal generator $Q = (q_{ij})$ and let $c = \sup_i q_i$. For any $\nu \geq c$, we define

$$P_\nu = I + \frac{1}{\nu}Q. \quad (1.28)$$

The off-diagonal elements of Q are all nonnegative and the diagonal elements of Q/ν are not less than -1. Further $Q\mathbf{1} = 0$ since Q is conservative. It follows that P_ν is a stochastic matrix for any $\nu \geq c$.

Now we assume that the state space is given by $\mathcal{N} = (1, \dots, N)$ where $N < \infty$. Since we can always choose $c = \max_{0 \leq i \leq N} q_i < \infty$, any finite Markov chain in continuous time is uniformizable.

For an ergodic Markov chain X_t with infinitesimal generator $Q = (q_{ij})$, let $\nu \geq c = \max_{0 \leq i \leq N} |q_{ii}|$ and define P_ν as in (1.28). We suppose that the following equation called the detailed balance equation holds for P_ν , i.e.,

$$\pi_i p_{ij}^\nu = \pi_j p_{ji}^\nu, \quad i, j \in \mathcal{N},$$

where $\pi = (\pi_i)$ is the stationary distribution of P_ν . Because

$$p_{ij}^\nu = \delta_{ij} + \frac{1}{\nu}q_{ij}, \quad i, j \in \mathcal{N},$$

the detailed balance equation can be written as

$$\pi_i q_{ij} = \pi_j q_{ji}, \quad i, j \in \mathcal{N}. \quad (1.29)$$

The left side describes the transition intensity from state i to state j , the right side describes the transition intensity from state j to state i . Let π_D denote the diagonal matrix with diagonal elements π_i and define

$$Q_R = \pi_D^{-1} Q^T \pi_D. \quad (1.30)$$

The diagonal elements of Q_R are negative, the off-diagonal elements of Q_R are nonnegative, and since $\pi^T Q = \mathbf{0}^T$ we have

$$Q_R \mathbf{1} = \pi_D^{-1} Q^T \pi = \pi_D^{-1} (\pi^T Q)^T = \mathbf{0}.$$

The matrix Q_R can be considered as an infinitesimal generator of a continuous-time Markov chain on \mathcal{N} . The generator Q_R is called dual of Q since $(Q_R)_R = Q$. Now, in matrix notation the detailed balance equation says that $\pi_D Q$ is symmetric, i.e.,

$$\pi_D Q = (\pi_D Q)^T = Q^T \pi_D,$$

so that $Q = Q_R$.

Definition 1.2.20 *An ergodic Markov chain or its infinitesimal generator Q is said to be reversible in time if Q and its dual Q_R are identical.*

Theorem 1.2.8 *Suppose that the Markov chain X_t is ergodic and reversible in time, then it follows that $\pi_D P(t)$ is symmetric for all $t \geq 0$:*

$$\pi_i p_{ij}(t) = \pi_j p_{ji}(t), \quad i, j \in \mathcal{N}; \quad t \geq 0.$$

Theorem 1.2.9 (Kolmogorov criterion) *An ergodic Markov chain with infinitesimal generator $Q = (q_{ij})$ is reversible in time if and only if*

$$q_{i,i_1} q_{i_1,i_2} \cdots q_{i_k,i} = q_{i,i_k} q_{i_k,i_{k-1}} \cdots q_{i_1,i}$$

for all $k \geq 2$ and all states i, i_1, \dots, i_k .

Proof 1.2.7 *For both theorems lookup [Kijima].*

1.2.6 Calculation of transition probability functions

This section is quite useful when trying to simulate a continuous time Markov chain and helps us to show the behavior of $P(t)$ for $t \rightarrow \infty$.

We consider an irreducible Markov chain X in continuous time with state space $\mathcal{N} = (0, 1, \dots, N)$, generator $Q = (q_{ij})$. The transition matrix function $P(t)$ is given by

$$P(t) = \exp(Qt) = \sum_{n=0}^{\infty} \frac{Q^n t^n}{n!}, \quad t \geq 0.$$

Now we use some Linear Algebra.

Let $\lambda_j, j = 0, 1, \dots, N$ be the eigenvalues of Q and suppose that Q can be written as

$$Q = M \lambda_D M^{-1} \tag{1.31}$$

with some nonsingular matrix M . λ_D is the diagonal matrix with diagonal elements λ_j . If such a matrix M exists, we say that the generator Q is diagonalizable. Let v_j be the eigenvector associated with λ_j . i.e.,

$$Q v_j = \lambda_j v_j, \quad j = 0, 1, \dots, N.$$

When we write $M = (v_0, \dots, v_N)$ we have

$$QM = M \lambda_D.$$

The set of eigenvectors $(v_i)_{i=0,1,\dots,N}$ is linearly independent if and only if M is nonsingular. Thus Q is diagonalizable if and only if Q has a linearly independent set of N eigenvectors. A sufficient condition for this is that the eigenvectors are all distinct.

If (1.31) holds then

$$Q^n = M \lambda_D^n M^{-1} \quad n = 0, 1, \dots,$$

so that

$$P(t) = \sum_{n=0}^{\infty} \frac{Q^n t^n}{n!} = M \left(\sum_{n=0}^{\infty} \frac{\lambda_D^n t^n}{n!} \right) M^{-1} = M \lambda_D(t) M^{-1}, \tag{1.32}$$

where $\lambda_D(t)$ is the diagonal matrix with diagonal elements $e^{\lambda_j t}$.

Definition 1.2.21 A quadratic matrix A is called symmetrizable if it can be written as $A = SH$, with S being a positive definite matrix and H being a hermite matrix i.e. H is a quadratic, complex matrix with $H = \overline{H}^T$.

If Q is symmetrizable, then the eigenvalues are all real and

$$P(t) = \sum_{j=0}^N e^{\lambda_j t} v_j u_j^T, \quad t \geq 0,$$

where

$$M^{-1} = \begin{pmatrix} \mu_0^T \\ \mu_1^T \\ \vdots \\ \mu_N^T \end{pmatrix}$$

In particular, if X_t is ergodic, then $Q\mathbf{1} = \mathbf{0}$ so that $\lambda_0 = 0, v_0 = \mathbf{1}$ and $u_0 = \pi$ the stationary distribution of X_t .

To compute the eigenvalues we can use the simple relation

$$\sum_{j=0}^N \lambda_j = - \sum_{j=0}^N q_j = \text{tr}(Q),$$

the trace of Q . The eigenvalues of Q^n are the n -th powers of the eigenvalues of Q . So it follows that

$$\text{tr}(Q^n) = \sum_{j=0}^N \lambda_j^n, \quad n = 1, 2, \dots$$

Therefore we can obtain the eigenvalues $\lambda_j, j = 0, 1, \dots, N$ as the solutions of the set of equations

$$\lambda_0^n + \lambda_1^n + \dots + \lambda_N^n = \text{tr}(Q^n), \quad n = 1, \dots, N+1.$$

If the Markov chain X_t is ergodic, we know $\lambda_0 = 0$. Hence in this case N equations suffice to determine the eigenvalues $\lambda_j, j = 1, \dots, N$.

1.3 Brownian motion

A very important example for a stochastic process is Brownian motion. It is used for many financial models, whatever we consider: stocks, currencies or interest rates. The most in this section is taken from [Lamberton/Lapeyre] and [Rogers/Williams]. The interested reader is referred to [Rogers/Williams] for further development of aspects of Brownian motion.

Definition 1.3.1 (Brownian motion) A Brownian motion is a real-valued, continuous stochastic process $B = (B_t)_{t \geq 0}$ with independent and stationary increments. In other words:

- (1) continuity: \mathbb{P} a.s the map $s \rightarrow B_s(\omega)$ is continuous.
- (2) independent increments: If $s \leq t$ then $B_t - B_s$ is independent of $\mathcal{F}_s = \sigma(B_u, u \leq s)$.
- (3) stationary increments: If $s \leq t$ then $B_t - B_s$ and $B_{t-s} - B_0$ have the same probability law.

This definition induces the distribution of the process B .

Theorem 1.3.1 If B is a Brownian motion, then $B_t - B_0$ is a normal random variable with mean rt and variance $\sigma^2 t$ where r and σ are constant real numbers.

Remark 1.3.1 A Brownian motion is called standard if:

$$B_0 = 0 \quad \mathbb{P} \quad a.s. \quad \mathbb{E}(B_t) = 0, \quad \mathbb{E}(B_t^2) = t.$$

In that case, the distribution of X_t is the following:

$$\frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{x^2}{2t}\right) dx,$$

where dx is the Lebesgue measure on \mathbb{R} .

Now we give the definition of a Brownian motion with respect to a filtration $(\mathcal{F}_t)_{t \geq 0}$.

Definition 1.3.2 A real-valued continuous stochastic process is an (\mathcal{F}_t) -Brownian motion if it satisfies:

- (1) For any $t \geq 0$, X_t is (\mathcal{F}_t) -measurable.
- (2) If $s \leq t$ then $B_t - B_s$ is independent of the σ -algebra (\mathcal{F}_s) .
- (3) If $s \leq t$ then $B_t - B_s$ and $B_{t-s} - B_0$ have the same law.

Remark 1.3.2 The first point of this definition shows that $\sigma(X_u, u \leq t) \subset \mathcal{F}_t$. Moreover, it is easy to check that an \mathcal{F}_t -Brownian motion is also a Brownian motion with respect to its natural filtration.

Remark 1.3.3 If the process $(B_t)_{t \geq 0}$ is a Brownian motion then

- (1) the process $(-B_t)_{t \geq 0}$ is a Brownian motion.
- (2) for any $a \geq 0$, the process $(B_{t+a} - B_a)_{t \geq 0}$ is a Brownian motion.
- (3) for any $c \neq 0$, $(cB_{t/c^2})_{t \geq 0}$ is a Brownian motion.
- (4) the process $(\tilde{B}_t)_{t \geq 0}$ defined by $\tilde{B}_0 = 0$, $\tilde{B}_t = tB_t$ for $t > 0$ is a Brownian motion.

Lemma 1.3.1 If the process $(B_t)_{t \geq 0}$ is a Brownian motion then we have

$$\mathbb{P}(\sup_t B_t = +\infty, \quad \inf_t B_t = -\infty) = 1.$$

Proof 1.3.1 Lookup [Rogers/Williams].

1.3.1 Brownian motion as a martingale

Proposition 1.3.1 If $B = (B_t)_{t \geq 0}$ is a standard \mathcal{F}_t -Brownian motion then

- (1) B_t is an \mathcal{F}_t -martingale.
- (2) $B_t^2 - t$ is an \mathcal{F}_t -martingale.
- (3) $\exp(\sigma B_t - (\sigma^2/2)t)$ is an \mathcal{F}_t -martingale.

Proof 1.3.2 ad(1) If $s \leq t$ then $B_t - B_s$ is independent of the σ -algebra \mathcal{F}_s . So $\mathbb{E}[B_t - B_s | \mathcal{F}_s] = \mathbb{E}[B_t - B_s]$ and because B is a standard Brownian motion we have $\mathbb{E}[B_t - B_s] = 0$.
ad(2)

$$\begin{aligned} \mathbb{E}[B_t^2 - B_s^2 | \mathcal{F}_s] &= \mathbb{E}[(B_t - B_s)^2 + 2B_s(B_t - B_s) | \mathcal{F}_s] \\ &= \mathbb{E}[(B_t - B_s)^2 | \mathcal{F}_s] + 2B_s \mathbb{E}[B_t - B_s | \mathcal{F}_s], \end{aligned}$$

and since B is a martingale $\mathbb{E}[B_t - B_s] = 0$, we get

$$\mathbb{E}[B_t^2 - B_s^2 | \mathcal{F}_s] = \mathbb{E}[(B_t - B_s)^2 | \mathcal{F}_s].$$

Because the Brownian motion has stationary and independent increments it follows that

$$\mathbb{E}[(B_t - B_s)^2 | \mathcal{F}_s] = \mathbb{E}[B_{t-s}^2] = t - s.$$

Thus we get $\mathbb{E}[B_t^2 - t | \mathcal{F}_s] = B_s^2 - s$, if $s < t$.

ad(3) If X is a standard normal random variable we know that

$$\mathbb{E}[e^{\lambda X}] = e^{\lambda^2/2}.$$

If $s < t$

$$\mathbb{E}[e^{\sigma B_t - \sigma^2 t/2} | \mathcal{F}_s] = e^{\sigma B_s - \sigma^2 s/2} \mathbb{E}[e^{\sigma(B_t - B_s)} | \mathcal{F}_s],$$

because B_s is \mathcal{F}_s measurable. Since $B_t - B_s$ is independent of \mathcal{F}_s we get that

$$\begin{aligned} \mathbb{E}[e^{\sigma(B_t - B_s)} | \mathcal{F}_s] &= \mathbb{E}[e^{\sigma(B_t - B_s)}] \\ &= \mathbb{E}[e^{\sigma B_{t-s}}] = \mathbb{E}[e^{\sigma x \sqrt{t-s}}] \\ &= \exp\left(\frac{1}{2} \sigma^2 (t-s)\right). \quad \square \end{aligned}$$

Theorem 1.3.2 (Levy) Let $(X_t)_{t \geq 0}$ be a continuous martingale, $X_0 = 0$, and suppose that $X_t^2 - t$ is a martingale. Then X is a Brownian motion.

Proof 1.3.3 Lookup [Rogers/Williams]

1.3.2 Brownian motion as a Markov process

Brownian motion is a (time-homogeneous) Markov process, for any bounded Borel $f : \mathbb{R} \rightarrow \mathbb{R}$, and $s, t \geq 0$,

$$\mathbb{E}[f(B_{t+s}) | \mathcal{B}_s] = P_t f(B_s), \quad (1.33)$$

where the transition semigroup $(P_t)_{t \geq 0}$ is defined by

$$P_t f(x) = \begin{cases} \int_{-\infty}^x p_t(x, y) f(y) dy & (t > 0) \\ f(x) & (t = 0) \end{cases}$$

where

$$p_t(x, y) = (2\pi t)^{-1/2} \exp\left[-\frac{(x-y)^2}{2t}\right] \quad (1.34)$$

is the Brownian transition density. The Markov property (1.33) is immediate from the definition of Brownian motion. It is easy to confirm that $(P_t)_{t \geq 0}$ is a semigroup:

$$P_{t+s} = P_t P_s = P_s P_t \quad (s, t \geq 0), \quad (1.35)$$

the so called Chapman-Kolmogorov equations. The semigroup property suggests that we have

$$\frac{d}{dt} P_t = \lim_{s \rightarrow 0} \frac{1}{s} (P_{t+s} - P_t) = P_t G = G P_t,$$

where

$$G = \lim_{s \rightarrow 0} \frac{1}{s} (P_s - I)$$

is the infinitesimal Generator of $(P_t)_{t \geq 0}$. Hence that G is not defined for all functions. We will now work out the generator Gf of Brownian motion for suitable f by

$$Gf = \lim_{t \rightarrow 0} \frac{1}{t} (P_t f - f),$$

if $f \in C_b^2(\mathbb{R})$ then

$$\begin{aligned} \lim_{t \rightarrow 0} (P_t f - f)(x) &= \lim_{t \rightarrow 0} \int_{-\infty}^{\infty} \frac{f(x + y\sqrt{t}) - f(x)}{t} \exp\left(-\frac{1}{2}y^2\right) \frac{dy}{\sqrt{2\pi}} \\ &= \lim_{t \rightarrow 0} \int_{-\infty}^{\infty} \frac{1}{t} \left\{ y\sqrt{t} f'(x) + \frac{1}{2} y^2 t f''(x + \theta y\sqrt{t}) \right\} \exp\left(-\frac{1}{2}y^2\right) \frac{dy}{\sqrt{2\pi}} \\ &= \frac{1}{2} f''(x). \end{aligned}$$

where $\theta \in (0, 1)$ depends on $y\sqrt{t}$.

Therefore the infinitesimal generator of Brownian motion is

$$G = \frac{1}{2} \frac{d^2}{dx^2}. \quad (1.36)$$

1.3.3 Brownian motion as a diffusion

For the exact definition of a diffusion lookup the Appendix. Heuristically a diffusion is a continuous time-homogeneous Markov process X that is characterised by its local infinitesimal drift b and variance a : for small h ,

$$\begin{aligned} \mathbb{E}[X_{t+h} - X_t | \mathcal{F}_t] &= hb(X_t), \\ \mathbb{E}[\{X_{t+h} - X_t - hb(X_t)\}^2 | \mathcal{F}_t] &= ha(X_t), \end{aligned}$$

If a and b were constant functions then

$$X_t = \sigma B_t + b_t, \quad \sigma = a^{1/2},$$

would satisfy the above description. By Levy's theorem Brownian motion is described by $b = 0, a = 1$.

2 Introduction to interest rate theory

2.1 Bond prices and interest rates

The concept of interest rates belongs to our every-day life. When depositing an amount of money in a bank account, everybody expects that the amount grows (at some rate) with time. Lending money must be rewarded somehow so that an amount of money today is not equivalent to the same amount tomorrow.

2.1.1 Short-term interest rate

We will now define such concepts in mathematical terms. Most traditional stochastic interest rate models are based on the exogenous specification of a short-term rate of interest.

Definition 2.1.1 (Instantaneous short-rate) *On a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and for some $T^* \geq 0$ we define a stochastic process r_t to denote the instantaneous interest rate (also referred to as a short-term interest rate or spot interest rate) for risk free borrowing or lending prevailing at time t over the infinitesimal time interval $[t, t + dt]$. We assume throughout that r_t is an adapted stochastic process with almost all sample paths integrable on $[0, T^*]$ with respect to the Lebesgue measure.*

In the following we will refer to this just as *spot-rate* or *short-rate*.

Definition 2.1.2 (Bank account) *We define C_t to be the value of a bank account at time $t \geq 0$, an adapted process of finite variation and with continuous sample path. We assume $C_0 = 1$ and that the bank account evolves according to the following differential equation:*

$$dC_t = r_t C_t dt, \quad C_0 = 1, \quad (2.1)$$

where r_t is a function of time. It follows that for almost all $\omega \in \Omega$ this equation can be solved by the function $C_t = C_t(\omega)$

$$C_t = \exp\left(\int_0^t r_s ds\right), \quad \forall t \in [0, T^*]. \quad (2.2)$$

In financial interpretation C_t represents the price process of a riskless asset which continuously compounds in value at rate r_t .

Assuming in general that r_t is a stochastic process leads to the following.

Definition 2.1.3 (Stochastic discount factor) *The stochastic discount factor $D(t, T)$ between two instants t and T is the amount at time t that is "equivalent" to one unit payable at time T , and given by*

$$D(t, T) = \frac{C_t}{C_T} = \exp\left(-\int_t^T r_s ds\right). \quad (2.3)$$

2.1.2 Zero-coupon bonds

We now take our focus on a very simple financial instrument, a bond. The buyer of a bond pays the issuer an initial price P in return for a predetermined sequence of coupon payments. We will just refer to government bonds which are, in opposite to corporate bonds (emitted by companies), securitized (no risk to default).

Definition 2.1.4 (Zero-Coupon Bond) *Let T^* be a fixed time horizon for all market activities. A zero-coupon bond (pure discount bond) is a contract that guarantees its holder the payment of one unit of cash at time T , without intermediate coupon payments. The price of a zero-bond at time $t < T$ is denoted by $P(t, T)$. Clearly $P(T, T) = 1$ for all $T \leq T^*$*

We will usually assume that for any fixed maturity $T \leq T^*$, the bond price $P(\cdot, T)$ follows a strictly positive and adapted process on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

When evaluating the price of a bond, the following is of high importance.

Definition 2.1.5 (Time to maturity) *The time to maturity $T - t$ is the amount of time (in years) from the present time to the maturity time $T > t$.*

There are several conventions that can be made called the *year fraction* and the *day-count convention*. For more details lookup the Appendix.

When we consider a zero bond with a fixed time horizon $T \leq T^*$, the simple rate of return from holding the bond over the period $[t, T]$ is equal to

$$\frac{1 - P(t, T)}{P(t, T)} = \frac{1}{P(t, T)} - 1.$$

The equivalent rate of return with continuous compounding is commonly referred to as the *yield to maturity* on a bond. Formally, we have following definition.

Definition 2.1.6 *An adapted process $Y(t, T)$ defined by*

$$Y(t, T) = -\frac{1}{T - t} \ln P(t, T), \quad \forall t \in [0, T] \quad (2.4)$$

is called the annualized yield to maturity on a zero-bond maturing at time T .

2.1.3 Forward interest rates

Forward rates are characterized by three time instants. The time t at which the rate is considered, its expiry T and its maturity U , with $t \leq T \leq U$. They are defined through a prototypical *Forward-rate agreement* which is widely used by companies to hedge their interest rate risk if they consider to borrow cash in the future.

Forward rates are interest rates that can be locked in today for an investment in a future time period. Their values can be derived directly from the prices of zero-coupon bonds.

Definition 2.1.7 (Continuously compounded forward rate) *We define $f(t, T, U)$ to be the Continuously compounded forward rate at time t for the expiry $T > t$ and the maturity $U > T$.*

We must have

$$e^{Y(t, U)(U-t)} = e^{Y(t, T)(T-t)} e^{f(t, T, U)(U-T)} \quad (2.5)$$

so that

$$f(t, T, U) = \frac{1}{U - T} (Y(t, U)(U - t) - Y(t, T)(T - t)) \quad (2.6)$$

or, in terms of bond prices

$$f(t, T, U) = \frac{\ln P(t, T) - \ln P(t, U)}{U - T} \quad (2.7)$$

otherwise an arbitrage would be possible.

If we take the $\lim_{U \rightarrow T}$ in the last expression we get the following definition.

Definition 2.1.8 (Instantaneous forward rate) *The instantaneous forward rate of maturity T is*

$$f(t, T) = f(t, T, T) = \lim_{U \rightarrow T} f(t, T, U),$$

and, by the definition of the differential (under suitable technical assumptions) we have

$$f(t, T) = -\frac{\partial \ln P(t, T)}{\partial T}. \quad (2.8)$$

The concept of an instantaneous forward rate is a mathematical idealization rather than a quantity observable in practice. A widely accepted approach to bond prices due to Heath, Jarrow and Morton is based on a family $f(t, T), t \leq T \leq U$, of forward rates. Given such a family $f(t, T)$, the bond prices are then defined by setting

$$P(t, T) = \exp\left(-\int_t^T f(t, u) du\right), \quad \forall t \in [0, T]. \quad (2.9)$$

Remark 2.1.1 *Observe that $Y(t, T) = f(t, t, T)$ as expected. Investing at time t in a T -maturity bond is clearly equivalent to lending money over the period $[t, T]$. The connection to the spot-rate is then given by $r_t = f(t, t) = Y(t, t)$.*

2.1.4 Arbitrage-free family of bond prices

We recall the definition of a zero-coupon bond. In our framework we assume that for any fixed maturity $T \leq T^*$ the price process $P(t, T)$, $t \in [0, T]$ follows a strictly positive and adapted process on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The first question we will consider is the absence of arbitrage between all bonds with different maturities and a savings account (riskless investment). An adapted process r_t on $(\Omega, \mathcal{F}, \mathbb{P})$ models the short-term interest rate, meaning that the savings account satisfies

$$C_t = \exp\left(\int_0^t r_s ds\right), \quad \forall t \in [0, T^*].$$

Definition 2.1.9 (Arbitrage-free family of bond prices) *A family $P(t, T), t \leq T \leq T^*$, of adapted processes is called an arbitrage-free family of bond prices if the following conditions are satisfied:*

- (a) $P(T, T) = 1$, for all $T \in [0, T^*]$ and
- (b) *There exists a probability measure \mathbb{P}^* on $(\Omega, \mathcal{F}_{T^*})$ equivalent to \mathbb{P} , such that for any maturity $T \in [0, T^*]$ the relative bond price*

$$Z^*(t, T) = \frac{P(t, T)}{C_t}, \quad \forall t \in [0, T^*], \quad (2.10)$$

follows a martingale under \mathbb{P}^* .

Any probability measure \mathbb{P}^* of this definition is called a martingale measure for the family $P(t, T)$. Under such an assumption we trivially have $Z^*(t, T) = \mathbb{E}_{P^*}(Z^*(T, T)|\mathcal{F}_t)$ for $t \leq T$ so that the bond price satisfies

$$P(t, T) = C_t \mathbb{E}_{P^*}(C_T^{-1} | \mathcal{F}_t), \quad \forall t \in [0, T]. \quad (2.11)$$

So, for any martingale measure P^* of an arbitrage-free family of bond prices, we have

$$P(t, T) = \mathbb{E}_{P^*}(e^{-\int_t^T r_s ds} | \mathcal{F}_t), \quad \forall t \in [0, T]. \quad (2.12)$$

Conversely, given any nonnegative short rate r_t on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and any probability measure \mathbb{P}^* on $(\Omega, \mathcal{F}_{T^*})$ equivalent to \mathbb{P} , the family $P(t, T)$ given by Equation (2.12) is easily seen to be an arbitrage-free family of bond prices relative to r . If a family $P(t, T)$ satisfies Definition (2.10), then necessarily the bond price process $P(\cdot, T)$ is a \mathbb{P}^* -semimartingale, as a product of a martingale and a process of finite variation (that is, a product of two \mathbb{P}^* -semimartingales). Therefore it is also a \mathbb{P} -semimartingale since \mathbb{P} and \mathbb{P}^* are assumed to be equivalent.

2.2 The term structure of interest rates

The *term structure of interest rates*, also known as the *yield curve*, is the function that relates the yield $Y(t, T)$ to maturity T . When we look back at Definition (2.4) we have defined a process $Y(t, T)$, the yield to maturity. For an arbitrary fixed maturity date T , there is a one-to-one correspondence between $P(t, T)$ and $Y(t, T)$. Given the yield to maturity process, the corresponding bond price is uniquely defined by

$$P(t, T) = e^{-Y(t, T)(T-t)}, \quad \forall t \in [0, T]. \quad (2.13)$$

The discount factor relates the discounted bond price $P(t, T)$ to maturity T . At a theoretical level the initial term structure of interest rates may be presented by the current bond prices $P(0, T)$, or by the initial yield curve $Y(0, T)$, as

$$P(0, T) = e^{-Y(0, T)T}, \quad \forall t \in [0, T^*]. \quad (2.14)$$

In practice, the term structure of interest rates is derived from the prices of several actively traded interest rate instruments, such as zero-coupon bonds, swaps and futures. The yield curve at any given day is determined by market prices for that day. Historically the yield curve and the yield volatilities vary over time. This shows how difficult it is to find a reliable model for the stochastic behavior.

2.2.1 The typical shape of the yield curve

The term structure of interests is very important for economists who want to get an idea of the future development of economy and financial markets, estimate the risk they carry and find strategies for investment. So they are very interested in the shape of the yield curve today to "predict" the future because their investments are very sensitive on changes in the interest rate.

Following shapes have been observed:

- **Normal yield curve** This means that the yields rise with maturity. The positive slope reflects the expectations for economy to grow in the future and also the need of a risk premium (liquidity premium) for longer maturities.

- **Flat yield curve** A flat yield curve means that the yields are independent to the maturity. Such a shape is a sign of uncertainty in economy. Under the assumption that a risk premium is paid it means that yields are expected to be falling.
- **Humped yield curve** This shape occurs when short-term and long-term yields are equal but medium-term yields are higher. It is also sometimes called the irregularly shape.
- **Inverted yield curve** An inverted yield curve occurs when long-term yields are lower than short-term yields. This means that the market is expecting yields to fall sharply, so the investors prefer to invest in long-term maturities. Because of the higher demand on the long end, long-term yields fall.
In the past this shape has often been an indicator of a recession.

Figure (2.1)-(2.4) show us the different shapes of the yield curve. There are three facts about the term structure of interest rates that empirical studies have shown.

- Interest rates on bonds of different maturities tend to move together over time.
- Yield curves usually slope upwards.
- When short term interests are low, yield curves are more likely to have a steep upward slope whereas when short term interest rates are high, yield curves are more likely to be inverted.

Now we present some economic theories which try to explain this behavior.

2.2.2 Theoretical Models

There are four main economic theories attempting to explain how yields vary with maturity. Two of the theories are extreme positions, while the others attempt to find a middle ground between the former two. All of these are in general not arbitrage free.

The fifth model presents the more modern idea of an arbitrage free market.

The following should just give an introduction and oversight. For more information lookup [Cairns] and [Mishkin].

We will now see what facts about the term structure the models are able to explain, and in which they fail.

- **The (pure) expectations hypothesis**

This hypothesis assumes, that the different maturities are perfect substitutes and suggests that the shape of the yield curve depends on the market participants expectations on future interest rates. For example, if investors have an expectation of what 1-year interest rates will be next year, the 2-year interest rate can be calculated as the compounding of this year's interest rate by next year's interest rate. Information is fully available for all subjects, which all act risk averse.

This leads to following result:

- If investors expect growing yields on the market, they prefer short investments and wait until the yields grow to invest in long-term securities. So the demand on the short end of the yield curve increases, which reduces the yield for short maturities. For long-term maturities the supply is higher than the demand which leads to higher yields on the long end. The resulting yield curve is growing with maturity (*normal yield curve*).

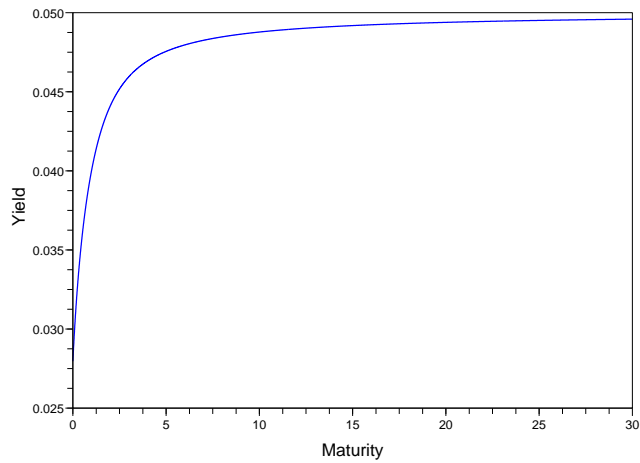


Figure 2.1: Normal term structure

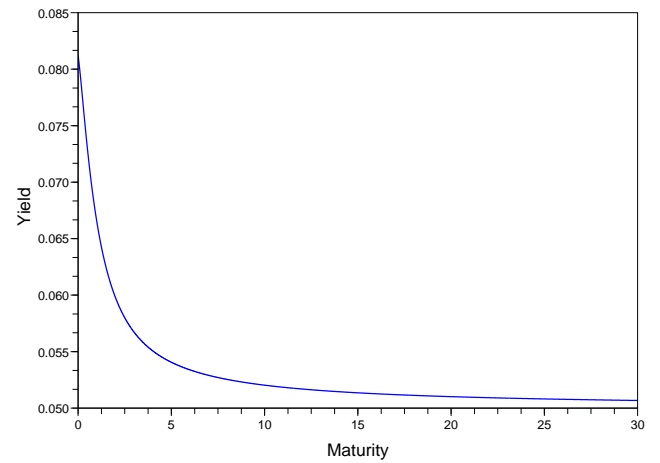


Figure 2.2: Inverse term structure

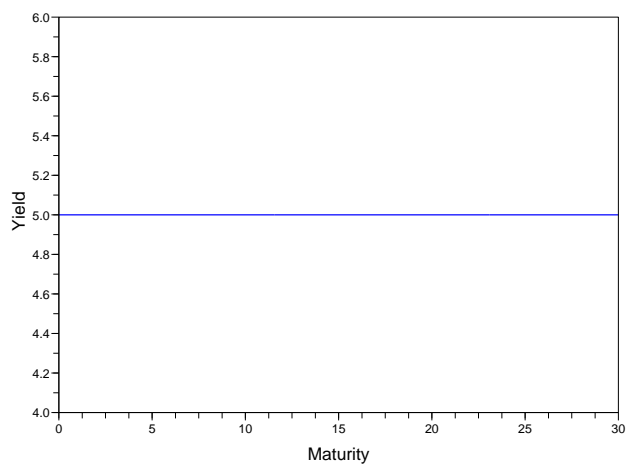


Figure 2.3: Flat term structure

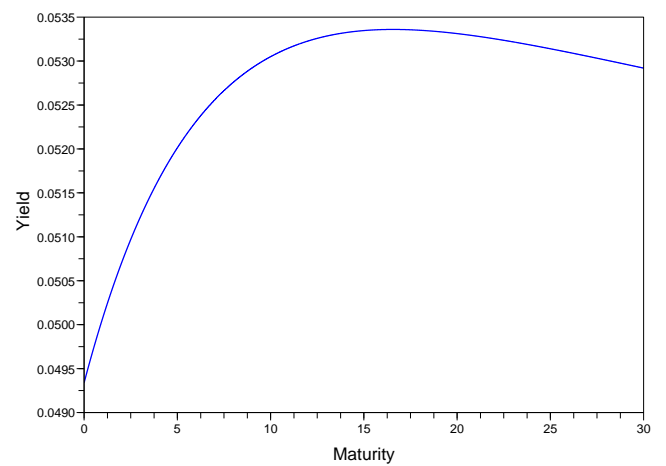


Figure 2.4: Humped term structure

- If falling yields are expected the opposite effect occurs. Investors would prefer the longer maturities which reduces the yields on the long end. The effect is the *inverted yield curve*, which often occurred in the past before recessions.
- If no change of yields is expected then the long-term yields are equal to short-term yields. The resulting yield curve is flat.

The expectations hypothesis explains why in a time of high yields the structure is often inverted and in a time of low yields the structure is normal (increasing).

It does not consider the fact that investments in longer maturities keep a higher risk than short ones.

- **Market segmentation theory**

This theory is also called the segmented market hypothesis. In this theory, financial instruments (bonds) of different terms are not suitable because each investor has in mind an appropriate set of bonds and maturities that are suitable for his purpose. For example, life insurance companies require long-term bonds to match their long-term liabilities. In contrast banks are likely to prefer short-term bonds to reflect the needs of their costumers.

As a result, the supply and demand in the markets for short-term and long-term instruments is determined independently. This behavior of the investors can lead to different yields in the different segments which can lead to a *humped yield curve*.

The market segmentation theory can also explain the *normal yield curve*:

Due to the fact investors prefer their portfolio to be liquid, they will prefer short-term instruments to long-term instruments. Therefore, the market for short term instruments will receive a higher demand. Higher demand for the instrument implies higher prices and lower yield. This explains why short-term yields are usually lower than long-term yields.

The theory fails to explain the observed fact that yields tend to move together (i.e. upward and downward shifts in the curve), because supply and demand of different markets (for example short-term and long-term bonds) are said to be independent.

- **The liquidity preference theory**

The background of this theory is that investors usually prefer short-term investments to long-term investments. They do not like to tie up their capital for too long. Investors will only invest in securities of longer maturity if they have a higher expected return, often referred to as the *risk premium* or *liquidity premium*, to offset the higher risk. Because of the premium for longer terms, the long-term bond yields tend to be higher than the short-term yields, and the yield curve slopes upwards.

We can see, a combination of the expectations hypothesis and the liquidity preference theory might explain what happens on the market:

- If the term structure is slowly growing we can say that for long-term investments only the liquidity premium is paid, and the market is not expecting increasing yields.
- A fast growing term structure means, the market expects increasing yields. More than only the liquidity premium is paid for long-term investments.

The liquidity preference theory on its own cannot explain an inverse term structure of interest rates.

- **Preferred habitat theory**

This theory states that, in difference to the liquidity preference theory (investors prefer short-term investments because they do not like to tie their capital up for too long), that investors prefer a special maturity for their investments in addition to interest rate expectations. They will only invest in other maturities outside their habitat if they receive an additional premium.

So, in this theory bonds of different maturities are substitutes for each other. But this degree of substitutability is lower than that assumed by the pure expectations theory because of the riskiness associated with leaving one's preferred habitat. This riskiness can be attributed not only to the possibilities of capital losses but also to the risk associated with income losses. In this later case, an investor would need to be compensated with a premium above the long-term rate in order to hold a shorter-term security.

The key difference between the preferred habitat and the liquidity preference theories is that the latter assumes that the investment with minimum risk is a short-term asset, whereas the former assumes that the least risky investment is the one whose maturity structure the investor prefers. We can say the liquidity preference theory is a special case of the preferred habitat theory, where the preferred habitat is in short-term bonds. In other words, the only risk of concern is capital risk.

With the help of the preferred habitat theory any occurring shape or change in the term structure can be explained.

These were the four main theories. Now we will just have a short look on another theory. For more information see [Cairns].

- **The arbitrage free pricing theory**

We consider the pricing of bonds in a market which is free of arbitrage. This theory (which is very extensive) pulls together the market expectations, liquidity preference and market segmentation theories in a mathematically precise way. In this framework we can usually decompose forward rates into three components.

- the expected future risk free rate of interest $r(t)$;
- an adjustment for the market price of risk¹;
- a convexity adjustment to reflect the fact that $\mathbb{E}(e^X) \geq e^{\mathbb{E}(X)}$.

The form of the two adjustments is not obvious. This is why we need arbitrage free pricing theory to derive prices.

For single-factor models with just one source of randomness, for example the Vasicek Model, there is no place for market segmentation theory. This theory can be incorporated in multifactor-models (more than one random factor) to some extent.

¹under the "real-world" probability measure \mathbb{P} .

3 The potential approach

3.1 Introduction

When we have a look at financial literature, most term structure models to date fall into one of two classes, either one models the short (spot)-term interest rate, as in Vasicek or Cox, Ingersoll and Ross for example, or alternatively one models the process of forward rates, as in Ho and Lee or Heath, Jarrow and Morton.

Another class are the so-called *market models* which try to find suitable processes to fit with the prices of various derivatives. For more information about this different models look up [Musielà] or [Brigo]. Here we will discuss the considerable advantages of another approach, called the *potential approach*. The key element of this approach by [Rogers 1] is to view the state price density process ζ (a positive supermartingale) as the modeling primitive, and to express the prices of derivatives directly in terms of this. For example the price of a zero-coupon bond which pays 1 at time T has the simple form

$$P(t, T) = \tilde{\mathbb{E}}_t^1[\zeta_T]/\zeta_t, \quad 0 \leq t \leq T.$$

$\tilde{\mathbb{P}}$ is a reference measure. Furthermore Rogers shows how easy yield curves in many countries can be handled together with the exchange rates between them. He proves that assuming complete markets the exchange rate between two currencies must be the ratio of their state-price densities. Namely

$$Y_t^{ij} = Y_0^{ij} \zeta_t^j / \zeta_t^i,$$

where Y_t^{ij} denotes the currency j 's value at time t in currency i .

To get a wide family of models we need to find a way of generating positive supermartingales. Therefore we use classical theory of Markov processes. We take a Markov process $(X_t, t \geq 0)$ with resolvent $(R_\alpha)_{\alpha \geq 0}$. Then we define the state-price density to be

$$\zeta_t = e^{-\alpha t} R_\alpha g(X_t).$$

This defines a positive supermartingale whenever the function g is positive and suitable integrable. When we think of the fact that, as the maturity T of the bond tends to infinity the price tends to zero, we get that the process ζ has to fulfill the further condition

$$\lim_{t \rightarrow \infty} \mathbb{E}[\zeta_t] = 0.$$

A positive supermartingale satisfying this additional condition is called a *potential*, whence the name of this approach.

There are a few examples where the resolvent can be written down in closed form. In the context of a finite Markov chain with finite state space and infinitesimal generator G the resolvent is

$$R_\alpha = (\alpha - G)^{-1}.$$

When we regard the transition semigroup $P(t)_{t \geq 0}$ as semigroup of matrices acting on the vector space \mathbb{R}^I and $P(t) = \exp(tQ)$. Therefore we can simply build examples by choosing

¹ \mathbb{E}_t is short for the conditional expectation $\mathbb{E}(\cdot | \mathcal{F}_t)$

$\alpha \geq 0$ and a positive $f = R_\alpha g$ and then recover g by $g = (\alpha - G)f$ with the demand that g should be non-negative because we are modeling nominal interest rates. Within this framework prices of zero-coupon bonds can simply be written down by

$$P(0, t) = \frac{\tilde{\mathbb{E}}_0[e^{-\alpha t} R_\alpha g(X_t)]}{R_\alpha g(X_0)} = \frac{e^{-\alpha t} \tilde{\mathbb{E}}_0[f(X_t)]}{f(X_0)},$$

where $\tilde{\mathbb{P}}^x$ denotes the law of the process started at x . Furthermore we can define the spot-rate process r_t by

$$r_t = \frac{g(X_t)}{R_\alpha g(X_t)}.$$

The main aim of this chapter is first to develop the basic theory and show two different ways of an approach. Then we discuss two special Markov processes (diffusions in \mathbb{R}^d) and try to come up with closed form expressions for the spot-rate, the state-price density and bond prices.

The most in this chapter is taken from [Rogers 1] and [Rogers 2].

3.2 Pricing

In a filtered probability space $(\Omega, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ we consider *market pricing operators* $(\pi_{st})_{0 \leq s \leq t \leq T}$ for contingent claims:

$$\pi_{st} : L^\infty(\mathcal{F}_t) \rightarrow L^\infty(\mathcal{F}_s), \quad (0 \leq s \leq t).^2$$

If Y is some bounded, \mathcal{F}_t -measurable, random variable, then the time- s market price of Y is $\pi_{st}(Y)$, again a random variable (because our observations up to time s affect what we think the price of the claim to be.)

The pricing operators $(\pi_{st})_{0 \leq s \leq t}$ should satisfy certain axioms:

- (A1) Each π_{st} is a bounded positive linear operator from $L^\infty(\mathcal{F}_t)$ to $L^\infty(\mathcal{F}_s)$;
- (A2) If $Y \in L^\infty(\mathcal{F}_t), Y \geq 0$, then

$$\pi_{0t}(Y) = 0 \Leftrightarrow \mathbb{P}(Y > 0) = 0.$$

(no arbitrage)

- (A3) For $0 \leq s \leq t \leq u, Y \in L^\infty(\mathcal{F}_u), X \in L^\infty(\mathcal{F}_t)$,

$$\pi_{su}(XY) = \pi_{st}(X\pi_{tu}(Y)).$$

(intertemporal consistency)

- (A4) If $(Y_n) \in L^\infty(\mathcal{F}_t), |Y_n| \leq 1, Y_n \rightarrow Y$ then $\pi_{st}(Y_n) \rightarrow \pi_{st}(Y)$. (continuity)

Remark 3.2.1 (A1) The price of a non-negative contingent claim will be non-negative, and the price of a linear combination of claims will be the linear combination of their prices.

(A2) No-Arbitrage: A claim that is almost surely worthless when paid, will be almost surely worthless at all earlier times and conversely.

(A3) Intertemporal consistency: The market prices at time s for XY at time u , or for X times the time- t market price for Y at time t , should be the same, for any X which is known at time t .

(A4) This is a natural continuity condition which is needed for technical reasons.

Theorem 3.2.1 Under the axioms (A1)-(A4), there exists a strictly positive process $(\zeta_t)_{t \geq 0}$ such that the pricing operators π_{st} can be expressed as

$$\pi_{st}(Y) = \frac{\mathbb{E}_s[\zeta_t Y]}{\zeta_s}, \quad (0 \leq s \leq t). \quad (3.1)$$

If we also assume

- (A5) For all $(0 \leq s \leq t), \pi_{st}(1) \leq 1$,

(where 1 denotes the constant function identically equal to 1) then ζ is a positive supermartingale:

$$\mathbb{E}_s[\zeta_t] \leq \zeta_s, \quad (0 \leq s \leq t).$$

²For the exact definition of L^∞ lookup the Appendix.

The process ζ is commonly known as the *state-price density*.

Proof 3.2.1 First, for any $T > 0$, the map

$$A \mapsto \pi_{0T}(I_A)$$

defines a non-negative measure on the σ -algebra \mathcal{F}_T . This follows from the linearity and positivity of (A1) and the continuity property (A4). Furthermore, this measure is absolutely continuous with respect to \mathbb{P} , in view of (A2). There is a non-negative \mathcal{F}_T measurable random variable ζ_T such that

$$\pi_{0T}(Y) = \mathbb{E}[\zeta_T Y]$$

for all $Y \in L^\infty(\mathcal{F}_T)$. Also, $\mathbb{P}[\zeta_t > 0] > 0$, because of (A2) again. Now we use the intertemporal consistency condition (A3): we get

$$\pi_{0T}(X_{\pi_{tT}}(Y)) = \mathbb{E}[X\zeta_t\pi_{tT}(Y)] = \pi_{0T}(Y) = \mathbb{E}[XY\zeta_t].$$

Because $Y \in L^\infty(\mathcal{F}_T)$ is arbitrary, we receive that

$$\pi_{tT}(Y) = \mathbb{E}_t[Y\zeta_T]/\zeta_t,$$

as claimed in theorem 3.2.1. The proof of ζ being a supermartingale under (A5) can now be followed straight forward. \square

Remark 3.2.2 Theorem (3.2.1) shows that if we write $Y_t = \pi_{tT}(Y)$ for some fixed $Y \in L^\infty(\mathcal{F}_T)$ then

$$\zeta_t Y_t = \mathbb{E}_t[\zeta_T Y]$$

is a martingale.

3.2.1 Pricing with the spot rate

Under the assumption that the market is arbitrage free, the price of a contingent claim Y at time- t payable at time $T > t$ is

$$\mathbb{E}[\exp(-\int_t^T r_s ds)Y|\mathcal{F}_t], \quad (3.2)$$

where $(r_t)_{t \geq 0}$ is the spot rate of interest and $\mathbb{E}_t = \mathbb{E}[\cdot|\mathcal{F}_t]$ is the conditional expectation under the filtration \mathcal{F}_t , taken with respect to a fixed risk-neutral measure. When pricing a zero bond, where $Y = 1$, we have

$$P(t, T) = \mathbb{E}_t \exp(-\int_t^T r_s ds). \quad (3.3)$$

What we will do now is to model the state-price density process ζ_t directly.

Definition 3.2.1 There is some reference probability measure $\tilde{\mathbb{P}}$ (equivalent to \mathbb{P}) in terms of which

$$\zeta_t = \exp(-\int_0^t r_s ds) \frac{d\mathbb{P}}{d\tilde{\mathbb{P}}} = \exp(-\int_0^t r_s ds) Z_t, \quad (3.4)$$

where Z_t is a positive martingale, which is interpreted as a change of measure, from the reference measure \mathbb{P} to some new pricing probability $\tilde{\mathbb{P}}$. This is an equivalent martingale measure to the original measure because both have the same null sets, and under $\tilde{\mathbb{P}}$ the discounted prices of all traded assets become martingales.

Theorem 3.2.2 *Using the process ζ_t we can write bond prices as*

$$P(t, T) = \frac{\tilde{\mathbb{E}}_t[\zeta_T]}{\zeta_t}. \quad (3.5)$$

Proof 3.2.2 *We use definition 3.3 of ζ_t to compute 3.4. We get*

$$\begin{aligned} P(t, T) &= \frac{\mathbb{E}_t[\exp(-\int_0^T r_s ds) Z_t]}{\exp(-\int_0^t r_s ds) Z_t} = \frac{Z_t \mathbb{E}_t[\exp(-\int_0^t r_s ds) \exp(-\int_t^T r_s ds)]}{Z_t \exp(-\int_0^t r_s ds)} \\ &= \frac{\exp(-\int_0^t r_s ds) \mathbb{E}_t[\exp(-\int_t^T r_s ds)]}{\exp(-\int_0^t r_s ds)} = \mathbb{E}_t[\exp(-\int_t^T r_s ds)]. \quad \square \end{aligned}$$

We compare this result to Theorem (3.2.1) where we got the same result from a no-arbitrage standpoint.

Assuming $r \geq 0$ (which we always do, because we are concerned with nominal interest rates and not real rates) the process ζ_t is a positive supermartingale.

3.3 The potential approach

Theorem (3.2.1) tells us how easy it is to model prices with ζ . When we use the state price density we can write prices of zero bonds in the form

$$P(t, T) = \frac{\tilde{\mathbb{E}}_t[\zeta_T]}{\zeta_t}.$$

When we now think about the fact that as the maturity T of the bond tends to infinity, the current value of it tends to zero (the value of a bond which is never paid back is 0), we find that the process ζ has to satisfy the further condition

$$\lim_{T \rightarrow \infty} \mathbb{E}[\zeta_T] = 0.$$

A positive supermartingale satisfying this condition is called a *potential* because of the very close links with the Markov process concept of a potential, whence the name of this approach.

Satisfying a mild further condition, a potential can be expressed as

$$\zeta_t = \mathbb{E}_t[A_\infty - A_t] = \mathbb{E}_t[A_\infty] - A_t, \quad (3.6)$$

where A is a previsible, increasing and integrable process.

When we look at (3.1) we see that prices are expressed in terms of conditional expectations which we want to be as simple as possible in terms of a few variables. This leads us to the concept of modeling with Markov processes.

3.4 Markov processes and potentials

During the whole section $(X_t)_{t \geq 0}$ denotes a Markov process with state space \mathcal{X} and resolvent $(R_\alpha)_{\alpha \geq 0}$. We now show two representations of the supermartingale ζ as $\zeta_t = h(t, X_t)$.

3.4.1 Generic approach 1

For fixed $g : \mathcal{X} \rightarrow (0, \infty]$ and $\alpha > 0$, we consider the increasing process

$$A_t = \int_0^t e^{-\alpha s} g(X_s) ds. \quad (3.7)$$

This is adapted and continuous (therefore previsible) and under further conditions on g (uniform boundedness) it is also integrable. We use this to build a pricing model for the state-price density:

$$\zeta_t = \mathbb{E} \left[\int_t^\infty e^{-\alpha s} g(X_s) ds \right] = e^{-\alpha t} R_\alpha g(X_t), \quad (3.8)$$

where $R_\alpha g(X_t)$ is the resolvent of the Markov process and equation (3.8) is exactly the definition of the resolvent. Sometimes the state price density is also represented as

$$\zeta_t = \frac{e^{-\alpha t} R_\alpha g(X_t)}{R_\alpha g(X_0)}.$$

This clearly does not change the behavior of the process ζ and norms ζ_0 to 1.

Equation (3.7) also shows us how simple and flexible this approach is. We can choose any non-negative function g on \mathcal{X} and any positive α to build a model. The decomposition (3.6) of ζ into a martingale less an increasing process takes a very simple form. If we interpret the state price density as

$$\zeta_t = \exp\left(-\int_0^t r_s ds\right) Z_t,$$

and apply Itô's formula we get two decompositions of ζ :

$$\begin{aligned} d\zeta_t &= \zeta_t (dM_t - r_t dt) \\ &= dN_t - e^{-\alpha t} g(X_t) dt, \end{aligned}$$

where M and N are two local martingales, so after equating the finite variation parts we get

$$r_t = \frac{g(X_t)}{R_\alpha g(X_t)} \quad (3.9)$$

for the spot-rate process.

In applications it is often very difficult to specify the resolvent of a given Markov process in a usable closed form, but there are a few examples where the resolvent of a given Markov process can be written down in closed form by

$$R_\alpha = (\alpha - G)^{-1}, \quad (3.10)$$

where G is the generator of the Markov process. When we use this relation we can build examples by first choosing a positive function $f : \mathcal{X} \rightarrow (0, \infty)$ and then define g by

$$g = (\alpha - G)f. \quad (3.11)$$

This means we have $R_\alpha g = f$ and if g is non-negative everywhere we have the same result described at (3.9):

$$r_t = \frac{g(X_t)}{R_\alpha g(X_t)} = \frac{g(X_t)}{f(X_t)}. \quad (3.12)$$

The demand that g should be non-negative arises because we are modeling nominal interest rates not real ones.

Hence: if G is a $N \times N$ matrix the resolvent must be read as $R_\alpha = (\alpha I - G)^{-1}$ where I is an identity matrix of size N .

When we compute bond prices within this framework we get

$$P(t, T) = \frac{\mathbb{E}_t[\zeta_T]}{\zeta_t} = \frac{\mathbb{E}_t[e^{-\alpha T} R_\alpha g(X_T)]}{e^{-\alpha t} R_\alpha g(X_t)} = \frac{e^{-\alpha T} \mathbb{E}_t[f(X_T)]}{e^{-\alpha t} f(X_t)}. \quad (3.13)$$

3.4.2 Generic approach 2

Another approach to create examples. Let $(\phi_i)_{i=1, \dots, n}$ be the eigenfunctions of G ,

$$G\phi_i = \lambda_i \phi_i,$$

where the λ_i are the eigenvalues with respect to ϕ_i . Further we choose $g = \sum_{i=1}^n c_i \phi_i$ nonnegative. Then if $\alpha > \max(\lambda_i)$

$$f = R_\alpha g = \sum_{i=1}^n c_i (\alpha - \lambda_i)^{-1} \phi_i$$

is an example of potential type, for which bond prices can simply be written down:

$$P(0, T) = \frac{e^{-\alpha T} \sum_{i=1}^n \tilde{c}_i \exp(\lambda_i T) \phi_i(X_0)}{\sum_{i=1}^n \tilde{c}_i \phi_i(X_0)}, \quad (3.14)$$

where $\tilde{c}_i = c_i / (\alpha - \lambda_i)$.

The processes r_t and ζ_t are defined by f and g as in Generic Approach 1.

On first sight this approach seems to be much easier because we only need to find the eigenfunctions, but do not need to compute the transition densities. But if we want to find the price of a caplet, we will still need to know the law of X_t given starting value X_0 to compute the price.

3.5 Application of Markov chain potential models

In theory we are looking for models which produce simple and closed-form expressions for derivatives, whereas in practice (where prices are computed numerically) we are looking for fast algorithms.

When using the potential approach we see from (3.1) that any pricing calculation is in fact an integration. When we do this numerically we compute a finite weighted sum over the state space of the underlying Markov process. To simplify this we work with a process with finite state space: a Markov chain.

Assuming \mathcal{X} to be a finite set of size N has following advantages:

- (i) The generator of the chain is an $N \times N$ matrix Q . The transition semigroup can be expressed as $P(t) = (p_t(x, y))_{x, y \in \mathcal{X}} = \exp(tQ)$.
- (ii) All calculations are made with finite matrices and are therefore fast.
- (iii) Pricing an American-option is an optimal-stopping problem for a finite Markov-chain and easy to handle.

and disadvantages:

- (i) the size of the parameter space is $\mathcal{O}(N^2)$ and so gets quite large quickly.
- (ii) The model will only admit N possible values for the price.

Point (ii) is quite restrictive. When working with an N -state chain, we can only observe N different yield curves which is simply incompatible with daily observations.

3.6 Hedging in the potential model

Just in short some ideas of hedging. In conventional models the standard way to hedge a derivative is to delta-hedge it. The idea here is to compute the differential of the price of the derivative with respect to the prices of the underlying instruments. For example in the case of a put option we differentiate with respect to the stock price. This tells us how many units of the underlying we have to hold to be secured against moves in the underlying. In a complete market this strategy is exact so the contingent claim can be perfectly replicated³.

When using a Markov chain potential model differentiating has no meaning. The idea that we want to protect our portfolio against moves in the underlying can be used here too.

Assume we have a derivative Z and hedging instruments $z^{(1)}, z^{(2)}, \dots$. If the chain jumps from state i to state j at time t the value changes by $\Delta Z_{ij}(t)$. What we will do now is to hold $w_r(t)$ units of asset r so that

$$\Delta Z_{ij}(t) + \sum_{r=1}^m w_r(t) \Delta z_{ij}^{(r)}(t) = 0 \quad \forall j \quad (X_t = i).$$

Now we are immune against any jumps that occur in the chain. Of course we are not able to know X_t in practice, but this does not change our strategy. We just have to make a portfolio of more hedging assets to be sure that

$$\Delta Z_{ij}(t) + \sum_{r=1}^M w_r(t) \Delta z_{ij}^{(r)}(t) = 0 \quad \forall i, j. \quad (3.15)$$

Applying this to a n -state chain would lead us to hold $\frac{n!}{(n-2)!}$ instruments⁴. So we see that practice is not as simple as theory because we cannot be sure if that many instruments are available on the market.

3.6.1 Mini-example

For illustration we use a Markov chain with just two states, therefore $\mathcal{N} = \{1, 2\}$. We suppose that $Z(t) = 0.9$ if we are in state 1 and $Z(t) = 1.1$ if we are in state 2 for simplicity. Then the hedging instruments are $z^{(1)}(t) = -0.2$ and $z^{(2)}(t) = 0.2$. From (3.15) we get

$$\begin{aligned} 0 &= \Delta Z_{12}(t) + w_1(t) \Delta z^{(1)}(t) + w_2(t) \Delta z^{(2)}(t) \\ 0 &= 0.2 + 1 \cdot (-0.2) + 0 \cdot 0.2 \end{aligned}$$

and

$$\begin{aligned} 0 &= \Delta Z_{21}(t) + w_1(t) \Delta z^{(1)}(t) + w_2(t) \Delta z^{(2)}(t) \\ 0 &= -0.2 + 0 \cdot (-0.2) + 1 \cdot 0.2. \end{aligned}$$

³A replicating strategy means we can build a portfolio to receive a value h at time T .

⁴This is a variation without repetition. Using a 9-state chain would lead us to hold 72 different instruments!

3.7 Examples

Now we present some examples with continuous state-space which will illustrate some of the points we discussed before. For the first four examples we take as the underlying Markov process the Gaussian diffusion in \mathbb{R}^d given by

$$dX_t = dW_t - BX_t dt, \quad (3.16)$$

where B is a general $d \times d$ Matrix. The distribution of X_t is given by $N(e^{-TB}X_0, V_T)$, where we define:

$$V_T = \int_0^T e^{-sB} (e^{-sB})^T ds.$$

For this special process the generator⁵ is given by:

$$Gf(x) = \frac{1}{2} \sum_{i=1}^d \frac{\partial^2 f}{\partial x_i^2}(x) - \sum_{i=1}^d (Bx)_i \frac{\partial f}{\partial x_i}(x). \quad (3.17)$$

Before we use this special process in examples let us just have a short look on the results we have already seen.

Given a positive function $f : \mathcal{X} \rightarrow (0, \infty)$ and taking the resolvent

$$R_\lambda = (\alpha - G)^{-1},$$

where G is the generator of the Markov process, we can define g by

$$g = (\alpha - G)f.$$

So we have $R_\alpha g = f$, and, provided that g is nonnegative we now get:

$$r_t = \frac{(\alpha - G)f(X_t)}{R_\alpha g(X_t)} = \frac{g(X_t)}{f(X_t)}.$$

Using the state price density, we have seen, that we can compute bond prices in the following way:

$$P(0, t) = \tilde{\mathbb{E}}_t[\zeta_t] / \zeta_0,$$

where $\zeta_t = e^{-\alpha} R_\alpha g(X_t)$.

3.7.1 Exponential-linear example

For this example we take $f(x) = \exp(a \cdot x)$ ⁶ as the function in *Generic Approach 1*, where a is some fixed, non zero element in \mathbb{R}^d . First we compute the generator.

The summed product of the first partial derivatives with $(Bx)_i$:

$$\begin{aligned} \sum_{i=1}^d (Bx)_i \frac{\partial f}{\partial x_i}(x) &= (Bx)_1 a_1 e^{a \cdot x} + \dots + (Bx)_d a_d e^{a \cdot x} \\ &= ((Bx)_1 a_1 + \dots + (Bx)_d a_d) e^{a \cdot x} = (a \cdot Bx) e^{a \cdot x}. \end{aligned}$$

⁵Lookup the Appendix for the exact definition.

⁶”.” stands of course for the scalar product

The summed product of the second partial derivatives:

$$\sum_{i=1}^d \frac{\partial^2 f}{\partial x_i^2}(x) = e^{a \cdot x} (a_1^2 + \dots + a_d^2) = |a|^2 e^{a \cdot x}.$$

so

$$Gf(x) = \frac{1}{2}|a|^2 e^{a \cdot x} - (a \cdot Bx) e^{a \cdot x} = e^{a \cdot x} \left(\frac{1}{2}|a|^2 - a \cdot Bx \right).$$

Now we can compute

$$\begin{aligned} g(x) = \alpha f(x) - Gf(x) &= \alpha e^{a \cdot x} - e^{a \cdot x} \left(\frac{1}{2}|a|^2 - a \cdot Bx \right) \\ &= e^{a \cdot x} \left(\alpha - \frac{1}{2}|a|^2 + a \cdot Bx \right). \end{aligned}$$

The spot rate process is given by

$$\begin{aligned} r_t = \frac{g(X_t)}{f(X_t)} &= \frac{e^{a \cdot X_t} \left(\alpha - \frac{1}{2}|a|^2 + a \cdot BX_t \right)}{e^{a \cdot X_t}} \\ &= \alpha - \frac{1}{2}|a|^2 + a \cdot BX_t. \end{aligned}$$

This is a multifactor Gaussian model. Such models are often used because they can easily be fitted to data. We reject such models, because here nominal interest rates can become negative.

3.7.2 Exponential-quadratic example

Here we use *Generic Approach 1* again with following function f:

$$f(x) = \exp\left(\frac{1}{2}(x - c)^T Q (x - c)\right),$$

where $c \in \mathbb{R}^d$ and Q is a $d \times d$ positive-definite and symmetric matrix. Now we have:

$$g(x) = f(x) \left[\frac{1}{2}(x - S^{-1}v)^T S (x - S^{-1}v) + \alpha - \frac{1}{2} \text{tr} Q - \frac{1}{2} |Qc|^2 - \frac{1}{2} v^T S^{-1}v \right],$$

where $S = B^T Q + QB - Q^2$, $v = (B^T - Q)Qc$. We shall take Q small enough to ensure that S is positive-definite, and take

$$\alpha = \frac{1}{2} \text{tr} Q + \frac{1}{2} |Qc|^2 + \frac{1}{2} v^T S^{-1}v,$$

which reduces $g(x)$ to

$$g(x) = f(x) \frac{1}{2} (x - S^{-1}v)^T S (x - S^{-1}v).$$

Then the spot-rate process is simply

$$r_t = \frac{1}{2} (X_t - S^{-1}v)^T S (X_t - S^{-1}v),$$

which is a squared Gaussian process. For the bond-prices we have

$$P(0, T) = e^{-\alpha t} \det(I - QV_T)^{-1/2} \exp \left(\frac{1}{2} \mu_T^T (I - QV_T)^{-1} Q \mu_T - \frac{1}{2} \mu_0^T Q \mu_0 \right),$$

where $\mu_T = e^{-TB} X_0 - c$.

3.7.3 Quadratic example

Again we use *Generic Approach 1* and now we take the function

$$f(x) = \gamma + \frac{1}{2}(x - c)^T Q (x - c),$$

for fixed $c, \gamma \in \mathbb{R}^d$ and Q is again a $d \times d$ positive-definite, symmetric matrix. Then

$$g(x) = \alpha\gamma - \frac{1}{2}\text{tr}Q + \frac{1}{2}\alpha c^T Q c + \frac{1}{2}(x - v)^T S (x - v) - \frac{1}{2}v^T S v,$$

where $S = \alpha Q + B^T Q + Q B$ and $v = S^{-1}(\alpha Q c + B^T Q c)$. When we choose γ so that

$$\gamma = \frac{\text{tr}Q + v^T S v}{2\alpha} - \frac{1}{2}c^T Q c,$$

then the spot-rate process is given by

$$r_t = \frac{(X_t - v)^T S (X_t - v)}{2\gamma + (X_t - c)^T Q (X_t - c)}.$$

The zero-coupon bond prices are given by

$$P(0, T) = \frac{e^{-\alpha T}}{f(X_0)} \left(\gamma + \frac{1}{2}(\text{tr}(Q V_T) + \mu_T^T Q \mu_T) \right),$$

where $\mu_T = e^{-TB} X_0 - c$.

3.7.4 cosh example

For this example we assume $B = \beta I$, and choose f to be

$$f(x) = \cosh \gamma \cdot (x + c),$$

for fixed $c, \gamma \in \mathbb{R}^d$. We now take $d = 1$ (without losing generality) for notional simplicity. Here the generator of the Markov process is

$$Gf(x) = \frac{1}{2}\gamma^2 \cosh(\gamma(x + c)) - \beta\gamma x \sinh(\gamma(x + c)),$$

so

$$\begin{aligned} g(x) &= \alpha \cosh(\gamma(x + c)) - \frac{1}{2}\gamma^2 \cosh(\gamma(x + c)) + \beta\gamma x \sinh(\gamma(x + c)) \\ &= \alpha \cosh(\gamma(x + c)) - \frac{1}{2}\gamma^2 \cosh(\gamma(x + c)) + \beta\gamma x \sinh(\gamma(x + c)) \\ &= \left(\alpha - \frac{1}{2}\gamma^2\right) \cosh(\gamma(x + c)) + \beta\gamma x \sinh(\gamma(x + c)), \end{aligned}$$

which will be nonnegative if and only if α is large enough. The spot rate process is

$$\begin{aligned} r_t = \frac{g(X_t)}{f(X_t)} &= \frac{(\alpha - \frac{1}{2}\gamma^2) \cosh(\gamma(X_t + c)) + \beta\gamma X_t \sinh(\gamma(X_t + c))}{\cosh(\gamma(X_t + c))} \\ &= \beta\gamma X_t \tanh(\gamma(X_t + c)). \end{aligned}$$

We can compute bond prices by

$$P(0, T) = \frac{\cosh \gamma(X_0 e^{-\beta T} + c)}{\cosh \gamma(X_0 + c)} \exp \left[-\alpha T + \frac{\gamma^2}{2} \frac{1 - e^{-2\beta T}}{2\beta} \right].$$

For the next two examples we take the following d-dimensional diffusion as the underlying Markov process :

$$dX_t = 2\sqrt{X_t}dW_t + (a + BX_t)dt, \quad (3.18)$$

where $a_i > 0$ for all i , $B = b_{ij} \geq 0$ for $i \neq j$, and $b_{ii} \leq 0$ for all i .

For this special process the generator is defined by

$$Gf(x) = \sum_{i=1}^d (a + Bx)_i \frac{\partial f}{\partial x_i}(x) + \frac{1}{2} \sum_{i=1}^d 4x_i \frac{\partial^2 f}{\partial x_i^2}(x).$$

3.7.5 Exponential-linear example 2

Here we take (3.18) as the underlying Markov process and some $\gamma \in \mathbb{R}^d$ fixed and consider

$$f(x) = e^{\gamma^T x},$$

which leads to

$$g(x) = f(x) \left[\alpha - 2 \sum_{i=1}^d \gamma_i^2 x_i - \gamma^T (a + Bx) \right].$$

For non-negativity we need

$$(B^T \gamma)_i + 2\gamma_i^2 \leq 0 \quad \forall i.$$

In the case where this inequality is strict for some i , we have that the spot-rate process is an affine function of the diffusion.

For this example bond-prices can seldom be written down in closed form.

3.7.6 Multi-type branching diffusion example

Again we take (3.18) as the underlying Markov process, but now we choose

$$f(x) = \gamma^T x,$$

where $\gamma \in \mathbb{R}^d$, and $\gamma_i > 0$ for all i . This time we get

$$g(x) = -\gamma^T a + x^T (a - B^T) \gamma.$$

We need for γ that $(\alpha - B^T)\gamma \geq 0$ and $-\gamma^T a \geq 0$ for non-negativity of the spot-rate which is given by

$$r_t = \frac{-\gamma^T a + x^T (a - B^T) \gamma}{\gamma^T x}.$$

This can take very large values if $-\gamma^T a > 0$ and small values if $\gamma^T (a - B)_i = 0$ for some i . The case $-\gamma^T a = 0$ results in a bounded function r .

This time the bond prices can be computed more easily and are given by

$$P(0, T) = \frac{e^{-\alpha t} \gamma^T (-B^{-1} a + e^{tB} (X_0 + B^{-1} a))}{\gamma^T X_0}.$$

3.7.7 Rational Log-normal Model

For the final example we use Generic Approach II and take the underlying Markov process to be Brownian motion. For any θ , $\phi(x) = \exp(\theta x)$ is an eigenfunction of the generator $\frac{1}{2}D^2$ with eigenvalue $\frac{1}{2}\theta^2$. Therefore we take

$$g(x) = \sum_{i=1}^n c_i \phi_i = \sum_{i=1}^n c_i \exp(\theta_i x),$$

if this is everywhere positive. For $\alpha > \max(\frac{1}{2}\theta_i^2)$

$$f(x) = \sum_{i=1}^n c_i (\alpha - \lambda_i)^{-1} \phi_i = \sum_{i=1}^n c_i (\alpha - \frac{1}{2}\theta_i^2)^{-1} \exp(\theta_i x).$$

We set $\tilde{c}_i = c_i (\alpha - \frac{1}{2}\theta_i^2)^{-1}$ and get a simple expression for bond prices:

$$P(0, t) = \frac{e^{-\alpha t} \sum_{i=1}^n \tilde{c}_i e^{\lambda_i t} \phi_i(X_0)}{\sum_{i=1}^n \tilde{c}_i \phi_i(X_0)} = \frac{e^{-\alpha t} \sum_{i=1}^n \tilde{c}_i e^{\frac{1}{2}\theta_i^2 t + \theta_i X_0}}{\sum_{i=1}^n \tilde{c}_i e^{\theta_i X_0}}. \quad (3.19)$$

3.8 Foreign Exchange in the Potential approach

Now we consider the pricing of assets in many countries at once and each asset's price is expressed in the currency of its home country. We are often confronted with this problem in practice.

For example: We are asked to price a swap which swaps floating USD interest payments for fixed EUR interest payments. In the conventional approach (This is how pricing and hedging is mostly done today) we would first build a model for the interest rates in the US, then a model for the interest rates in Europe, and then try to model the exchange rate USD/EUR. Even using an extremely simple model, we would need one driving Brownian motion for the USD yield curve, one for the EUR yield curve and one for the exchange rate. So we need three Brownian motions in total.

In our example we have to integrate over three dimensions, and for example the pricing of an American style option is an optimal stopping problem in three dimensions.

The next problem we have to consider is, that no-one would use a one factor model for the yield curves and the exchange rate between them, because they might not be very tractable. But when using multi-factor models we are again beginning to hit problems of dimensionality, so we see the pricing of such an asset gets very complex.

At least it does in the conventional approach. Now we will see how easy it becomes in the potential approach when we consider a complete market⁷.

3.8.1 Foreign Exchange in complete markets

First we introduce some notation. We suppose that

$$1 \text{ unit of currency } j = Y_t^{ij} \text{ units of currency } i.$$

Now if S_t^j is a traded asset in country j , then

$$\zeta_t^j S_t^j \text{ is a martingale.}$$

When we convert its currency- j price into currency i , it becomes a traded asset in country i , and so

⁷ $\exists!$ martingale measure

$\zeta_t^i Y_t^{ij} S_t^j$ is a martingale.

Now we set

$$N_t^{ij} = \frac{\zeta_t^i Y_t^{ij}}{\zeta_t^j}.$$

This is a martingale orthogonal to the space of martingales of the form $\zeta^j S^j$. With the help of N_t^{ij} we can express the exchange rate Y_t^{ij} as

$$Y_t^{ij} = \frac{N_t^{ij} \zeta_t^j}{\zeta_t^i}.$$

When we consider a complete market, N_t^{ij} must be constant. This is the very simple and "nice" result, that *in a complete market the exchange rate between two countries is the ratio of the state-price densities of the two countries*. In general we can say, that there is the possibility of some exchange risk not hedgeable through other assets, represented by the martingale N_t^{ij} .

The biggest advantage of the potential approach based on a Markov process is that *adding another country* does not mean adding more sources of randomness. We just have to build another state-price density over the *same* Markov process. In fact we just have to choose a new f and α . So adding new countries to our model does not cause problems of dimensionality.

3.8.2 Example

The exchange rate between two currencies is given as the ratio of their state-price densities. In Section (3.7) of this chapter we described general methods for creating examples. Here we will discuss the exponential-quadratic example.

Given the process

$$dX_t = dW_t - BX_t dt,$$

and taking

$$f_i(x) = \exp\left(\frac{1}{2}(x - c_i)^T Q_i (x - c_i)\right),$$

where i is the country (currency) index. As before

$$\alpha_i = \frac{1}{2} \text{tr} Q_i + \frac{1}{2} |Q_i c_i|^2 + \frac{1}{2} v_i^T S_i^{-1} v_i,$$

where $S = B^T Q_i + Q_i B - Q_i^2$, $v = (B^T - Q_i) Q_i c_i$, and we get

$$\begin{aligned} g_i(x) &= (\alpha_i - G) f_i(x) \\ &= f(x) \cdot \frac{1}{2} (x - a_i)^T S_i (x - a_i), \end{aligned}$$

where $a_i = S_i^{-1} v_i$. This gives us a spot rate for country i which is again a squared Gaussian process.

To illustrate our results, we will now have a detailed look at two cases:

$$(a) Q_i = Q \quad \forall i \quad (b) c_i = 0 \quad \forall i.$$

Case (a)

$Q_i = Q$ for all i . By applying a rotation to X we can suppose that Q is diagonal. We recall that

$$\zeta_t = \frac{e^{-\alpha t} R_\alpha g(X_t)}{R_\alpha g(X_0)} = \frac{e^{-\alpha t} f(X_t)}{f(X_0)}.$$

So taking f from above the state-price density in country i is

$$\zeta_t^i = \frac{\exp(-\alpha_i t + \frac{1}{2}(X_t - c_i)^T Q (X_t - c_i))}{\exp(\frac{1}{2}(X_0 - c_i)^T Q (X_0 - c_i))} \quad (3.20)$$

$$= \exp(-\alpha_i t + \frac{1}{2}(X_t - c_i)^T Q (X_t - c_i) - \frac{1}{2}(X_0 - c_i)^T Q (X_0 - c_i)). \quad (3.21)$$

Picking the same Q for all i and computing the exchange rate leads us to

$$\frac{Y_t^{ij}}{Y_0^{ij}} = \frac{\zeta_t^j}{\zeta_t^i}. \quad (3.22)$$

Now we insert the ζ_t from above for each country. Because $Q = (q_{ij})_{i,j=1,\dots,d}$ is diagonal we only have the quadratic terms left after multiplying. We can write the last fraction as

$$\exp[t(\alpha_i - \alpha_j) + \sum_{k=1}^d \frac{1}{2} q_{kk} ((X_{t,k} - c_{j,k})^2 - (X_{0,k} - c_{j,k})^2 - (X_{t,k} - c_{i,k})^2 + (X_{0,k} - c_{i,k})^2)].$$

Simplifying this term (the quadratic terms all vanish) leads to

$$\begin{aligned} \exp[t(\alpha_i - \alpha_j) + \sum_{k=1}^d q_{kk} (-X_{t,k} c_{j,k} + X_{0,k} c_{j,k} + X_{t,k} c_{i,k} - X_{0,k} c_{i,k})] = \\ \exp[t(\alpha_i - \alpha_j) + \sum_{k=1}^d q_{kk} ((c_{i,k} - c_{j,k})(X_{t,k} - X_{0,k}))], \end{aligned}$$

where $X_{t,k}, c_{i,k}$ denotes the k -th component of the vector X_t , respectively c_i .

The last result written in matrix form leads us to

$$\frac{Y_t^{ij}}{Y_0^{ij}} = \exp[t(\alpha_i - \alpha_j) + (c_i - c_j)^T Q (X_t - X_0)]. \quad (3.23)$$

In this example we have a structure where the exchange rates between two countries are log-brownian motions, and all spot rates are squared-gaussian. In section 7 we computed bond prices for this example, so the yield curve in country i is

$$\begin{aligned} y_t^i &= -\frac{1}{t} \log P(0, t) \\ &= \alpha_i + \frac{1}{2t} \log \det(I - QV_t) + \frac{1}{2t} [(x - c_i)^T Q (x - c_i) - \xi_t^T (I - QV_t)^{-1} Q \xi_t], \end{aligned}$$

where $\xi_t = e^{-tB} x - c_i$, and we write x instead of X_0 .

Case (b)

$c_i = 0$ for all i . This simplification centers all the quadratic forms which appear to 0. So we get

$$f_i(x) = \exp[\frac{1}{2} x^T Q_i x].$$

S_i does not change, $v_i = 0$ and clearly $a_i = S_i v_i = 0$, so $\alpha_i = \frac{1}{2} \text{tr}(Q_i)$ and

$$g_i(x) = f_i(x) x^T S_i^{-1} x.$$

The state-price density in country i is

$$\begin{aligned} \zeta_t^i &= \frac{e^{-\alpha t} f(X_t)}{f(X_0)} \\ &= \exp[-\alpha_i t + \frac{1}{2}(X_t^T Q_i X_t - X_0^T Q_i X_0)]. \end{aligned} \quad (3.24)$$

For the exchange rate we get

$$\begin{aligned} \frac{\zeta_t^j}{\zeta_t^i} &= \frac{\exp[-\alpha_j t + \frac{1}{2}(X_t^T Q_j X_t - X_0^T Q_j X_0)]}{\exp[-\alpha_i t + \frac{1}{2}(X_t^T Q_i X_t - X_0^T Q_i X_0)]} \\ &= \exp[t(\alpha_i - \alpha_j) + \frac{1}{2}(X_t^T (Q_j - Q_i) X_t + X_0^T (Q_i - Q_j) X_0)]. \end{aligned}$$

The resulting yield curve in country i at time t is

$$y_t^i = \alpha_i + \frac{1}{2t} \log(\det(I - Q_i V_t)) + \frac{1}{2t} [x^T Q_i x - (e^{-tB} x)^T (I - Q_i V_t)^{-1} Q_i e^{-tB} x].$$

4 Simulation

Now we arrived at a point where we want to get results and see if the theory we discussed before really works. At first we will "choose" some parameters and "play" around to get some basic results. Then we will try a fit on some real data and see if we are able to come up with results close to reality. I used the open source program scilab-4.1.2 which you can get at "www.scilab.org". Scilab is a free Matlab "clone". The code of the programmed functions can be viewed in the Appendix.

4.1 Simulation of a finite Markov chain

At first we are trying to simulate the process our framework is all about: a Markov process. We simulate a finite Markov chain in continuous time with state space $\mathcal{N} = \{1, 2, \dots, N\}$. All that we have to do to simulate a continuous-time Markov chain is to choose a generator Q and an initial distribution α^1 , the distribution of X_0 .

This can be determined by setting $X_0 = i$. Therefore the initial distribution $\alpha = (0, \dots, 1, \dots, 0)$ with 1 being the i -th component of the vector α . Otherwise, if the distribution of α is given then X_0 is a random variable generated after the law of α . For my simulations i used the discrete uniform distribution. Thus $\alpha = (\frac{1}{N}, \dots, \frac{1}{N})$ of size N . To generate an X_0 of this distribution α we can use the following algorithm:

Algorithm 4.1.1 (Generating X_0) (1a) $X_0 = i$ is determined. Therefore

$\alpha = (0, \dots, 1, \dots, 0)$ as described above and we are finished.

(1b) Otherwise: $\alpha = (\frac{1}{N}, \dots, \frac{1}{N})$, and we we set $\beta = (\alpha_1, \alpha_1 + \alpha_2, \dots, 1) = (\frac{1}{N}, \frac{2}{N}, \dots, 1)$.

(2) We use the implemented generator of Scilab to generate an $U(0, 1)$ random variable z .

(3) We make a loop from 1 to N to check if $z \leq \beta_i$. We set $\beta_0 = 0$ and if $z \leq \beta_i$ then z is in the interval $(\beta_{i-1}, \beta_i]$ and we set $X_0 = i$.

(4) Return: the value of X_0 .

Hence: $\mathbb{P}(\beta_{i-1} < z \leq \beta_i) = \frac{1}{N}$.

Clearly any other distribution can be taken for α . For generating an X_i for $i > 0$ we use that the transition densities at time t are given by $P_t = (p_{ij}(t))_{i,j \in \mathcal{N}} = \exp(tQ)$ and clearly we use the Markov property (1.13). The simulation of a continuous-time Markov chain with finite state space can be summed in following algorithm:

Algorithm 4.1.2 (Generation of a continuous-time Markov chain) (1) We choose a generating Matrix Q that fulfills the properties of Theorem (1.2.5).

(2) If $\text{state_0} \in \{1, 2, \dots, N\}$ then $X_0 = \text{state_0}$ else if $\text{state_0} = 0$, then X_0 is generated using Algorithm 4.1.1.

(3) We set $n > 0$ to be the maximum time of our simulation and choose a discretisation $d \geq 1^2$. For the beginning we set the time $t = 0$.

(4) We set $t = t + 1/d$ and compute $P_t = \exp(tQ)$. The conditional distribution of X_t given $X_{t-1/d} = i$ is given by the i -th row of P_t which we set γ .

(5) We set $\beta = (\gamma_1, \gamma_1 + \gamma_2, \dots, 1)$.

¹I go along with the notation of the initial distribution in [Kijima]. Do not change this α with the α used in the potential framework.

²This was chosen for more comfort when working with the functions. In the code we set $d = 1/d$.

- (6) We generate an $U(0,1)$ random variable z .
 (7) We make a loop from 1 to N to check if $z \leq \beta_j$. We set $\beta_0 = 0$ and if $z \leq \beta_j$ then z is in the interval $(\beta_{j-1}, \beta_j]$ and we set $X_t = j$.
 Hence: $\mathbb{P}(\beta_{j-1} < z \leq \beta_j) = p_{ij}(t)$.
 (8) We go back to point (4) until $t = n$ is reached.
 (9) Return: a vector of size $nd + 1$ with the states of $(X_t)_{t \geq 0}$ as entries.

These two algorithms are implemented in the Scilab function "markov.sci". The function markov has the parameters Q (the generator), state_0 (the state of X_0 , can be chosen from 1 to N , if chosen 0, then X_0 is generated as in 4.1.1 point (1b)-(2)), n (the maximum time of the simulation) and d (the step size of discretisation).

For the first simulations i choose a matrix

$$Q = \begin{pmatrix} -3 & 1 & 0 & 1 & 1 \\ 0 & -2 & 0.5 & 0.5 & 1 \\ 1 & 2 & -4 & 0 & 1 \\ 0.25 & 0 & 0.25 & -1 & 0.5 \\ 0 & 1 & 1 & 3 & -5 \end{pmatrix},$$

which fulfills the properties of Theorem (1.2.5). Clearly our Markov chain has state space $\mathcal{N} = \{1, 2, 3, 4, 5\}$. The eigenvalues λ_i of Q rounded to 3 steps after comma are

$$\lambda_i = 0, -5.656, -3.435, -4.23, -1.679$$

and the matrix of the eigenvectors ϕ_i to eigenvalue λ_i (again rounded) is

$$V = \begin{pmatrix} -0.447 & -0.236 & 0.724 & -0.11 & 0.231 \\ -0.447 & -0.223 & -0.187 & 0.303 & 0.618 \\ -0.447 & -0.149 & 0.648 & -0.893 & 0.655 \\ -0.447 & -0.079 & -0.144 & 0.122 & -0.365 \\ -0.447 & 0.93 & 0.017 & -0.29 & 0.053 \end{pmatrix}.$$

We recall Section (1.2.6) and especially that Q can be diagonalized by

$$Q = VDV^{-1},$$

and we can compute P_t by

$$P_t = \exp(tQ) = VD(t)V^{-1}.$$

where D is a matrix with the eigenvalues of Q as the diagonal entries and $D(t)$ is the matrix with diagonal entries $e^{\lambda_i t}$. The eigenvalues of Q are quite big negative and so the convergence of P_t against

$$\lim_{t \rightarrow \infty} P_t = \begin{pmatrix} 0.073 & 0.185 & 0.087 & 0.533 & 0.122 \\ 0.073 & 0.185 & 0.087 & 0.533 & 0.122 \\ 0.073 & 0.185 & 0.087 & 0.533 & 0.122 \\ 0.073 & 0.185 & 0.087 & 0.533 & 0.122 \\ 0.073 & 0.185 & 0.087 & 0.533 & 0.122 \end{pmatrix}$$

is rather fast. We see that the conditional distribution of X_t for great t given $X_{t-d} = i$ is independent from i . This means that there is no more dynamic in the Markov chain and we will see later how this especially influences the behavior of the yield curves.

For illustration we look at two trajectories of the Markov chain produced by the Scilab-function "markov.sci" both starting with $X_0 = 1$, with parameters $\text{markov}(Q, 1, 30, 2)$: Figure (4.1) and (4.2).

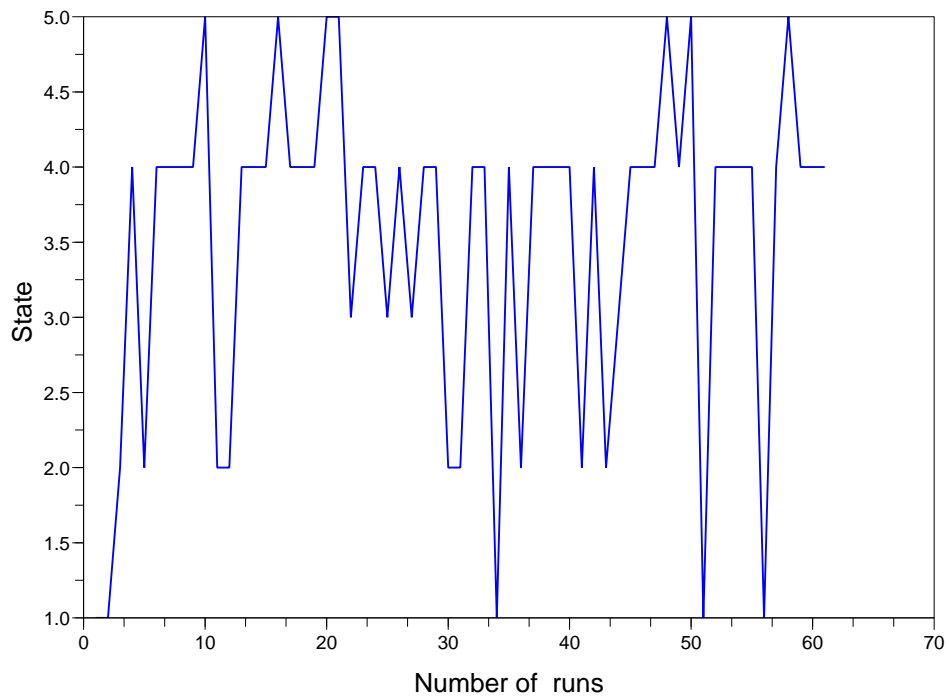


Figure 4.1: Trajectory of a finite Markov chain

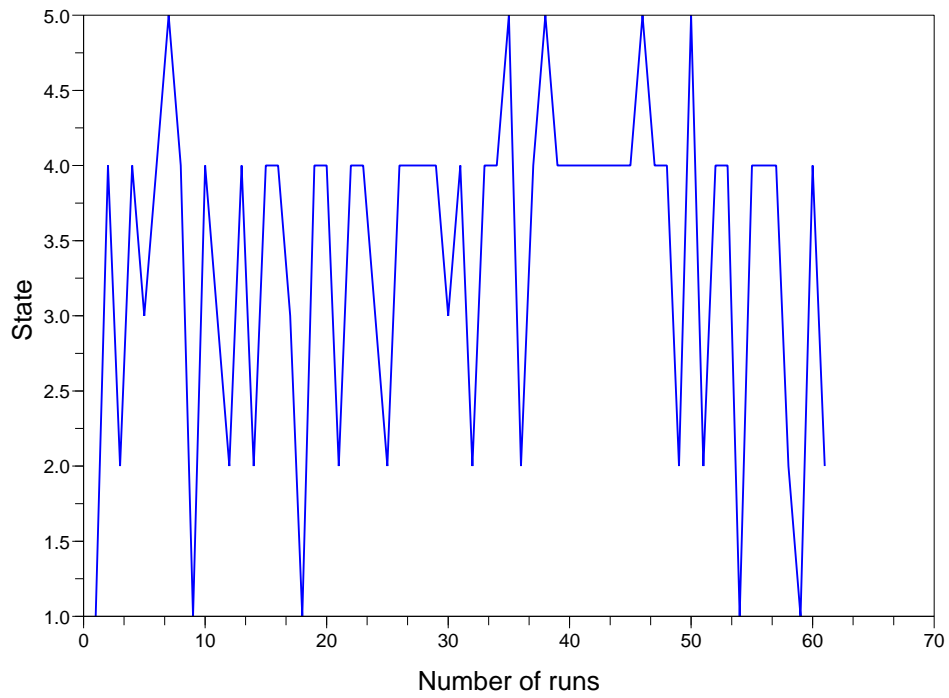


Figure 4.2: Trajectory of a finite Markov chain

4.2 Calculation of bond-prices and yields

In this section we will calculate prices of zero-coupon bonds $P(0, t)$ and the yield to maturity $Y(0, t)$ by the two approaches shown in Section (3.4).

4.2.1 Generic approach 1

From Section (3.4.1) we know that

$$P(0, t) = \frac{e^{-\alpha t} \mathbb{E}_0[f(X_t)]}{f(X_0)}.$$

In the case of a finite state space Markov chain the conditional expectation is a finite weighted sum over the state space of the Markov chain, so that

$$P(0, t) = \frac{e^{-\alpha t} P_t f_{X_0}}{f_{X_0}}, \quad (4.1)$$

where clearly $P_t = \exp(tQ)$. For a Markov chain with finite state space, f is a vector of size N and the term $P_t f_{X_0}$ means that we take the i -th component of the vector $P_t f$ when $X_0 = i, i \in \mathcal{N}$. In this case the price is a weighted sum of exponential functions.

After having chosen a generator Q we just have to choose an $\alpha > 0$ and a positive function $f = (\alpha - Q)^{-1}g$. We use following theorem:

Theorem 4.2.1 *If $\alpha > 0$ and Q is of the form (1.2.5) then $(\alpha - Q)^{-1} \geq \mathbf{0}$.*

Proof 4.2.1 *When we calculate $(\alpha - Q)^{-1}$ we can do this by transforming*

$$\begin{pmatrix} \alpha + q_{11} & \cdots & -q_{1n} \\ \vdots & \vdots & \vdots \\ -q_{n1} & \cdots & \alpha + q_{nn} \\ \hline 1 & \cdots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \cdots & 1 \end{pmatrix}, \text{ to } \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \cdots & 1 \\ \hline & & (\alpha - Q)^{-1} \end{pmatrix}.$$

The upper matrix is transformed to identity and we do the same transformation for the lower matrix which is identity at the beginning. The non-diagonal elements $(-q_{ij})_{i \neq j}$ in row i can be transformed to 0 by adding positive multiplies of $(\alpha + q_{ii})$. Therefore we just have adding of positive multiplies and therefore $(\alpha - Q)^{-1} \geq \mathbf{0}$. \square

So when we pick a positive vector g we can be sure that $f = (\alpha - Q)^{-1}g \geq \mathbf{0}$.

For the first simulations i picked a vector

$$g = [2, 3, 1, 1, 4]$$

and $\alpha = 0.05$. Therefore

$$f = [36.483, 36.938, 36.528, 35.756, 36.581].$$

The calculation of bond-prices is implemented in the Scilab-function "bond.sci". This function has the parameters Q for the generator, **state_0** (the state of X_0), α , **n** for the maximum maturity of simulation and **d** as the step size (again $d \geq 1$ and set to $d = 1/d$ in code). We use following algorithm to compute the prices for the different maturities of a zero-bond.

Algorithm 4.2.1 (Bond-prices) (1) We choose the parameters $Q, g, \text{state_0}, \alpha, n, d$ as described above.
(2) If $\text{state_0} \in \{1, 2, \dots, N\}$ then $X_0 = \text{state_0}$ else if $\text{state_0} = 0$, then X_0 is generated using Algorithm 4.1.1.
(3) We set $P(0, 0) = 1$, $f = (\alpha - Q)^{-1}g$ and $t = 0$.
(4) We set $t = t + 1/d$.
(5) We calculate $P_t = \exp(tQ)$.
(6) We compute $P(0, t)$ using (4.1).
(7) We go back to (4) until $t = n$ is reached.
(8) Return: a vector of size $nd + 1$ with the values of $P(0, t)$ for $0 \leq t \leq n$ as entries.

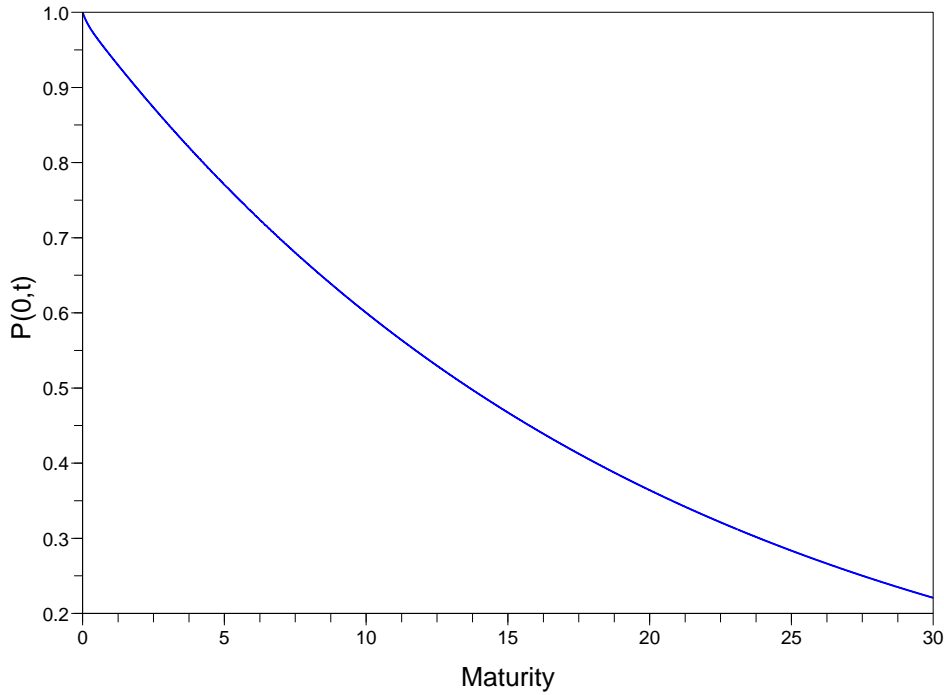


Figure 4.3: $P(0, t)$ for bond($Q, g, 1, 0.05, 30, 360$)

We are much more interested in the different yield curves that occur. When we look at Equation (4.1) we see that in our case we can have 5 different values for $P(0, t)$ for fixed $t \in [0, n]$ depending on X_0 . From (2.4) we know that the yield to maturity is

$$Y(t, T) = -\frac{\ln(P(t, T))}{T - t}. \quad (4.2)$$

So if using a 5-state chain, only 5 different yield curves can occur. When we look at (4.2) we see that in this case the yield to maturity is $-1/t$ times the \ln of a weighted sum of exponential functions. The calculation of yields is implemented in the short Scilab-function "yield.sci" and has the same parameters as "bond.sci". We use (4.1) and (4.2) to compute the yield by following algorithm:

Algorithm 4.2.2 (Yields) (1) We choose the parameters $Q, g, \text{state_0}, \alpha, n, d$ as described above.
(2) If $\text{state_0} \in \{1, 2, \dots, N\}$ then $X_0 = \text{state_0}$ else if $\text{state_0} = 0$, then X_0 is generated

using Algorithm 4.1.1.

- (3) We set $f = (\alpha - Q)^{-1}g$, $Y(0,0) = g_{X_0}/f_{X_0}$ and $t = 0$.
- (4) We calculate a vector of size $nd + 1$ of bond-prices $P(0,t)$ using Algorithm (4.2.1).
- (4) We set $t=t+1/d$.
- (5) We use $P(0,t)$ to compute $Y(0,t)$ with (4.2).
- (6) We go back to (4) until $t = n$ is reached.
- (7) Return: a vector of size $nd + 1$ with the values of $Y(0,t)$ for $0 \leq t \leq n$ as entries.

Remark 4.2.1 Point (3): $Y(0,0) = g_{X_0}/f_{X_0}$ is the spot-rate at time 0.

Proof 4.2.2

$$Y(0,0) = \lim_{t \rightarrow 0} Y(0,t) = \lim_{t \rightarrow 0} -\frac{\ln P(0,t)}{t} = \text{„} \frac{0}{0} \text{„}$$

and we use de L'Hospital's Rule to get

$$= \lim_{t \rightarrow 0} -\frac{\frac{1}{P(0,t)}P(0,t)'}{1} = \lim_{t \rightarrow 0} -P(0,t)'$$

as $P(0,0)=1$. So we need to find what $P(0,t)'$ is. We start from (4.1) and use $P_t = \exp(tQ)$ and receive

$$P(0,t)' = \frac{1}{f_{X_0}}[-\alpha \exp(tQ)f_{X_0} + e^{-\alpha t}Q \exp(tQ)f_{X_0}]$$

When we now compute $-\lim_{t \rightarrow 0}$ of this term where $\lim_{t \rightarrow 0} \exp(tQ) = I$ we get

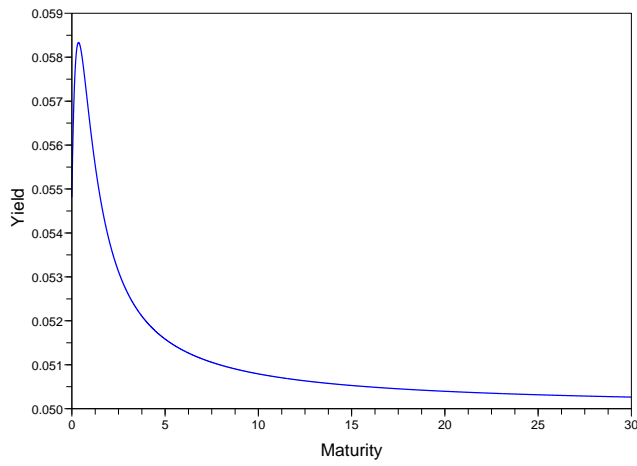
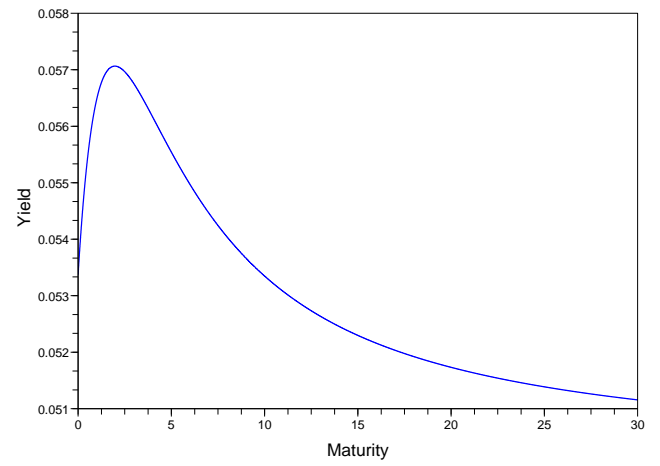
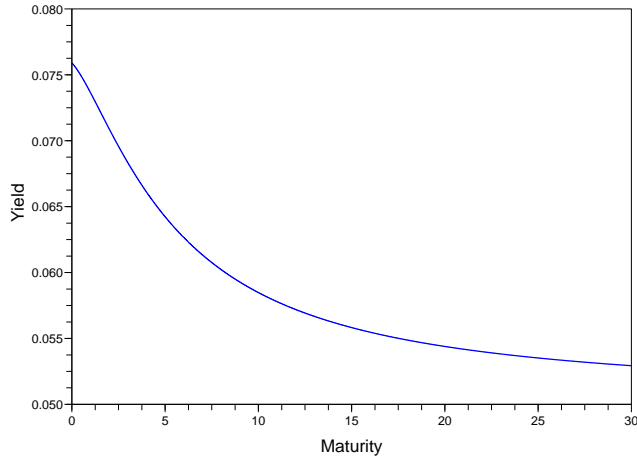
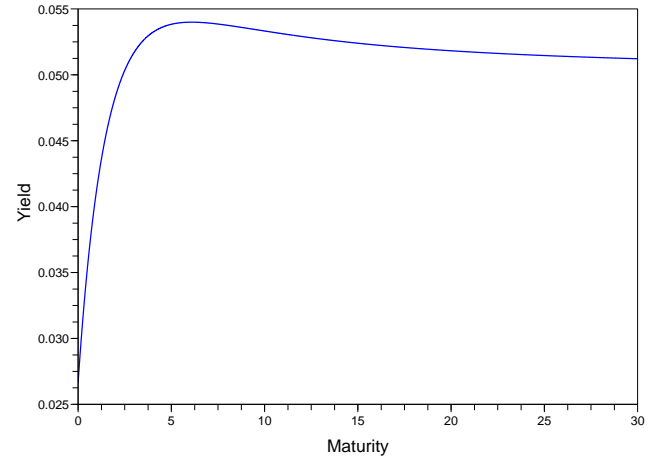
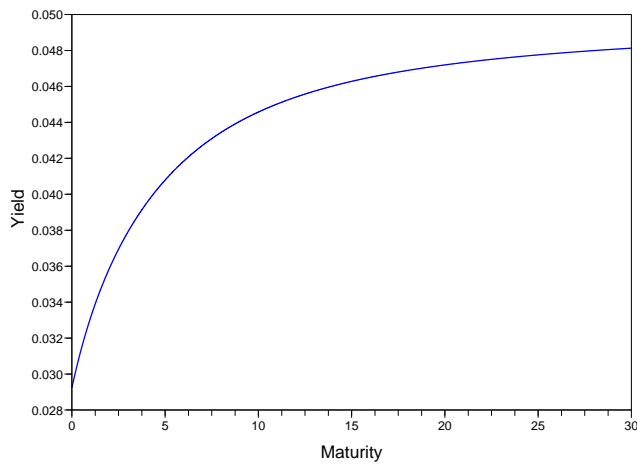
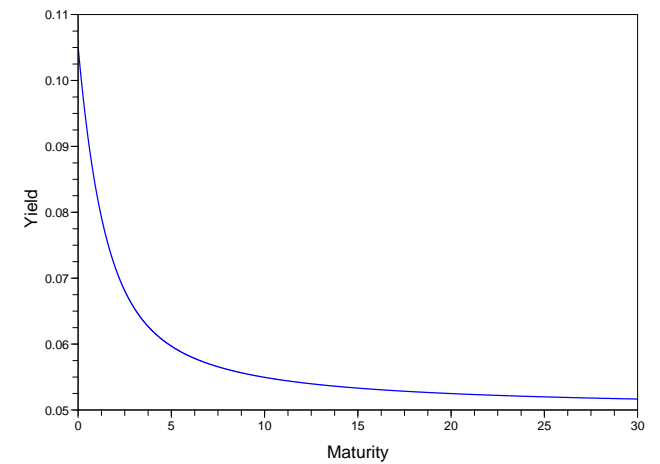
$$-\frac{1}{f_{X_0}}[-\alpha I f_{X_0} + Q I f_{X_0}] = \frac{(\alpha I - Q)f_{X_0}}{f_{X_0}} = \frac{g_{X_0}}{f_{X_0}}. \quad \square$$

Hence: In General we clearly have $Y(t,t) = r_t$.

In Figure (4.4) we see a plot for yield($Q, g, 1, 0.05, 30, 360$). The structure is sharply humped because of the fact that the eigenvalues of the generating matrix Q are quite big negative and therefore the convergence of $\lim_{t \rightarrow \infty} P_t = \lim_{t \rightarrow \infty} \exp(tQ)$ is fast. Lets see what happens if we take $G = Q/5$ which causes that the eigenvalues μ_i of G are $1/5\lambda_i$ the eigenvalues of Q . In Figure (4.5) we can see that the structure is also humped but not as sharply as before.

The other yield curves that are possible for $X_0 = 2, 3, 4, 5$ can be seen from Figure (4.6) to (4.9). Again we use $G = Q/5$.

Remark 4.2.2 As we see a matrix with eigenvalues close to 0 produces flat yield curves. The parameter α shifts the curve up and down.

Figure 4.4: Yield curve for $\text{yield}(Q, g, 1, 0.05, 30, 360)$ Figure 4.5: Yield curve for $\text{yield}(Q/5, g, 1, 0.05, 30, 360)$ Figure 4.6: Yield curve for $\text{yield}(Q/5, g, 2, 0.05, 30, 360)$ Figure 4.7: Yield curve for $\text{yield}(Q/5, g, 3, 0.05, 30, 360)$ Figure 4.8: Yield curve for $\text{yield}(Q/5, g, 4, 0.05, 30, 360)$ Figure 4.9: Yield curve for $\text{yield}(Q/5, g, 5, 0.05, 30, 360)$

4.2.2 Generic approach 2

Now we look at the results when we use Generic approach 2. We recall Section (3.4.2). In this case the price of a zero-bond is written as

$$P(0, t) = \frac{e^{-\alpha t} \sum_{i=1}^n \tilde{c}_i \exp(\lambda_i t) \phi_i(X_0)}{\sum_{i=1}^n \tilde{c}_i \phi_i(X_0)}, \quad (4.3)$$

in matrix notation we can write this as

$$P(0, t) = \frac{e^{-\alpha t} V \exp(tD) \tilde{c}_{X_0}}{V \tilde{c}_{X_0}}, \quad (4.4)$$

where $g = Vc$ is nonnegative, $\alpha > \max(\lambda_i)$ and $f = V\tilde{c}$ with $\tilde{c} = c(\alpha - D)^{-1}$. D is a diagonal matrix with the eigenvalues λ_i as diagonal entries, V is the matrix of the eigenvectors ϕ_i of the generating matrix Q .

In difference to Generic Approach 1 we only need to find the eigenvectors, but do not need to compute the transition densities.

For a first simulation we choose the generating Matrix as in Generic Approach 1, and

$$g = [1, 3, 5, 2, 1].$$

Then we recover c by solving $Vc = g$. Therefore we receive

$$c = [-5.031, -2.708, -3.679, -4.37, 1.261].$$

Choosing $\alpha = 0.05$ again (all $\lambda_i \leq 0$ so $0.05 > \max(\lambda_i)$) leads to

$$f = [44.629, 45.445, 45.776, 44.799, 44.875].$$

Now we can use (4.4) to compute $P(0, t)$.

For simplicity i computed the bond-prices and the yields to maturity in one Scilab-function implemented in "bond_yield_gen2.sci". Following algorithm is implemented:

Algorithm 4.2.3 (Bond-prices and Yields for Generic Approach 2) (1) We choose the parameters $Q, c, \text{state_0}, \alpha, n, d$ as described above.

(2) If $\text{state_0} \in \{1, 2, \dots, N\}$ then $X_0 = \text{state_0}$ else if $\text{state_0} = 0$, then X_0 is generated by Algorithm 4.1.1.

(3) We compute the matrices V and D using the Scilab-function $[V, D] = \text{spec}(Q)$.

(4) We set $g = Vc$, $f = V(\alpha - D)^{-1}c$, $\tilde{c} = (\alpha - D)^{-1}c$ and $t = 0$.

(5) We set $P(0, 0) = 1$ and $Y(0, 0) = g_{X_0}/f_{X_0}$.

(6) We set $t = t + 1/d$.

(7) We calculate $P(0, t)$ by (4.4) and $Y(0, t)$ by (4.2).

(8) We go back to (6) until $t = n$ is reached.

(9) Return: 2 vectors of size $nd + 1$ with the values of $P(0, t)$ and $Y(0, t)$ for $0 \leq t \leq n$ as entries.

In Scilab you can call this function in the workspace by

$[b, y] = \text{bond_yield_gen2}(Q, c, \text{state_0}, \alpha, n, d)$. Then the vector of bond prices is stored in b , the vector of yields is stored in y .

Again we are more interested in the yield curves than in the bond-prices. As in Generic Approach 1 we get the same result when we choose a generating matrix G with eigenvalues close to 0. The curve becomes flatter. We compare the plots for $\text{bond_yield_gen2}(Q, c, 1, 0.05, 30, 360)$ to choosing $G = Q/5$ in Figure (4.10) and (4.11).

The other yield curves that occur for $X_0 = 2, 3, 4, 5$ can be seen from Figure (4.12) to (4.15). Again with $G = Q/5$, the other parameters stay the same. Now we will discuss what would have happened if we used the same vector g as in Generic approach 1.

Remark 4.2.3 *If we choose c as the solution of $Vc = g$ for the vector g used in Generic approach 1, then $P(0, t)$ of the two approaches is the same for all $t \in [0, n]$.*

Proof 4.2.3 *We are going to show that (4.4) is equal to (4.1). The denominator of (4.4) is f_{X_0} per definition, so all that is left to show is that*

$$V \exp(tD) \tilde{c}_{X_0} = P_t f_{X_0}.$$

We recall that we can write $P_t = \exp(tQ) = V \lambda_D(t) V^{-1}$ where $\lambda_D(t) = \exp(tD)$ as described in Section (1.2.6). From Generic Approach 2 we write $f = V \tilde{c}$. Then we get

$$V \lambda_D(t) \tilde{c}_{X_0} = V \lambda_D(t) V^{-1} V \tilde{c}_{X_0},$$

and clearly

$$V \lambda_D(t) \tilde{c}_{X_0} = V \lambda_D(t) \tilde{c}_{X_0}. \quad \square$$

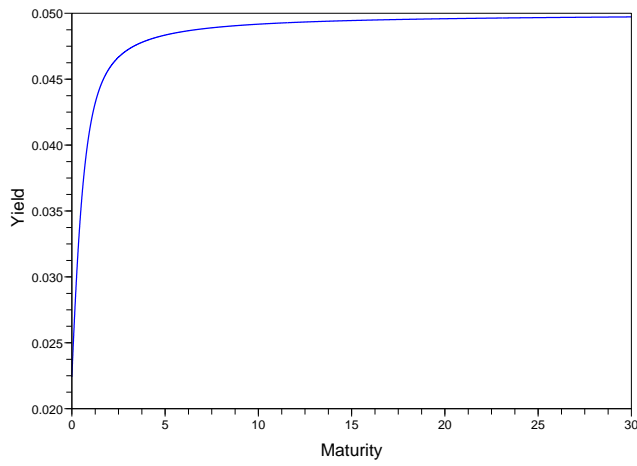
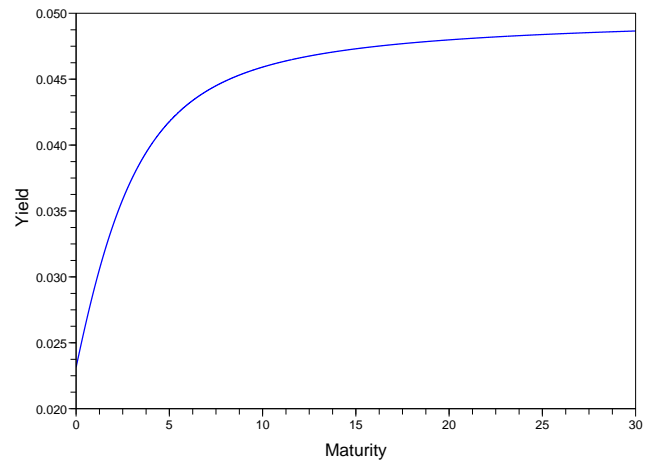
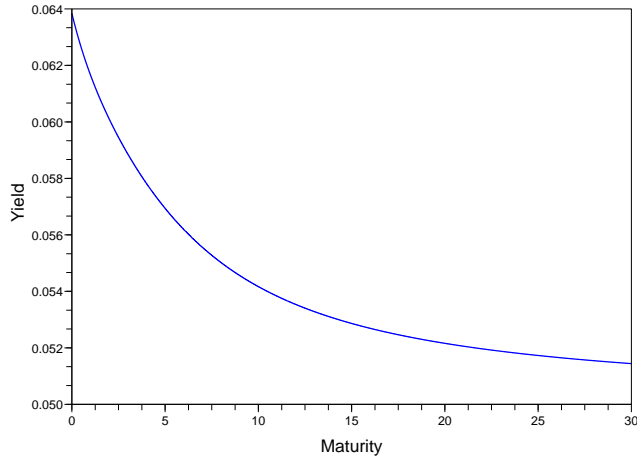
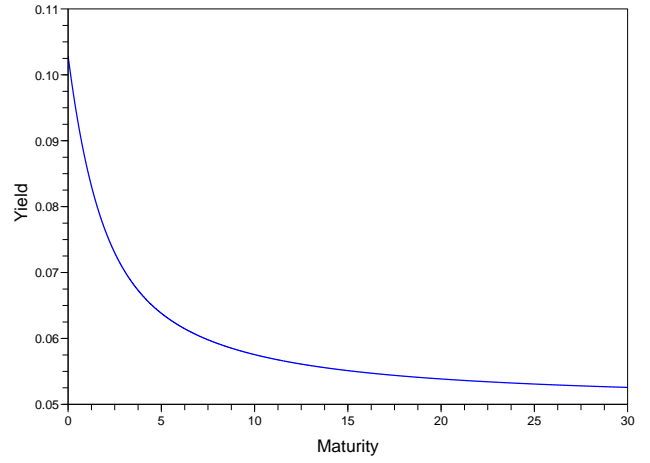
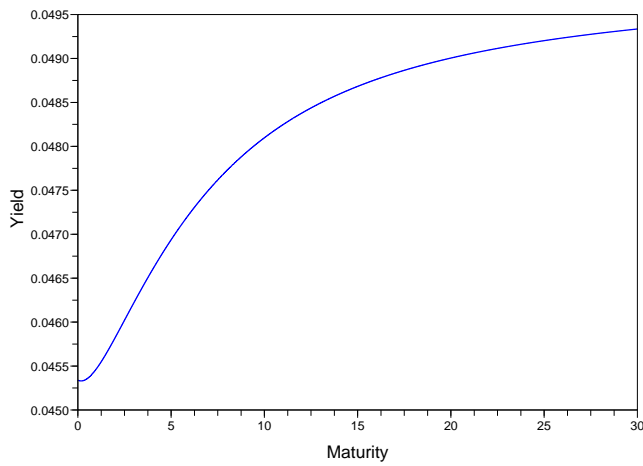
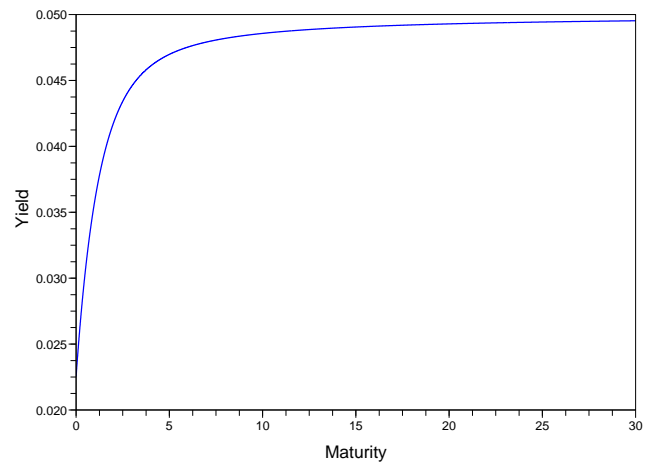
When we look at the resulting yield curves we see that we have some interesting shapes in the Figures (4.5), (4.6) and (4.14). We now try to work the occurring humps more sharply by just manipulating the vector g . The other parameters $(Q/5, \alpha, n, d)$ stay the same. As shown in (4.2.3) the two approaches are equal using the same g . For Generic Approach 1 we had

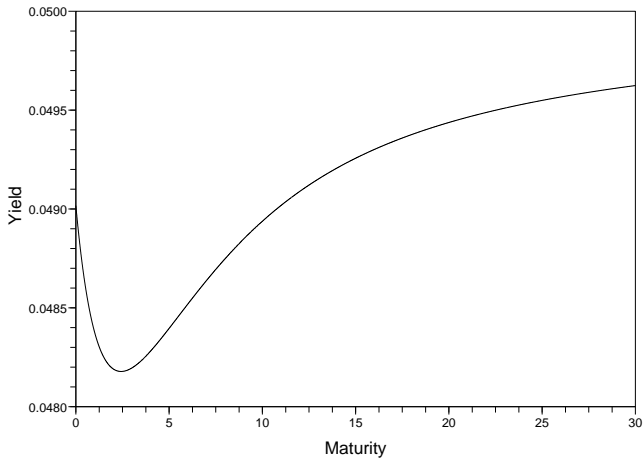
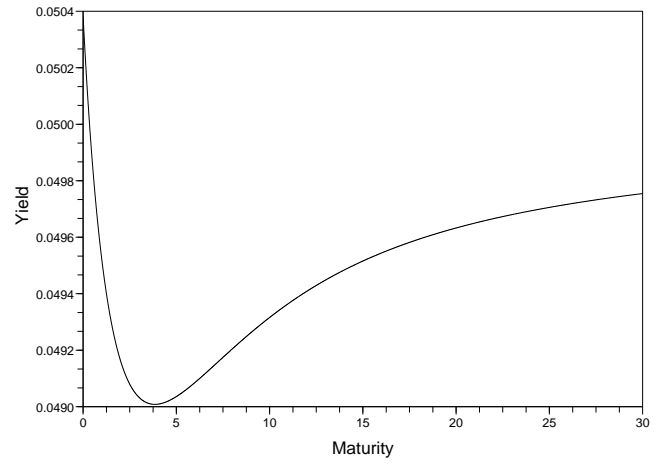
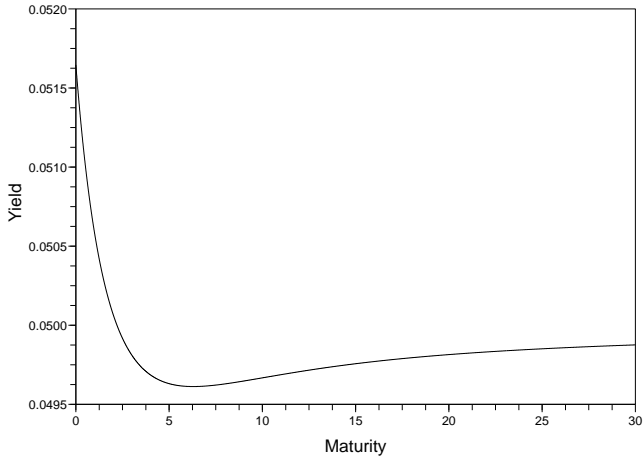
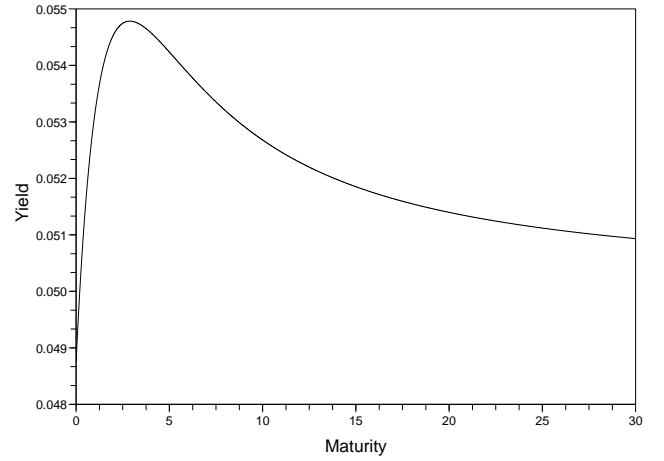
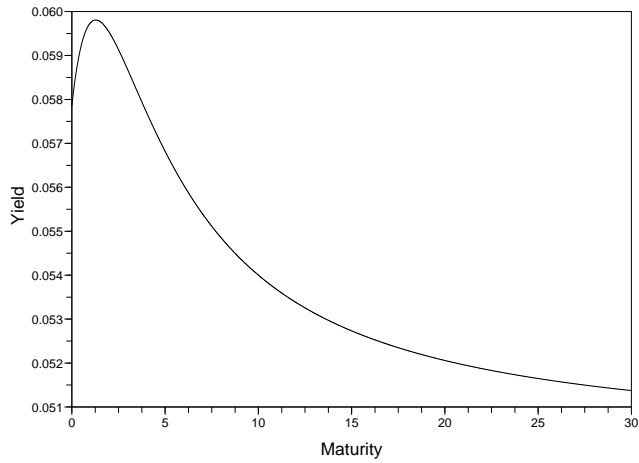
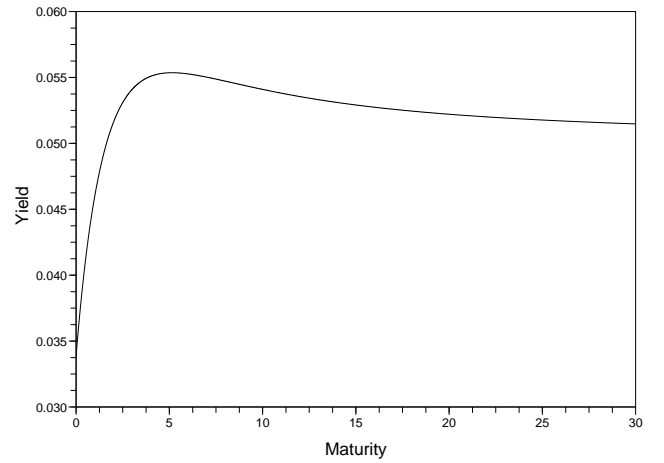
$$g = [2, 3, 1, 1, 4],$$

for Generic Approach 2 we had

$$g = [1, 3, 5, 2, 1].$$

If $X_0 = i, i \in \{1, \dots, N\}$ then we try to get a more irregular, humped structure by using a vector $k = g$ and changing the i -th component of this vector. The results can be seen in Figure (4.16) to (4.21). Using special parameters we can get some really "nice" irregular shapes.

Figure 4.10: $\text{bond_yield_gen2}(Q, c, 1, 0.05, 30, 360)$ Figure 4.11: $\text{bond_yield_gen2}(Q/5, c, 1, 0.05, 30, 360)$ Figure 4.12: $\text{bond_yield_gen2}(Q/5, c, 2, 0.05, 30, 360)$ Figure 4.13: $\text{bond_yield_gen2}(Q/5, c, 3, 0.05, 30, 360)$ Figure 4.14: $\text{bond_yield_gen2}(Q/5, c, 4, 0.05, 30, 360)$ Figure 4.15: $\text{bond_yield_gen2}(Q/5, c, 5, 0.05, 30, 360)$

Figure 4.16: $k = [1, 3, 5, 2.2, 1]$, $X_0 = 4$ Figure 4.17: $k = [1, 3, 5, 2.35, 1]$, $X_0 = 4$ Figure 4.18: $k = [1, 3, 5, 2.5, 1]$, $X_0 = 4$ Figure 4.19: $k = [1.8, 3, 1, 1, 4]$, $X_0 = 1$ Figure 4.20: $k = [2.2, 3, 1, 1, 4]$, $X_0 = 1$ Figure 4.21: $k = [2, 3, 1.3, 1, 4]$, $X_0 = 3$

4.3 Simulation of spot-rates and state-price densities

Now we simulate the positive supermartingale ζ known as state price density and the spot-rate process r_t . We recall their definitions in Section (3.4):

$$\zeta_t = \frac{e^{-\alpha t} R_\alpha g(X_t)}{R_\alpha g(X_0)} = \frac{e^{-\alpha t} f(X_t)}{f(X_0)}, \quad (4.5)$$

when we take the standardisation of $\zeta_0 = 1$, and

$$r_t = \frac{g(X_t)}{R_\alpha g(X_t)} = \frac{g(X_t)}{f(X_t)}, \quad (4.6)$$

where $f = (\alpha - Q)^{-1}g$.

So all we have to do to simulate the two processes is to pick a generating matrix Q , a positive vector g and $\alpha > 0$. The calculation was implemented in the functions "state-price.sci" and "spot-rate.sci". Both have the same parameters namely $Q, g, \text{state_0}, \alpha, n, d$ which have the same characteristics as in "bond.sci" or "yield.sci". We use following algorithms:

Algorithm 4.3.1 (State-price density) (1) We choose the parameters

$Q, c, \text{state_0}, \alpha, n, d$ as described above.

(2) We use Algorithm (4.1.2) with parameters $Q, \text{state_0}, n, d$ to simulate a Markov chain of size $nd + 1$.

(3) We set $t = 0$ and $f = (\alpha - Q)^{-1}g$.

(4) We calculate ζ_t by (4.5).

(5) We set $t = t + 1/d$.

(6) We go back to (4) until $t = n$ is reached.

(7) Return: a vector of size $nd + 1$ with the values of ζ_t for $0 \leq t \leq n$ as entries.

Algorithm 4.3.2 (Spot-rate) (1) We choose the parameters

$Q, c, \text{state_0}, \alpha, n, d$ as described above.

(2) We use Algorithm (4.1.2) with parameters $Q, \text{state_0}, n, d$ to simulate a Markov chain of size $nd + 1$.

(3) We set $t = 0$ and $f = (\alpha - Q)^{-1}g$.

(4) We calculate r_t by (4.6).

(5) We set $t = t + 1/d$.

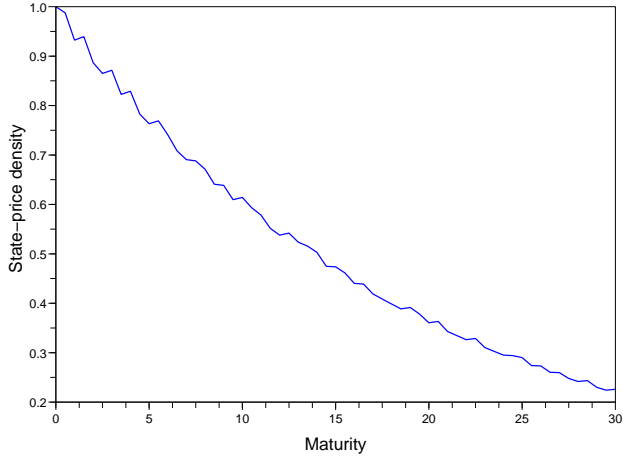
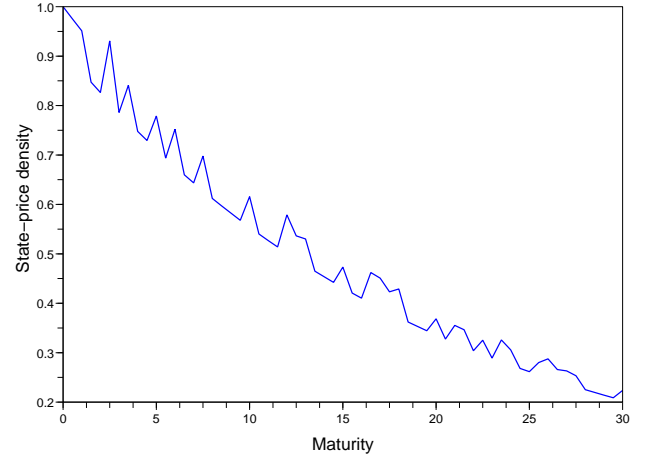
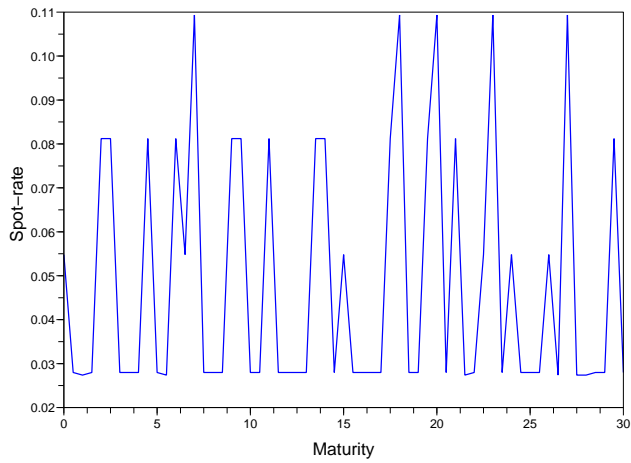
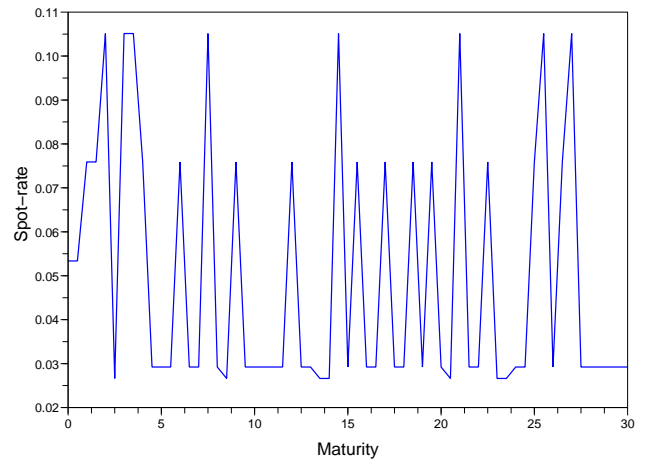
(6) We go back to (4) until $t = n$ is reached.

(7) Return: a vector of size $nd + 1$ with the values of r_t for $0 \leq t \leq n$ as entries.

Now we will look at some plots. We choose Q and g to be the same as for Generic approach 1. First we see two trajectories of the state-price density process with `state_price(Q,g,1,0.05,30,2)`. The step size is only taken 2 for better visibility of the results. In Figure (4.22) the generator is Q in Figure (4.23) the generator is $Q/5$. The fact that the second plot has bigger jumps can also be explained by the faster convergence of $\exp(tQ)$ than $\exp(tQ/5)$ for $t \rightarrow \infty$. The value of `state_0` does not influence the behavior of the process.

Figures (4.24) and (4.25) show two trajectories of the spot-rate process for `spot_rate(Q,g,1,0.05,30,2)`.

In (4.24) the generator is Q in (4.25) the generator is $Q/5$. The difference between the results can be seen, but not as obvious as before. Again the value of `state_0` does not influence the behavior of the process.

Figure 4.22: $\text{state_price}(Q, g, 1, 0.05, 30, 2)$ Figure 4.23: $\text{state_price}(Q/5, g, 1, 0.05, 30, 2)$ Figure 4.24: $\text{spot_rate}(Q, g, 1, 0.05, 30, 2)$ Figure 4.25: $\text{spot_rate}(Q/5, g, 1, 0.05, 30, 2)$

4.4 Yields in calendar time

In Section (4.2) we calculated the yields $Y(t, T)$ for fixed $t = 0$. From (4.1, Generic Approach 1) and (4.4, Generic Approach 2) we know that this calculation is only dependent on the state of X_0 and so using an N -state chain only N yield curves can ever occur. Now we try a simulation of yields over calendar time. Therefore we compute $Y(t, T)$ for $0 \leq t \leq kal$, with kal being the time horizon of calendar time and therefore get a yield curve for every $0 \leq t \leq kal$.

When we look at the history of yield curves we know that the yields of the different maturities vary over time depending on the economic development described in Section (2.2.2). For example we take the yield curves of US-treasury bonds³ from 2002 to 2007 and get following plot in Figure (4.26). The maturities are $[1/12, 0.25, 0.5, 1, 2, 3, 5, 7, 10, 20]$. I left out the yields for maturity 30 years because there was so much data missing. We see that at the beginning the structure is normal whereas at the end the structure gets flat or even inverted.

We will now simulate such structure with the help of a continuous time Markov chain.

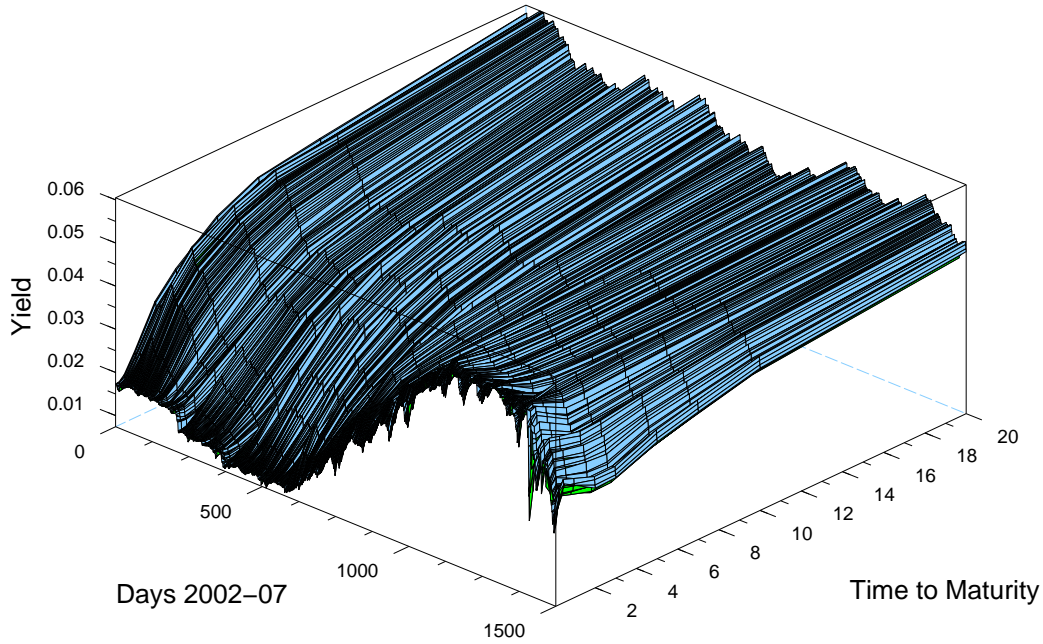


Figure 4.26: US-Dollar Yields 2002-07

Again we first choose a generating matrix Q and `state_0` for the state X_0 is in. We set kal to be the calendar time of how long we want the simulation of the chain to be. Choosing $kal = 1$ and $d = 360$ for example would mean that we do a daily simulation over 1 year. Starting from X_0 we now simulate X_t with Algorithm (4.1.2) until the maximum calendar-time kal is reached by taking step size d . So we now know which state our Markov chain is in for each point of simulation. Then we use (4.1) and (4.4) again to calculate $Y(t, T)$ depending on X_t for $0 \leq t \leq kal$.

Again we differentiate between Generic Approach 1 and Generic approach 2.

³The data was taken from

"<http://www.ustreas.gov/offices/domestic-finance/debt-management/interest-rate/yield.shtml>".

4.4.1 Generic Approach 1

The generator Q and `state_0`, the state of X_0 have already been chosen. For Generic Approach 1 we additionally need parameters g, α as in Section (4.2) as well as parameters for the maximum of calendar time and time to maturity. In this case we have to be a bit careful because a rather large amount of data is produced. So in conflict to Section (4.2), where we had a maximum time to maturity and worked with discretisation d , we restrict the maturities T_i , we want to calculate $Y(t, T)$ for, to a given vector, for example $T \in \text{mat} = [1M, 3M, 6M, 1Y, 2Y, 3Y, 5Y, 7Y, 10Y, 20Y, 30Y] = [1/12, 0.25, 0.5, 1, 2, 3, 5, 7, 10, 20, 30]$. This is also plausible as long as no bonds with maturity 5.36 years are traded on the market. When we now take the maximum calendar time kal and the step size d the matrix of the values $Y(t, T)$, $0 \leq t \leq kal$ has size $(kal*d+1)*\text{length}(\text{mat})$.

Remark 4.4.1 Hence the notation: T is the time to maturity and not the absolute time, so we compute $Y(0, T)$ for all $T_i \in \text{mat}$, depending on $X_0 = X_t, 0 \leq t \leq kal$.

These ideas are implemented in the Scilab-function "data3d_gen1..sci" which has parameters $Q, \text{state}_0, g, \alpha, \text{mat}, kal, d$ as described above. Following algorithm is implemented to calculate $Y(t, T)$, $0 \leq t \leq kal, T \in \text{mat}$:

Algorithm 4.4.1 ($Y(t, T)$ in Generic Approach 1) (1) We choose the parameters $Q, g, \text{state}_0, \alpha, \text{mat}, kal, d$ as described above.
 (2) We use Algorithm (4.1.2) with parameters $Q, \text{state}_0, kal, d$ to simulate a Markov chain of size $kal * d + 1$ and store it in vector v .
 (3) We set $i = 1$ and $j = 1$ and $f = (\alpha - Q)^{-1}g$.
 (4) We set $X_0 = v_i$.
 (5) We set $T = \text{mat}_j$ and $P_T = \exp(TQ)$.
 (6) We compute $Y(0, T)$ by using (4.1) and (4.2).
 (7) We set $j=j+1$ and go back to (5) until $\text{length}(\text{mat})$ is reached.
 (8) We set $i = i + 1$ and go back to (4) until $i=kal*d+1$.
 (9) Return: a matrix of size $(kal*d+1)*\text{length}(\text{mat})$ with the values of $Y(0, T), T \in \text{mat}$, depending on $X_0 = X_t$ for $0 \leq t \leq kal$, as entries.

4.4.2 Generic Approach 2

We use the same ideas as in Generic Approach 1. The only thing that is different is that we use a vector c that fulfills $Vc = g \geq 0$, where V is the matrix of eigenvectors of Q . The calculation that uses Generic Approach 2 is implemented in the Scilab-function "data3d_gen2.sci". Following Algorithm is implemented:

Algorithm 4.4.2 ($Y(t, T)$ in Generic Approach 2) (1) We choose the parameters $Q, c, \text{state}_0, \alpha, \text{mat}, kal, d$ as described above.
 (2) We use Algorithm (4.1.2) with parameters $Q, \text{state}_0, kal, d$ to simulate a Markov chain of size $kal * d + 1$ and store it in vector v .
 (3) We compute the matrices V and D using the Scilab-function $[V, D]=\text{spec}(Q)$.
 (4) We set $\tilde{c} = (\alpha - D)^{-1}c$.
 (5) We set $i = 1$ and $j = 1$.
 (6) We set $X_0 = v_i$.
 (7) We set $T = \text{mat}_j$.
 (8) We compute $Y(0, T)$ by using (4.4) and (4.2).
 (9) We set $j=j+1$ and go back to (7) until $\text{length}(\text{mat})$ is reached.
 (10) We set $i = i + 1$ and go back to (6) until $i=kal*d+1$.

(11) Return: a matrix of size $(kal*d+1)*length(mat)$ with the values of $Y(0,T), T \in mat$, depending on $X_0 = X_t$ for $0 \leq t \leq kal$, as entries.

4.4.3 Results

In Figures (4.27) and (4.28) we see two plots for the two functions. Q, g and c where chosen as before, the other parameters are

α	0.05
kal	3
d	50

$$mat = [1M, 3M, 6M, 1Y, 2Y, 3Y, 5Y, 7Y, 10Y, 20Y, 30Y].$$

For both plots we start with $X_0 = 1$.

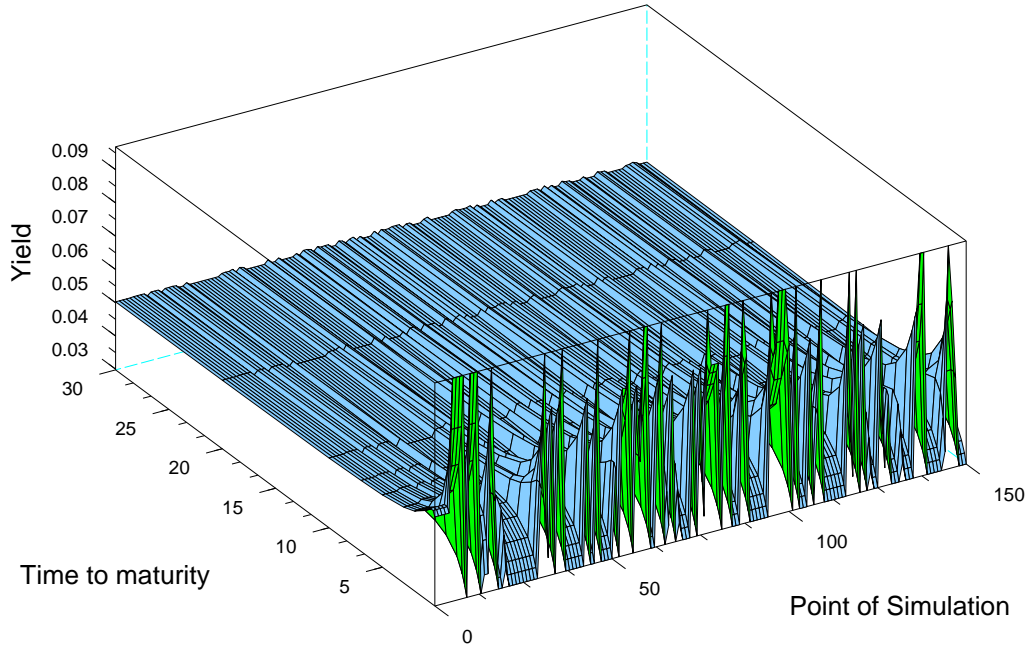


Figure 4.27: Term Structure for Generic Approach 1.

We see that for both plots the structure is very rough especially on the short end although we choose $d = 50$ which is rather big when we think of a daily simulation which means $d = 360$. We can also see that with this parameters an inverse structure can follow a normal term structure directly (or in reverse) which is not likely to occur on the market. Choosing such a simple Markov chain with just 5 states it will perhaps be possible to fit a single initial yield curve, but never a structure of yield curves for more than one day. For such a fit we would need a fine tuned Markov chain with more states (we always keep in mind that the parameter space gets large quickly), and the chain should not change state that often. Another possibility would be to use a Markov process with continuous state space.

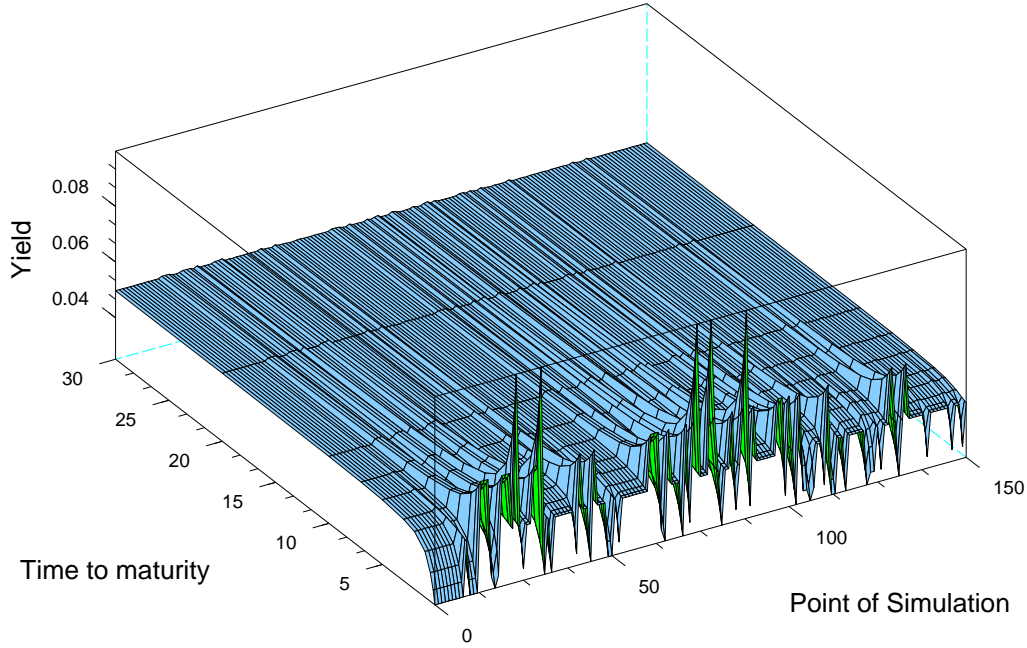


Figure 4.28: Term Structure for Generic Approach 2.

4.5 A continuous example

We recall the cosh-example of Section (3.7.4) in one dimension. Here we choose a function

$$f(x) = \cosh \gamma(x + c),$$

and get

$$g(x) = \left(\alpha - \frac{1}{2}\gamma^2\right) \cosh(\gamma(x + c)) + \beta\gamma x \sinh(\gamma(x + c)),$$

which is non-negative if α is big enough. The price of a zero-bond can be written down in closed form by

$$P(0, t) = \frac{\cosh \gamma(X_0 e^{-\beta t} + c)}{\cosh \gamma(X_0 + c)} \exp \left[-\alpha t + \frac{\gamma^2}{2} \frac{1 - e^{-2\beta t}}{2\beta} \right]. \quad (4.7)$$

So all that is left to do is to simulate the diffusion

$$dX_t = dW_t - BX_t dt, \quad (4.8)$$

in one dimension. We can do this by using the *Euler-Approximation*⁴ of an SDE. We use following theorem:

Theorem 4.5.1 *Given an Itô-diffusion*

$$dx = a(x, t)dt + b(x, t)dW_t,$$

⁴For a full account lookup [Seydel].

we can calculate approximations x_j of $x(t_j)$ by choosing t_0, x_0 , a step size Δt , $W_0 = 0$, and setting

$$t_{j+1} = t_j + \Delta t, \quad \text{and}, \quad \Delta W = Z\sqrt{\Delta t} \quad \text{with} \quad Z \sim \mathcal{N}(0, 1).$$

Then we define the approximation x_j of $x(t_j)$ by

$$x_{j+1} = x_j + a(x_j, t_j)\Delta t + b(x_j, t_j)\Delta W. \quad (4.9)$$

In our case where the diffusion is given by (4.8) we can write (4.9) as

$$x_{j+1} = x_j(1 - \beta\Delta t) + Z\sqrt{\Delta t}. \quad (4.10)$$

For this special example i computed the Markov-process together with the spot-rate process, the state-price density process and the yield curves $Y(t, T)$ together in the Scilab-function "example4.sci". This function has the parameters `state0` (the value of X_0), α, β, γ, c , as the parameters needed to compute (4.7), `mat` as the vector of maturities we want to calculate $Y(t, T)$ for, the maximum time of simulation `kal` (therefore $0 \leq t \leq kal$) and the step-size of discretisation `d`. We can use following algorithm:

Algorithm 4.5.1 (Simulation of Example (3.7.4)) (1) We choose the parameters `state0`, $\alpha, \beta, \gamma, c, mat, kal, d$ as described above.
(2) We set $X_0 = state0$.
(3) We set $i = 1$.
(4) We calculate X_i by (4.10) and set $X_0 = X_i$.
(5) We compute r_i by (4.6) and ζ_i by (4.5).
(6) We set $j = 1$.
(7) We set $T = mat_j$.
(8) We compute $Y(0, T)$ by using (4.7) and (4.2).
(9) We set $j=j+1$ and go back to (7) until `length(mat)` is reached.
(10) We set $i = i + 1$ and go back to (4) until $i=kal*d+1$.
(11) Return: a matrix of size $(kal*d+1)*length(mat)$ with the values of $Y(0, T)$, $T \in mat$, depending on $X_0 = X_t$ for $0 \leq t \leq kal$, as entries. 3 Vectors of size $kal * d + 1$ with X_j, r_j, ζ_j ; $0 \leq j \leq kal * d + 1$, as entries.

In Figure (4.29) to (4.32) we see a realisation of this special case to an underlying Markov-process with continuous state-space. The parameters are:

state0	1
α	0.05
β	0.1
γ	0.2
c	2
n	3
d	360

$$mat = [1M, 3M, 6M, 1Y, 2Y, 3Y, 5Y, 7Y, 10Y, 20Y, 30Y].$$

When we compare Figure (4.32) to (4.26), we see that using this framework we can get results very close to reality.

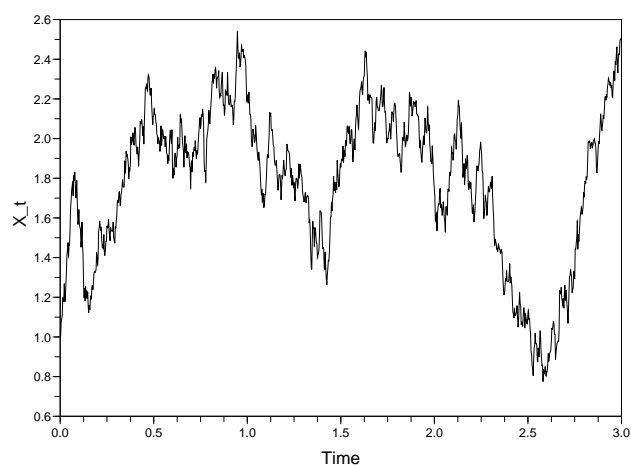


Figure 4.29: Trajectory of a continuous state-space process.

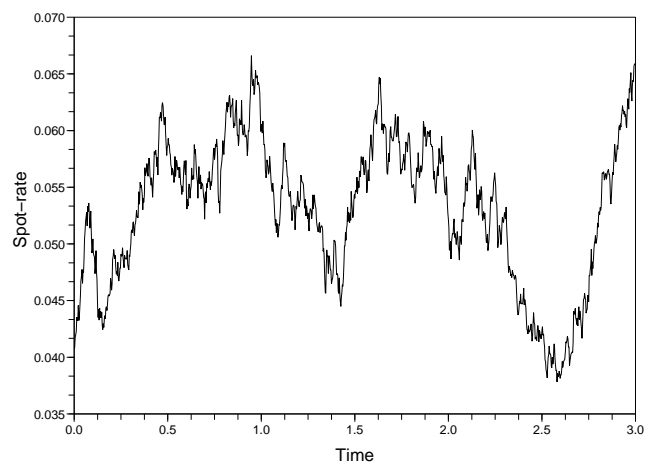


Figure 4.30: Spot-rate process, continuous case.

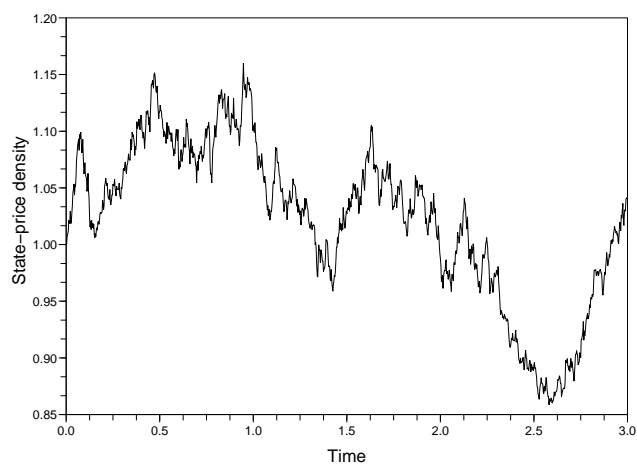


Figure 4.31: State-price density process, continuous case.

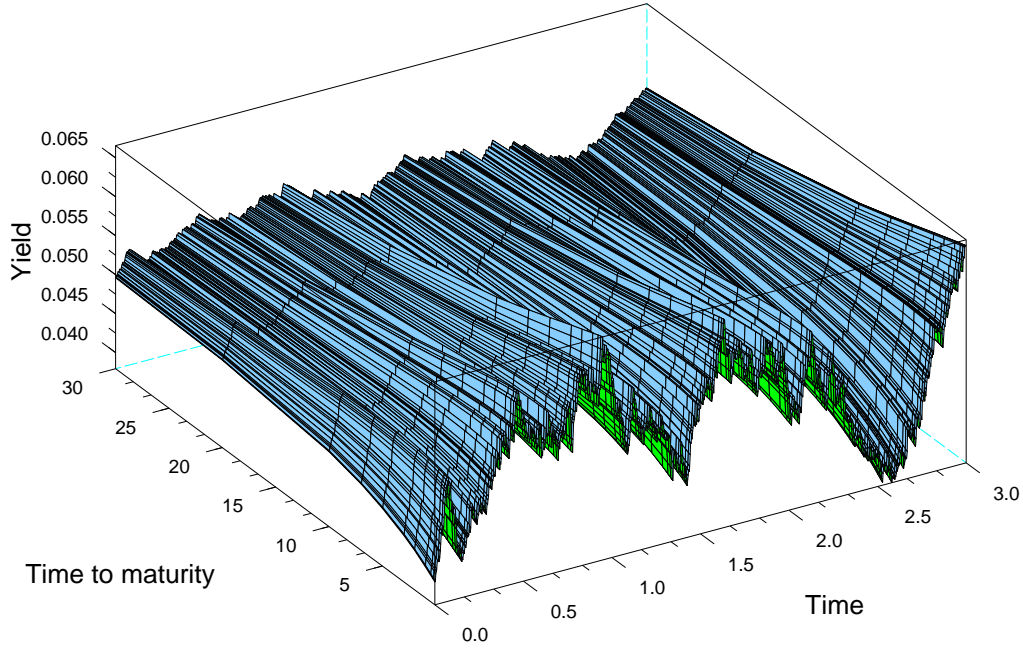


Figure 4.32: Term-structure, continuous case.

4.6 A trial fit to real data

In this section we try a short and simple fit to some data shown in Figure (4.26). We do a day-by-day calibration and simply ignore all the earlier information. Given the observations y_n on day n , we just compute

$$\min_{\theta} b(y_n - Y(x; \Theta)). \quad (4.11)$$

Therefore we use the implemented Scilab-function "datafit"⁵ to compute the solution in the least square sense.

We do the calibration for Period 1: day 1 to 100 and Period 2: day 1300 to 1399. In the first period the structure is normal, for the second the yields are more volatile.

We do not fit the parameters for every day, but make an initial fit on day 1 for each observation period and then minimize by just changing X_t , the value of the Markov process on day t . With the solutions for the parameters of the "datafit" function we again compute the prices for the different maturities and compare it to the real data. We compute the errors per day of observation and per maturity over the whole period in the sense of the supremum norm⁶. The errors in the figures are all given in basis-points⁷ (bp).

For further information about calibration lookup [Rogers 4].

⁵For an exact description lookup the Scilab Online-Help.

⁶In the case of the vector space \mathbb{R}^n , $n < \infty$ this is $\|x\|_{\infty} = \max\{|x_i| : x \in \mathbb{R}^n\}$.

⁷1bp= 10^{-4} .

4.6.1 Fit with a Markov chain

Here we make the convention that x is some distinguished state (say, the first) in the state space. The labeling of states is irrelevant under this simplifying assumption. When fitting a Markov chain with state-space $\mathcal{N} = (1, \dots, N)$ we have $N^2 + 1$ parameters to fit. Hence: the diagonal entries of the matrix Q are determined by $q_{ii} = -\sum_{i \neq j} q_{ij}$ by Theorem (1.2.5). So we have $N^2 - N$ parameters for the matrix Q , N -parameters for the vector f and one parameter for α . We can simplify the problem to time reversible chains⁸. Then for some positive vector m we have

$$m_i q_{ij} = m_j q_{ji} \quad \forall i, j.$$

Then we have $N(N-1)/2$ entries for the above diagonal elements of Q and $N-1$ parameters of m because the last entry is fixed by the fact that the entries have to sum to 1. With f and α we have in total $\frac{N^2+3N}{2}$ parameters to estimate. By restricting to time reversible chains we can also be sure that the eigenvalues of Q are all real.

In Figure (4.33) to (4.36) we see the plots by choosing a chain with 10 states which is not reversible. The result is not very satisfying because the errors are quite big, and we might better use another routine for minimization than the "datafit" function (After some experience I can say that "datafit" is very sensitive on the starting value, the initial guess, and given bounds.). Another problem is, that, in this case, we do not minimize the parameters directly to approximate $Y(0, t)$ but have to extract the generator Q out of the parameter vector and then calculate $P_t = \exp(tQ)$. Using more states than 10 would surely improve the fit, but the time of computation will grow potentially. For Period 2 I received smaller errors than for Period 1 by choosing other values for the initial guess and the bounds. I also tried to improve the fit for Period 1, but did not get a better result. The fit is done in the Scilab-script "Fit_chain.sci".

4.6.2 Continuous case

At first we have a look at Example (3.7.4). Here the bond-price was given by

$$P(0, t) = \frac{\cosh \gamma(X_0 e^{-\beta t} + c)}{\cosh \gamma(X_0 + c)} \exp \left[-\alpha t + \frac{\gamma^2}{2} \frac{1 - e^{-2\beta t}}{2\beta} \right]. \quad (4.12)$$

Including X_0 we have 5 parameters to fit. When we look at (4.12) we see that we have some restrictions to make. Of course $\beta \neq 0$, but also $\gamma \neq 0$ because otherwise (4.12) would be totally independent of X_0 . The parameters after the initial fit are:

Parameters	Period 1	Period 2
α	0.064	0.048
β	0.217	0.135
γ	0.315	0.03
c	1	1

The fit is done in the script "Fit_example4.sci". Figure (4.37) to (4.42) show the results. The errors for the more volatile Period 2 are higher than for Period 1.

When we compare the results to using a 10-state chain we see that, although we only have 5 parameters to estimate, the result is a lot better than using the chain with 101 parameters in total and also the algorithm runs faster.

⁸See Section (1.2.5) for the exact definition.

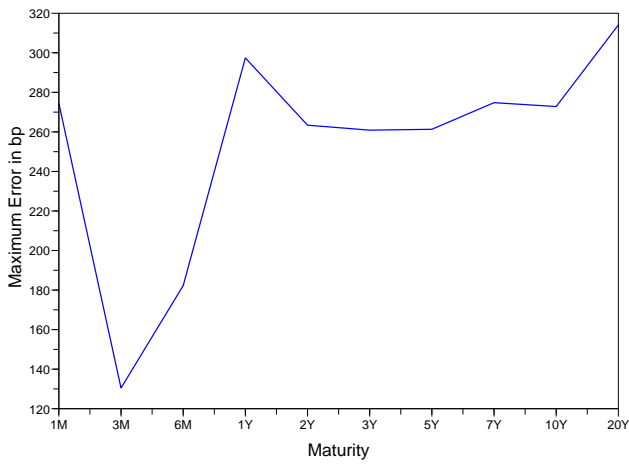


Figure 4.33: Error per maturity, Period 1

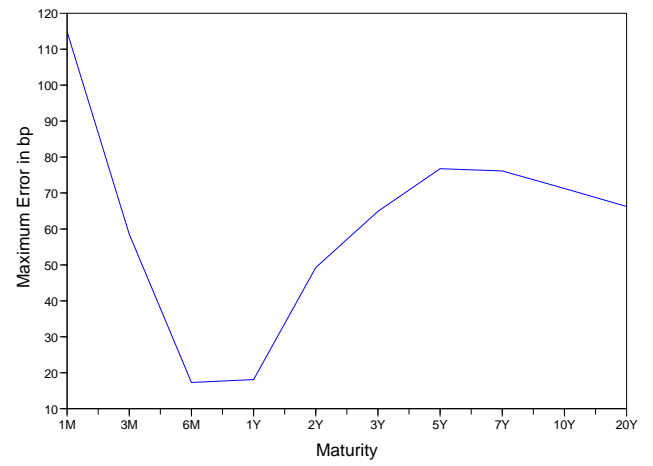


Figure 4.34: Error per maturity, Period 2

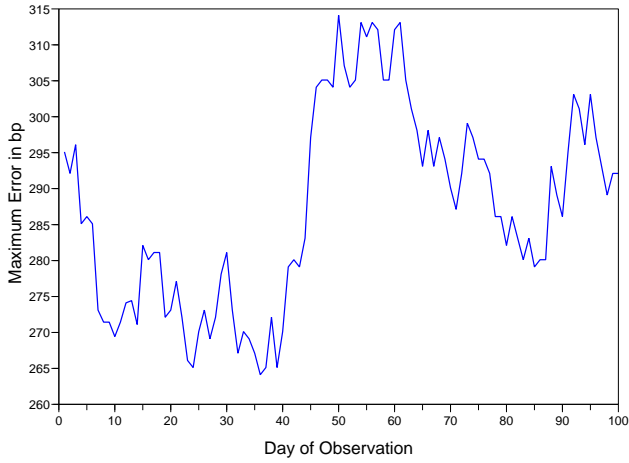


Figure 4.35: Error per day, Period 1

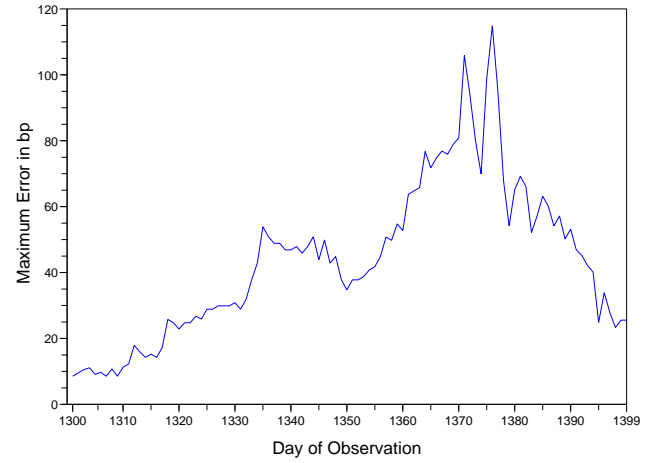


Figure 4.36: Error per day, Period 2

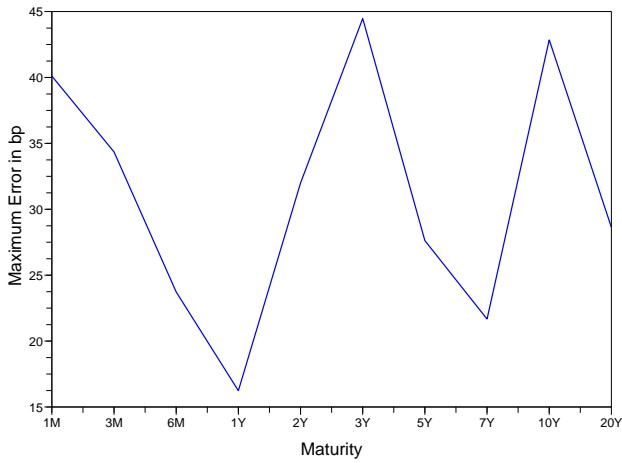


Figure 4.37: Error per maturity, Period 1

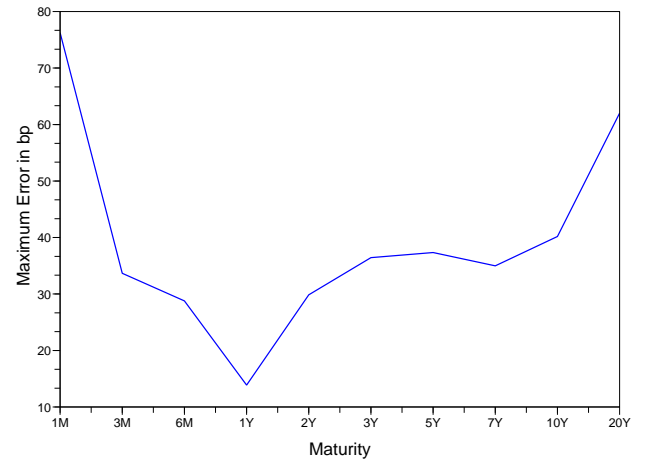


Figure 4.38: Error per maturity, Period 2

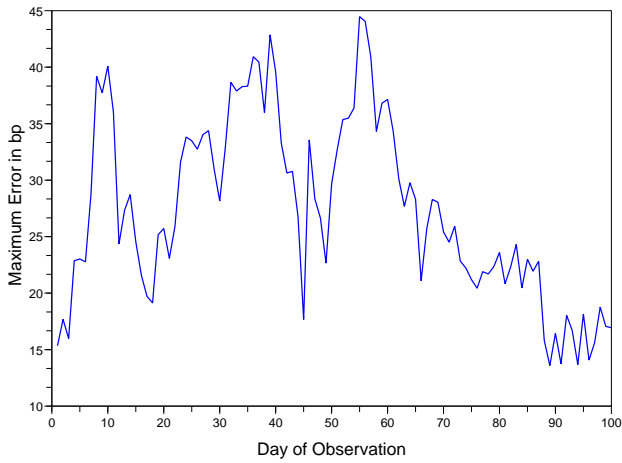


Figure 4.39: Error per day, Period 1

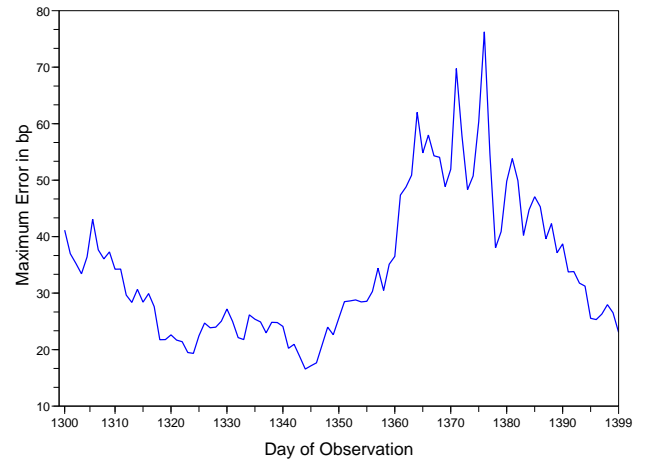


Figure 4.40: Error per day, Period 2

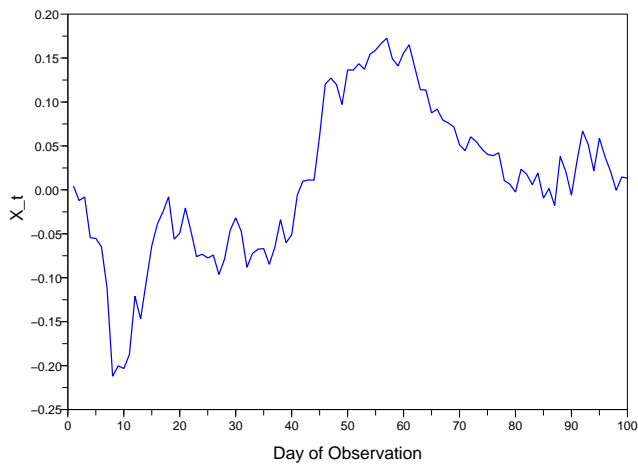


Figure 4.41: Underlying Markov process, Period 1

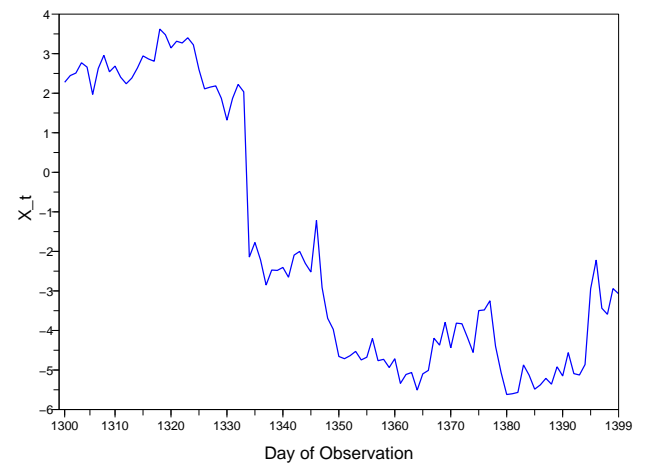


Figure 4.42: Underlying Markov process, Period 2

A Appendix

A.1 Year fraction, Day-count convention

In the following, by slight abuse of notation, t and T will denote both times, as measured by a real number from an instant chosen as time origin 0, and dates expressed as days/months/years.

In Definition (2.1.5) we defined the time to maturity $T - t$ as the amount of time (in years) from the present time t to the maturity time $T > t$.

This definition makes sense, as long as t and T are real numbers associated to two time instants. If they denote two dates expressed as day/month/year, say $D_1 = (d_1, m_1, y_1)$, $D_2 = (d_2, m_2, y_2)$, we need to define the amount of time as the number of days between. This choice is not unique and the market has several conventions, which tell us how to count and especially how to deal with holidays and weekends (non-trading days).

Definition A.1.1 (Year fraction, Day-count convention) *We denote by $\tau(t, T)$ the chosen time measure between t and T , which is usually referred to as year fraction between the dates t and T . When t and T are less than one-day distant (typically when dealing with limit quantities involving time to maturities tending to zero), $\tau(t, T)$ is to be interpreted as the time difference $T - t$ (in years), the year fraction according to the day-count convention. The particular choice that is made to measure the time between two dates reflects what is known as the day-count convention.*

To give an introduction, we now mention some examples of the day-count convention. We will then compute the year fraction between 4.1.2000 and 4.7.2000 (a leap year) and see the differences.

- **Actual/365**

With this convention a year is 365 days long and the year fraction between two dates is

$$\tau(D_1, D_2) = \frac{D_2 - D_1}{365},$$

if we denote $D_2 - D_1$ the actual number of days between the two dates, D_1 included and D_2 excluded.

In our example we get $182/365 = 0.49863$.

- **Actual/360** A year is assumed to be 360 days long. The corresponding year fraction is

$$\tau(D_1, D_2) = \frac{D_2 - D_1}{360}.$$

No we have $182/360 = 0.50556$.

- **Actual/Actual** It counts the number of whole calendar days between two dates and adds on the fractions of the year at the start and end of a period. Leap years count for 366 days, non-leap years count for 365 days.

$$\tau(D_1, D_2) = \frac{f_2 - D_1}{f_2 - f_1} + (n - 3) + \frac{D_2 - f_{n-1}}{f_n - f_{n-1}},$$

where f_i are year end dates, and $f_1 \leq D_1 \leq f_2 < \dots < f_{n-1} \leq D_2 \leq f_n$. When we apply this to our example we get $182/366 = 0.49727$.

- **30/360** With this convention, months are assumed to be 30 days long and years are 360 days long. In this case the year fraction between D_1 and D_2 is given by the following formula:

$$\tau(D_1, D_2) = \frac{\max(30 - d_1, 0) + \min(d_2, 30) + 360(y_2 - y_1) + 30(m_2 - m_1 - 1)}{360}.$$

In our example this is

$$\frac{(30 - 4) + 4 + 150}{360} = 0.5.$$

As already mentioned, several adjustments may be included, in order to leave out holidays and weekends. If D_2 is a non-trading day, it can be replaced with the first following business day. This convention is called *Modified Following* and changes the evaluation of the year fractions.

A.2 The L^p -spaces

Let M be a set and μ a positive measure on M . Further $p \in [1, \infty)$ is a constant. Then we call $L^p(\mu)$ to be the set of all (equivalence-classes μ a.e. equal) measurable functions $f : M \rightarrow \mathbb{C}$, which fulfill

$$\|f\|_p = \left(\int_M |f|^p d\mu \right)^{\frac{1}{p}} < \infty.$$

Then

$$\|\cdot\|_p : L^p(\mu) \rightarrow [0, \infty)$$

is a norm and $(L^p(\mu), \|\cdot\|_p)$ is a Banach space.

For a measurable function $f : M \rightarrow \mathbb{C}$ we call $\|f\|_\infty$ the *essential supremum* of $|f|$

$$\|f\|_\infty = \inf\{\alpha \in \mathbb{R} : \mu(f^{-1}(\alpha, \infty]) = 0\}$$

and $L^\infty(\mu)$ the set of all (equivalence-classes μ a.e. equal) measurable functions f with $\|f\|_\infty < \infty$. Again $(L^\infty(\mu), \|\cdot\|_\infty)$ is a Banach space.

If $M \subseteq \mathbb{R}^n$ and μ is the Lebesgue-measure on M we can also write $L^p(M)$ for $L^p(\mu)$.

A.3 Generator of an Itô-diffusion

Before we define what an Itô-diffusion is, we define the Itô-process and then develop Itô-formula.

Definition A.3.1 Let $(\Omega, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ be a filtered probability space and $(W_t)_{t \geq 0}$ an \mathcal{F}_t -Brownian motion. $(X_t)_{0 \leq t \leq T}$ is an \mathbb{R} -valued Itô-process if it can be written as

$$\mathbb{P} \text{ a.s. } \forall t \leq T \quad X_t = X_0 + \int_0^t K_s ds + \int_0^t H_s dW_s, \quad (\text{A.1})$$

where

- (1) X_0 is \mathcal{F}_0 -measurable.
- (2) $(K_t)_{0 \leq t \leq T}$ and $(H_t)_{0 \leq t \leq T}$ are \mathcal{F}_t -adapted processes.
- (3) $\int_0^T |K_s| ds < +\infty$ \mathbb{P} a.s.
- (4) $\int_0^T |H_s|^2 ds < +\infty$ \mathbb{P} a.s.

When we set $u(s, \omega) = K_s$ and $v(s, \omega) = H_s$ then, if X_t is an Itô-process of the form (A.1), it is often written in the shorter differential form

$$dX_t = u dt + v dB_t.$$

This notation is also the same for a multidimensional Itô-process, where

$$u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}, v = \begin{pmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{n1} & \cdots & v_{nn} \end{pmatrix},$$

and also

$$X_t = \begin{pmatrix} X_1(t) \\ \vdots \\ X_n(t) \end{pmatrix}, dB_t = \begin{pmatrix} dB_{11}(t) & \cdots & dB_{1n}(t) \\ \vdots & \ddots & \vdots \\ dB_{n1}(t) & \cdots & dB_{nn}(t) \end{pmatrix}.$$

Theorem A.3.1 Let $(X_t)_{0 \leq t \leq T}$ be an Itô-process

$$X_t = X_0 + \int_0^t K_s ds + \int_0^t H_s dW_s,$$

and f be a twice continuously differentiable function, then

$$f(X_t) = f(X_0) + \int_0^t f'(X_s) dX_s + \frac{1}{2} \int_0^t f''(X_s) d\langle X, X \rangle_s,$$

where, by definition

$$\langle X, X \rangle_t = \int_0^t H_s^2 ds,$$

and

$$\int_0^t f'(X_s) dX_s = \int_0^t f'(X_s) K_s ds + \int_0^t f'(X_s) H_s dW_s.$$

Likewise, if $(t, x) \rightarrow f(t, x)$ is a function which is twice differentiable with respect to x and once with respect to t , and if these partial derivatives are continuous with respect to (t, x) (i.e. f is a function of class $C^{1,2}$), Itô formula becomes

$$f(t, X_t) = f(0, X_0) + \int_0^t f'_s(s, X_s) ds + \int_0^t f'_x(s, X_s) dX_s + \frac{1}{2} \int_0^t f''_{xx}(s, X_s) d\langle X, X \rangle_s.$$

Now we define the Itô-diffusion:

Definition A.3.2 A (time-homogeneous) Itô-diffusion is a stochastic process $X_t(\omega) = X(t, \omega) : [0, \infty) \times \Omega \rightarrow \mathbb{R}^n$ satisfying a stochastic differential equation of the form

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t, \quad t \geq s; \quad X_s = x,$$

where B_t is a m -dimensional Brownian motion and $b : \mathbb{R}^n \rightarrow \mathbb{R}^n, \sigma : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times m}$ satisfy the condition

$$|b(x) - b(y)| + |\sigma(x) - \sigma(y)| \leq D|x - y|; \quad x, y \in \mathbb{R}^n,$$

for some constant D .

For many applications it is important that we can associate a second order partial differential operator A to an Itô-diffusion X_t . The basic connection between A and X_t is that A is the generator of the process X_t :

Definition A.3.3 Let X_t be a (time-homogeneous) Itô diffusion in \mathbb{R}^n . The (infinitesimal) generator A of X_t is defined by

$$Af(x) = \lim_{t \rightarrow 0} \frac{\mathbb{E}^x[f(X_t)] - f(x)}{t}; \quad x \in \mathbb{R}^n.$$

The set of functions $f : \mathbb{R}^n \rightarrow \mathbb{R}$ such that the limit exists at x is denoted by $\mathcal{D}_A(x)$ while \mathcal{D}_A denotes the set of functions for which the limit exists for all $x \in \mathbb{R}^n$.

After some calculation (lookup [Øksendal]) we can define the generator A of an Itô diffusion.

Theorem A.3.2 Let X_t be the Itô diffusion

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t.$$

If $f \in C_0^2(\mathbb{R}^n)$ then $f \in \mathcal{D}_A$ and

$$Af(x) = \sum_i b_i(x) \frac{\partial f}{\partial x_i} + \frac{1}{2} \sum_{i,j} (\sigma \sigma^T)_{i,j}(x) \frac{\partial^2 f}{\partial x_i \partial x_j}. \quad (\text{A.2})$$

Proof A.3.1 Lookup [Øksendal]

A.4 Code

```
//+++++ pstart.sci ++++++
//Input: stochastic vector v
//Return: vector (v_1,v_1+v_2, ... , 1)
//+++++
function vert=pstart(v)
    summe=0;
    for i=2:length(v)
        summe=v(i)+v(i-1);
        v(i)=summe;
    end
    vert=v;
endfunction

//+++++ checkq.sci ++++++
//Input: matrix Q
//Return: checks Theorem 1.2.5, 1 if true, 0 if false,
//        -1 no quadratic matrix
//+++++
function gen_check =checkq(a)
    s=size(a);
    if s(1)==s(2) then
        for i=1:s(1)
            for j=1:s(1)
                if i~=j then
                    if a(i,j)<0 then//nondiagonal elements>=0
                        gen_check=0;
                        return
                    end
                else
                    if a(i,j)>=0 then//diagonalelements <0
                        gen_check=0;
                        return
                    end
                end
            end
        end
        if abs(sum(a(i,:)))>0.0001 then//row sum=0 (rounding errors!)
            gen_check=0;
            return
        end
        gen_check=1;
    else// no quadratic matrix
        gen_check=-1;
    end
endfunction

//+++++ markov.sci ++++++
//Input: matrix Q,int state_0,int n, int d (d=1/d in code!)
//Return: vector of size nd+1 with values of X_t as entries
//+++++
function sim=markov(Q,state_0,n,d)
    select checkq(Q)//checks matrix Q
        case 0
            disp('Check matrix Q entries!');
            return
        case -1
            disp('Check matrix Q dimension!');
            return
    end
    z=size(Q);
    z=z(1);
    start=zeros(1,z);
    anzahl=n*d+1;
    sim=zeros(1,anzahl);
    d=1/d;
    if state_0==0 then
        for i=1:z
            start(i)=i/z;
        end
        beginn=1;
        t=0;
    else
        if state_0<0|state_0>z then
            disp('Wrong starting state!');
            return
        end
        beginn=2;
        t=d;
        sim(1)=state_0;
        P=expm(d*Q);
        start=pstart(P(state_0,:));
    end

    zven=rand(1,anzahl);
```

```

        for j=beginn:anzahl
            for k=1:z
                if zven(j)<=start(k) then
                    sim(j)=k;
                    break
                end
            end
            t=t+d;
            P=expm(t*Q);
            start=pstart(P(k,:));
        end
    endfunction

//+++++ bond.sci +++++//
//Input:  matrix Q, vector g, int state_0, double a,
//        int n, int d (d=1/d in code!)
//Return: vector of size nd+1 with values of P(0,t)
//        calculated with Gen.Approach 1 as entries
//+++++//
function price=bond(Q,g,state_0,a,n,d)
    select checkq(Q)//checks matrix Q
        case 0
            disp('Check matrix Q entries!');
            return
        case -1
            disp('Check matrix Q dimension!');
            return
    end
    z=size(Q);
    z=z(1);
    if state_0==0 then
        zven=rand(1,1);
        for i=1:z
            if zven<i/z then
                x_0=i;
                break
            end
        end
    else
        if state_0<0|state_0>z then
            disp('Wrong starting state!');
            return
        end
        x_0=state_0;
    end
    anzahl=n*d+1;
    price=zeros(1,anzahl);
    price(1)=1;
    f=inv(a*eye(z,z)-Q)*g;
    f_0=f(x_0);
    d=1/d;
    t=0;

    for i=2:anzahl
        t=t+d;
        P=expm(t*Q);
        product=P*f;
        price(i)=exp(-a*t)*product(x_0)/f_0;
    end
endfunction

//+++++ yield.sci +++++//
//Input:  matrix Q, vector g, int state_0, double a,
//        int n, int d (d=1/d in code!)
//Return: vector of size nd+1 with values of Y(0,t)
//        calculated with Gen.Approach 1 as entries
//+++++//
function zins=yield(Q,g,state_0,a,n,d)
    bond_vector=bond(Q,g,state_0,a,n,d);
    l=length(bond_vector);
    zins=zeros(1,l);
    z=size(Q);
    z=z(1);
    if state_0<0|state_0>z then
        disp('Wrong starting state!');
        return
    end
    f=inv(a*eye(z,z)-Q)*g;
    zins(1)=g(state_0)/f(state_0);//Y(0,0)=r_0
    d=1/d;
    t=0;
    for i=2:l
        t=t+d;
        zins(i)=-log(bond_vector(i))/t;
    end
endfunction

```

```

//+++++++ bond_yield_gen2.sci ++++++//
//Input: matrix Q, vector c, int state_0, double a, int n,
//        int d (d=1/d in code!)
//Return: 2 vectors of size nd+1 with values of P(0,t) and Y(0,t)
//        calculated by Gen.Approach 2 as entries
//+++++++//
function [gen2_b,gen2_y]=bond_yield_gen2(Q,c,state_0,a,n,d)
    select checkq(Q)
        case 0
            disp('Check matrix Q entries!');
            return
        case -1
            disp('Check matrix Q dimension!');
            return
    end

    z=size(Q);
    z=z(1);
    start=zeros(1,z);
    anzahl=n*d+1;
    gen2_b=zeros(1,anzahl);
    gen2_b(1)=1;
    gen2_y=zeros(1,anzahl);
    [V,D]=spec(Q);
    t=0;
    d=1/d;

    if state_0==0 then
        zven=rand(1,1);
        for i=1:z
            if zven<i/z then
                x_0=i;
                break
            end
        end
    else
        if state_0<0|state_0>z then
            disp('Wrong starting state!');
            return
        end
        x_0=state_0;
    end

    c=c*inv(a*eye(z,z)-D);//c snail
    phi=V(x_0,:);
    gen2_y(1)=(c*(a*eye(z,z)-D))*phi/(c*phi);//Y(0,0)=r_0

    for j=2:anzahl
        t=t+d;
        gen2_b(j)=exp(-a*t)*(c*expm(t*D))*phi/(c*phi);
        gen2_y(j)=-log(gen2_b(j))/t;
    end
endfunction

//+++++++ spot_rate.sci ++++++//
//Input: matrix Q, int state_0, double a,int n,
//        int d (d=1/d in code!)
//Return: vector of size nd+1 with values of r_t
//        using 4.6 as entries
//+++++++//
function rate=spot_rate(Q,g,state_0,a,n,d)
    anzahl=n*d+1;
    process=zeros(1,anzahl);
    process=markov(Q,state_0,n,d);
    rate=zeros(1,anzahl);
    z=size(Q);
    z=z(1);
    f=inv(a*eye(z,z)-Q)*g;

    for i=1:anzahl
        rate(i)=g(process(i))/f(process(i));
    end
endfunction

//+++++++ state_price.sci ++++++//
//Input: matrix Q, int state_0, double a,int n,
//        int d (d=1/d in code!)
//Return: vector of size nd+1 with values of zeta_t
//        using 4.5 as entries
//+++++++//
function zeta=state_price(Q,g,state_0,a,n,d)
    anzahl=n*d+1;
    process=zeros(1,anzahl);
    process=markov(Q,state_0,n,d);
    zeta=zeros(1,anzahl);
    zeta(1)=1;
    z=size(Q);

```

```

    z=z(1);
    f=inv(a*eye(z,z)-Q)*g;
    d=1/d;
    t=0;

    for i=2:anzahl
        t=t+d;
        zeta(i)=exp(-a*t)*f(process(i))/f(process(1));
    end
endfunction

//+++++ data3d_gen1.sci +++++//
//Input: matrix Q, vector g, int state_0, double a,vector mat,
//        int kal, integer d (d=1/d in code!)
//Return: 1)data = matrix of size size(mat)*(kal*d)+1 with values of Y(t,T)
//         calculated with Gen.Approach 1 as row entries
//         2)kal_vec= vector of size kal*d+1 with calendar times as entries
//+++++//
function [kal_vec,data]=data3d_gen1(Q,g,state_0,a,mat,kal,d)
    kal_zeit=kal*d+1;
    process=zeros(1,kal_zeit);
    process=markov(Q,state_0,kal,d);
    len=length(mat);
    data=zeros(kal_zeit,len);
    yield_v=zeros(1,len);
    z=size(Q);
    z=z(1);
    f=inv(a*eye(z,z)-Q)*g;
    m=max(mat);
    kal_vec=zeros(1,kal_zeit);
    kal_vec=0:1:kal_zeit-1;

    for i=1:kal_zeit
        for j=1:len
            t=mat(j);
            P=expm(t*Q);
            product=P*f;
            yield_v(j)=a-log(product(process(i))/f(process(1)))/t;
        end
        data(i,:)=yield_v;
    end
endfunction

//+++++ data3d_gen2.sci +++++//
//Input: matrix Q, vector c, int state_0, double a,vector mat,
//        int kal, int d (d=1/d in code!)
//Return: 1)data = matrix of size size(mat)*(kal*d)+1 with values of Y(t,T)
//         calculated with Gen.Approach 2 as row entries
//         2)kal_vec= vector of size kal*d+1 with calendar times as entries
//+++++//
function [kal_vec,data]=data3d_gen2(Q,c,state_0,a,mat,kal,d)
    z=size(Q);
    z=z(1);
    kal_zeit=kal*d+1;
    process=zeros(1,kal_zeit);
    process=markov(Q,state_0,kal,d);
    [V,D]=spec(Q);
    len=length(mat);
    m=max(mat);
    kal_vec=zeros(1,kal_zeit);
    kal_vec=0:1:kal_zeit-1;
    data=zeros(kal_zeit,len);
    yield_v=zeros(1,len);
    c_sl=c*inv(a*eye(z,z)-D);

    for i=1:kal_zeit
        phi=V(process(i),:);
        f_0=V*c_sl';
        g_0=V*c';
        for j=1:len
            t=mat(j);
            yield_v(j)=a-log((c_sl*expm(t*D)*phi)/(c_sl*phi))/t;
        end
        data(i,:)=yield_v;
    end
endfunction

```

```

//+++++ example4.sci ++++++
//Input: double \texts{state0},a,b,g,c,vector mat, int n, int d (d=1/d in code!)
//Return: 1)yields = matrix of size size(mat)*(n*d)+1 with values of Y(t,T)
//         calculated with Gen.Approach 1 as row entries
//         2)3 Vectors of size n*d+1 with X_j,r_j and zeta_j as entries
//+++++
function [proc,r,zeta,yields,anz]=example4(\texts{state0},a,b,gam,c,mat,n,d)
    anzahl=n*d+1;
    proc=zeros(1,anzahl); //vector of the X_t
    r=zeros(1,anzahl); //vector of the r_t
    zeta=zeros(1,anzahl); //vector of the zeta_t
    yields=zeros(anzahl,length(mat)); // matrix of the Y(t,T)
    anz=zeros(1,anzahl); //timestamp
    anz=0:1:anzahl-1;
    d=1/d;
    t=0;
    f_0=cosh(gam*(\texts{state0}+c));
    proc(1)=\texts{state0};

    for i=1:anzahl
        rve=rand(1,1,'normal');
        if i==1 then
            proc(i)=\texts{state0};
        else
            proc(i)=proc(i-1)*(1-b*d)+rve*sqrt(d);
        end
        f=cosh(gam*(proc(i)+c));
        g=cosh(gam*(proc(i)+c))*(a-1/2*gam^2)+sinh(gam*(proc(i)+c))*b*gam*proc(i);

        if g<0|f<=0 then //f>=0,g>0!!
            disp('Wrong parameters, negative f,g!!');
            return
        end

        r(i)=g/f; //spot-rate
        zeta(i)=exp(-a*t)*f/f_0; //state-price density
        t=t+d;

        for j=1:length(mat) //calculate the yield curve for Y(t,T)
            yields(i,j)=-(log(cosh(gam*(proc(i)*exp(-b*mat(j))+c))/cosh(gam*(proc(i)+c)))
                -a*mat(j)+gam^2/2*(1-exp(-2*b*mat(j)))/(2*b))/mat(j);
        end
    end
endfunction

//+++++ Fit_example4.sci ++++++
//Fits example 4 to real data
//+++++

//initialize the used functions
function yy=exmp4(x,p) //pricing function for example 4, Initial fit
    yy=-(log(cosh(p(4)*(p(1)*exp(-p(3)*x)+p(5)))/cosh(p(4)*(p(1)
        +p(5))))-p(2)*x+(p(4)^2/2*(1-exp(-2*p(3)*x))/(2*p(3)))/x;
endfunction

function e=Gg(p,z) //criterion function, Initial case
    y=z(1),x=z(2);
    e=y-exmp4(x,p);
endfunction

//load the data
load('C:\Programme\scilab-4.1.2\Diplom\data.sav'); disp('file loaded');
//choose period
M=y_data(1:100,:);
Z=zeros(1,length(mat)); //maturities vector
//errors
total_error=zeros(1,100); //error for yield curve on day i
mat_error=zeros(1,length(mat)); //error per maturity
//bounds
xinf=[-20,0,0.0000000001,0.03,-20]'; //beta and gamma ~0!!
xsup=[20,10,10,10,20]';
//initial guess
p0=[0,0.05,0.1,0.05,1]';

Z(1,:)=M(1,:); //Initial fit for all parameters
[p,err]=datafit(Gg,Z,"b",xinf,xsup,p0);
disp(p);
//define parameters ~x_0
alpha=p(2);
bet=p(3);
gamm=p(4);
c_=p(5);

function y=x0_fit(x,p) //fit x0
    y=-(log(cosh(gamm*(p(1)*exp(-bet*x)+c_))/cosh(gamm*(p(1)+c_)))
        -alpha*x+(gamm^2/2*(1-exp(-2*bet*x))/(2*bet)))/x;

```

```

endfunction

function e=Ggg(p,z)//criterion function for fitting x_0
    y=z(1),x=z(2);
    e=y-x0_fit(x,p);
endfunction

//bounds and guess for x_0
p0=[p(1)];
xinf=[-50];
xsup=[50];

x_t=zeros(1,100);//vector of the fitted x_t

for i=1:100
    Z(1,:)=M(i,:);
    [x_0,err]=datafit(Ggg,Z,'b',xinf,xsup,p0);
    x_t(i)=x_0;
    for j=1:length(mat)
        err=abs(Z(1,j)-x0_fit(mat(j),x_0));
        if err>total_error(i) then//take sup_norm for the errors
            total_error(i)=err;
        end
        if err>mat_error(j) then
            mat_error(j)=err;
        end
    end
    p0=x_0;//take new parameter values as new initial guess
    disp(i);
end

//error in basis points!!
total_error=total_error*10000;
mat_error=mat_error*10000;
disp('successfull !!');

//+++++ Fit_chain.sci ++++++//
//Fits continous time Markov chain to real data
//+++++//
number_of_states=10;//set number of states
anz_par=number_of_states^2+1;//set parameter size

//italize the used functions
function fit=gen1(x,p)
    dim=sqrt(length(p)-1);//choose size of generating matrix!
    Q=zeros(dim,dim);
    f=zeros(1,dim);
    for i=1:dim//extract Q and f from parameter vector
        l=(i-1)*(dim-1)+1;
        r=i*(dim-1);
        Q(i,i)=-sum(p(l:r));
        f_s=dim*(dim-1)+i;
        f(i)=p(f_s);
        count=1;
        for j=1:dim
            if i~=j then
                q_j=(i-1)*(dim-1)+count;
                Q(i,j)=p(q_j);
                count=count+1;
            end
        end
    end
    a_s=dim^2+1
    a=p(a_s);//extract alpha
    product=expm(x*Q)*f';
    fit=-log(exp(-a*x)*product(1)/f(1))/x;//we postulate we are in state 1
endfunction

function e=Gc(p,z)//criterion function
    y=z(1),x=z(2);
    e=y-gen1(x,p);
endfunction

//load the data
load('C:\Programme\scilab-4.1.2\Diplom\data.sav');disp('file loaded');
//choose period
M=y_data(1300:1399,:);
Z=[zeros(1,length(mat));mat];//
//errors
total_error=zeros(1,100);//error for yield curve on day i
mat_error=zeros(1,length(mat));//error per maturity over 100 days
//bounds
xinf=0.01*ones(1,anz_par)';
xsup=20*ones(1,anz_par)';//P1: 50
xsup(anz_par)=0.5;//max alpha=0.5//P1: 0.2
//initial guess
p0=0.1*ones(1,anz_par)';//P1: 0.4

```



```

Z(1,:)=M(1,:); //initial fit for all parameters
[p,err]=datafit(Gc,Z,"b",xinf,xsup,p0);

function [a,g,q]=extract(v)//extracts Q,f,alpha from parameter vector
    dim=sqrt(length(v)-1);
    a=v(length(v));
    g=zeros(1,dim);
    q=zeros(dim,dim);
    count=1;
    for i=1:dim
        s=dim*(dim-1)+i;
        g(i)=v(s);
        for j=1:dim
            if i~=j then
                q(i,j)=v(count);
                count=count+1;
            end
        end
        q(i,i)=-sum(q(i,:));
    end
endfunction

[alpha,f,Q]=extract(p);
x_t=zeros(1,100); //vector of the fitted x_t

function w=price(Q,f,alpha,t,state)
    product=expm(t*Q)*f';
    w=-log(exp(-alpha*t)*product(state)/f(state))/t;
endfunction

materrs=zeros(number_of_states,length(mat));

for i=1:100
    Z(1,:)=M(i,:); //take data row i
    ma=1000;
    for z=1:number_of_states
        for j=1:length(mat)
            err=abs(Z(1,j)-price(Q,f,alpha,mat(j),z));
            materrs(z,j)=err;
        end
        sup=max(materrs(z,:));
        if sup<ma then
            ma=sup;
            opti=z;
        end
    end
    total_error(i)=ma;
    x_t(i)=opti;
    for j=1:length(mat)
        if materrs(opti,j)>mat_error(j) then
            mat_error(j)=materrs(opti,j);
        end
    end
end

//error in basis points!!
total_error=total_error*10000;
mat_error=mat_error*10000;
disp('successfull !!');

```

Bibliography

- [Bauer 1] Heinz Bauer, *Maß- und Integrationstheorie*, (De Gruyter, 1990).
- [Bauer 2] Heinz Bauer, *Wahrscheinlichkeitstheorie 4. Auflage*, (De Gruyter, 1991).
- [Blumenthal/Gettoor] R. Blumenthal, R. Gettoor, *Markov Processes and Potential Theory*, (Academic Press, 1968).
- [Brigo] D. Brigo, F. Mercurio, *Interest Rate Models-Theory and Practice*, (Springer Finance, 2001).
- [Cairns] A. Cairns, *Interest Rate Models, An Introduction*, (Princeton Paperbacks, 2004).
- [James/Webber] J. James, N. Webber, *Interest rate modeling*, (John Wiley and Sons, 2000).
- [Kijima] M. Kijima, *Markov processes for stochastic modeling*, (Chapman & Hall, 1997).
- [Lamberton/Lapeyre] D. Lamberton, B. Lapeyre, *Introduction to Stochastic Calculus Applied to Finance*, (Chapman & Hall, 1996).
- [Miller] K. Miller, *Multidimensional Gaussian Distributions*, (John Wiley and Sons, 1964).
- [Mishkin] F. Mishkin, *Financial markets and institutions*, (Addison-Wesley, 1998).
- [Musiela] M. Musiela, M. Rutkowski, *Martingale Methods in Financial Modelling*, (Springer, 1997).
- [Øksendal] B. Øksendal, *Stochastic Differential Equations*, (Springer, 2002).
- [Rogers 1] L.C.G. Rogers, *One for All*, (European Congress on Mathematics Industry, 2004).
- [Rogers 2] L.C.G. Rogers, *The Potential Approach to the term structure of interest rates and foreign exchange rates*, (Mathematical Finance 7, 1997).
- [Rogers 3] L.C.G. Rogers, *Fitting Potential Models to Interest rate and foreign exchange Data*, (Risk Publications, London, 1997).
- [Rogers 4] L.C.G. Rogers, *Markov chains and the potential approach to modeling interest rates and exchange rates*, (Mathematical Finance-Bachelier Congress, 2000).
- [Rogers/Williams] L. Rogers, D. Williams, *Diffusions, Markov Processes and Martingales Vol.1,2*, (Cambridge Mathematical Library, 1993).
- [Seydel] R. Seydel, *Einführung in die numerische Berechnung von Finanz-Derivaten*, (Springer, 1999).
- [Wertz] W. Wertz, *Theorie stochastischer Prozesse*, (Lecture at TU Wien SS, 2004).

List of Figures

2.1	Normal term structure	24
2.2	Inverse term structure	24
2.3	Flat term structure	24
2.4	Humped term structure	24
4.1	Trajectory of a finite Markov chain	45
4.2	Trajectory of a finite Markov chain	45
4.3	Bond Price	47
4.4	Yield curve	49
4.5	Yield curve	49
4.6	Yield curve	49
4.7	Yield curve	49
4.8	Yield curve	49
4.9	Yield curve	49
4.10	Yield curve	52
4.11	Yield curve	52
4.12	Yield curve	52
4.13	Yield curve	52
4.14	Yield curve	52
4.15	Yield curve	52
4.16	Yield curve	53
4.17	Yield curve	53
4.18	Yield curve	53
4.19	Yield curve	53
4.20	Yield curve	53
4.21	Yield curve	53
4.22	State-price density	55
4.23	State-price density	55
4.24	Spot-rate	55
4.25	Spot-rate	55
4.26	US-Dollar Yields 2002-07	56
4.27	Yield Structure	58
4.28	Yield Structure	59
4.29	continuous State-space Process	61
4.30	Spot-rate, continuous case	61
4.31	State-price density, continuous case	61
4.32	Term-structure, continuous case	62
4.33	Fit to 10-state chain	64
4.34	Fit to 10-state chain	64
4.35	Fit to 10-state chain	64
4.36	Fit to 10-state chain	64
4.37	Fit to continuous example	64
4.38	Fit to continuous example	64
4.39	Fit to continuous example	65

4.40	Fit to continuous example	65
4.41	Fit to continuous example	65
4.42	Fit to continuous example	65