# Notes on stochastic processes

## James Stokes

### December 11, 2016

# Contents

# 1  Probability theory

## 1.1  Probability space

The sample space $\Omega$ is the set of all outcomes of an experiment. An event $A$ is a subset of $\Omega$. An event $A$ is said to occur if the outcome of the experiment lies in $A$. A probability distribution is a function on the collection of events such that $P(\Omega) = 1$, $P(A) \geq 0$ for all $A$ and for any countable union of disjoint sets

$$P\Big(\coprod_i A_i\Big) = \sum_i P(A_i). \tag{1}$$

It follows from the above axioms that $P(A) \leq 1$ since $1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$ and thus since $P(A^c) \geq 0$. By expressing $A \cup B$ as a disjoint union of sets,

$$A \cup B = (A \cap B^c) \sqcup (A^c \cap B) \sqcup (A \cap B) \tag{2}$$

we obtain

$$P(A \cup B) = P(A \cap B^c) + P(A^c \cap B) + P(A \cap B) \tag{3}$$
$$= [P(A) + P(B) - 2P(A \cap B)+] + P(A \cap B) \tag{4}$$
$$= P(A) + P(B) - P(A \cap B). \tag{5}$$

## 1.2   Conditional probability

Denote by $P(A|B)$ the probability of $A$ occurring given that $B$ has occurred. This is the probability that the outcome of the experiment lies in $A$ given that the outcome is known to lie in $B$ It is given by the formula

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{6}$$

The above formula has the corollary

$$P(A \cap B) = P(B)P(A|B) = P(A)P(B|A). \tag{7}$$

This can be thought of as a generalization of the formula $P(A \cap B) = P(A)P(B)$ for independent events. Iterating the definition of conditional probability we obtain

$$P(A \cap B \cap C) = P(A \cap B)P(C|A \cap B) \tag{8}$$
$$= P(A)P(B|A)P(C|A \cap B). \tag{9}$$

**Example 1.1.** A coin is tossed twice. Find the probability of the event of two heads given (a) the first toss is a head (b) at least one toss is a head. We have

$$P(H_1 \cap H_2|H_1) = \frac{P(H_1 \cap H_2 \cap H_1)}{P(H_1)} = \frac{P(H_1 \cap H_2)}{P(H_1)} \tag{10}$$

$$= \frac{p^2}{p^2 + p(1-p)} = p \tag{11}$$

$$P(H_1 \cap H_2|H_1 \cup H_2) = \frac{P(H_1 \cap H_2 \cap (H_1 \cup H_2))}{P(H_1 \cup H_2)} = \frac{P(H_1 \cap H_2)}{P(H_1 \cup H_2)} \tag{12}$$

$$= \frac{p^2}{p^2 + 2p(1-p)} = \frac{p}{1-p}. \tag{13}$$

Since $0 \leq p \leq 1$ we see that (a) is at least as likely as (b).

### 1.2.1 Bayes' theorem

Suppose we partition the sample space into a set of disjoint events called priors $\sqcup_i A_i = \Omega$. Then the partition theorem states that

$$P(B) = \sum_i P(B|A_i)P(A_i). \tag{14}$$

Using this fact allows us to express the conditional probability $P(A_i|B)$ in terms of the conditional probabilities $P(B|A_j)$ as follows

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum_j P(A_j)P(B|A_j)}. \tag{15}$$

Applying the partition theorem to the expectation of a discrete random variable $X$ gives

$$\mathbb{E}[X] = \sum_i \sum_x x P(X = x|A_i)P(A_i) \tag{16}$$

$$\sum_i \mathbb{E}[X \mid A_i]P(A_i). \tag{17}$$

Let $f_{X,Y}(x,y)$ be the joint probability density for two random variables $X$ and $Y$. The conditional probability density is defined as

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}. \tag{18}$$

**Example 1.2.** (Monty hall problem). Suppose without loss of generality that the contestant chooses door 1 (if the contestant chooses any other door then we just relabel the doors so that the contestant chose door 1). We clearly have

$$P(\text{behind } 1) = P(\text{behind } 2) = P(\text{behind } 3) = \frac{1}{3}. \tag{19}$$

Now consider the probabilities conditioned on the fact that the host chooses a door. Note that the host can only choose door 2 or 3 since the contestant has selected door 1. We have

$$P(\text{behind } 1|\text{chooses } 2) = \frac{P(\text{chooses } 2|\text{behind } 1)P(\text{behind } 1)}{P(\text{chooses } 2)} = \frac{\frac{1}{2}\frac{1}{3}}{\frac{1}{2}} = 1/3 \tag{20}$$

$$P(\text{behind } 1|\text{chooses } 3) = \frac{P(\text{chooses } 3|\text{behind } 1)P(\text{behind } 1)}{P(\text{chooses } 3)} = \frac{\frac{1}{2}\frac{1}{3}}{\frac{1}{2}} = 1/3. \tag{21}$$

Now

$$P(\text{behind } 2|\text{chooses } 3) = \frac{P(\text{chooses } 3|\text{behind } 2)P(\text{behind } 2)}{P(\text{chooses } 3)} = \frac{1\frac{1}{3}}{\frac{1}{2}} = 2/3 \tag{22}$$

$$P(\text{behind } 3|\text{chooses } 2) = \frac{P(\text{chooses } 2|\text{behind } 3)P(\text{behind } 3)}{P(\text{chooses } 2)} = \frac{1\frac{1}{3}}{\frac{1}{2}} = 2/3. \tag{23}$$

The only remaining possibilities have vanishing probability because the host will never choose the door behind which the prize lies; that is,

$$P(\text{behind } 2|\text{chooses } 2) = P(\text{behind } 3|\text{chooses } 3) = 0 \tag{24}$$

## 1.3 Discrete probability distributions

### 1.3.1 Bernoulli

Consider a biased coin that lands heads with probability $p$. This experiment can be modeled by a Bernoulli distribution which describes the probability of success in a single experiment with success probability $p$. A Bernoulli random variable $X \sim \mathrm{B}(1,p)$ has probability distribution

$$P(X = 1) = p, \qquad P(X = 0) = 1 - p. \tag{25}$$

The mean and variance are easily calculated to be

$$\mathbb{E}[X] = p1 + (1-p)0 = p, \qquad \mathrm{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = p - p^2 = p(1-p). \tag{26}$$

### 1.3.2 Binomial and Geometric

Now consider a sequence of $k$ biased coin tosses where heads lands with probability $p$. This can can be modeled as a sequence of $k$ independent and identically distributed Bernoulli random variables $(X_1, \ldots, X_k)$ with common success probability $P(X_i = 1) = p$. The binomial distribution $\mathrm{B}(k,p)$ characterizes the number of heads that land after $k$ tosses. It also describes the number of successes in $k$ draws with replacement, where the success probability on each draw is $p$. A binomial random variable $X \sim \mathrm{B}(k,p)$ has probability distribution

$$P(X = j) = \binom{k}{j} p^j (1-p)^{k-j}. \tag{27}$$

This gives the probability of $j$ heads appearing in $k$ flips of an unfair coin. The $\binom{k}{i}$ factor counts the number of ways that the $i$ heads can be distributed in the sequence of $k$ trials. Note that the binomial random variable can be represented as a sum of $k$ iid Bernoulli random variables

$$X = \sum_{j=1}^{k} X_j \tag{28}$$

and thus

$$\mathbb{E}[X] = kp, \tag{29}$$

$$\mathrm{var}(X) = \sum_{i=1}^{k} \sum_{j=1}^{k} \mathrm{cov}(X_i, X_j) = \sum_{i=1}^{k} \mathrm{var}(X_i) = kp(1-p). \tag{30}$$

The geometric distribution $\mathrm{Geom}(p)$ describes the probability of success on the $k$th trial after $k-1$ failures with success probability $p$. Given that a biasses coin toss yields tails, the binomial distribution characterizes the number of tosses until the next head. A geometric random variable $X \sim \mathrm{Geom}(p)$ has probability distribution

$$P(X = k) = (1-p)^{k-1} p. \tag{31}$$

An example is the probability that an unfair coin yields heads on the $k$th flip if the probability of heads is $p$. The expectation value

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} P(X = k)k. \tag{32}$$

can be computed by differentiating the following identity with respect to $p$,

$$1 = \sum_{k=1}^{\infty} (1-p)^{k-1}p \tag{33}$$

$$0 = \sum_{k=1}^{\infty} \left[ (1-p)^{k-1} - (k-1)p(1-p)^{k-2} \right] \tag{34}$$

$$= 1/p - \mathbb{E}[X]. \tag{35}$$

It can also be computed by representing it in terms of conditional expectations

$$\mathbb{E}[X] = P(X = 1)\mathbb{E}[X \mid X = 1] + P(X > 1)\mathbb{E}[X \mid X > 1] \tag{36}$$

$$= P(X = 1)\mathbb{E}[X \mid X = 1] + P(X > 1)\left(\mathbb{E}[X - 1 \mid X - 1 > 0] + 1\right) \tag{37}$$

$$= p + (1-p)\left(\mathbb{E}[X] + 1\right) \tag{38}$$

where we have used the memorylessness property $P(X - n | X > n) = P(X)$. Solving gives

$$\mathbb{E}[X] = 1/p. \tag{39}$$

The variance can be computed using the trick

$$\mathrm{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2 \tag{40}$$

We have

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} k(k-1)(1-p)^{k-1}p \tag{41}$$

$$= p(1-p)[1 \cdot 2 + 2 \cdot 3(1-p)^2 + \cdots] \tag{42}$$

$$= \frac{2p(1-p)}{[1-(1-p)]^3} \tag{43}$$

and thus

$$\mathrm{var}(X) = \frac{1-p}{p^2}. \tag{44}$$

**Example 1.3.** Let $X$ represent the number of tosses until two successive heads appear in a row. We partition on first and second tosses and use the memorylessness property $\mathbb{E}[X \mid T] = \mathbb{E}[X] + 1$, $\mathbb{E}[X \mid HT] = \mathbb{E}[X] + 2$ to give

$$\mathbb{E}[X] = \mathbb{E}[X \mid T]P(T) + \mathbb{E}[X \mid T]P(T) \tag{45}$$

$$= \mathbb{E}[X \mid T]P(T) + \mathbb{E}[X \mid HT]P(HT) + \mathbb{E}[X \mid HH]P(HH) \tag{46}$$

$$= (1 + \mathbb{E}[X])(1-p) + (2 + \mathbb{E}[X])p(1-p) + 2p^2. \tag{47}$$

This implies

$$\mathbb{E}[X] = 1/p + 1/p^2. \tag{48}$$

Alternatively let $X_n$ denote the number of flips required to obtain $n$ successive heads. We have

$$\mathbb{E}[X_2] = \sum_{k=1}^{\infty} \mathbb{E}[X_2 \mid X_1 = k] P(X_1 = k) \tag{49}$$

Now

$$\mathbb{E}[X_2 \mid X_1 = k] = p(k+1) + (1-p)(k+1+\mathbb{E}[X_2]) \tag{50}$$
$$= 1 + k + (1-p)\mathbb{E}[X_2] \tag{51}$$

and thus

$$\mathbb{E}[X_2] = \sum_{k=1}^{\infty} P(X_1 = k) \left[1 + k + (1-p)\mathbb{E}[X_2]\right] \tag{52}$$
$$= 1 + \mathbb{E}[X_1] + (1-p)\mathbb{E}[X_2]. \tag{53}$$

Setting $\mathbb{E}[X_1] = 1/p$ and solving we obtain the result.

### 1.3.3 Uniform

The discrete uniform distribution on $[a, b]$ (where $a, b \in \mathbb{Z}$) is given by

$$P(X = k) = \begin{cases} \frac{1}{b-a+1}, & k = a, \ldots, b \\ 0, & \text{o.w.} \end{cases} \tag{54}$$

We have

$$\mathbb{E}[X] = \frac{1}{b-a+1} \sum_{k=a}^{b} k = \frac{a+b}{2} \tag{55}$$

and

$$\text{var}(X) = \frac{(a-b)(a-b-2)}{12}. \tag{56}$$

### 1.3.4 Multinomial

Consider a sequence of $k$ independent and identically distributed random variables $(X_1, \ldots, X_k)$ with common probability distribution,

$$P(X_i = j) = p_j, \qquad j \in \{1, \ldots, n\}, \qquad \sum_{j=1}^{n} p_j = 1. \tag{57}$$

The multinomial distribution is thus a generalization of the binomial distribution where each random variable can take on $n$ values. We have,

$$P(X_1 = j_1, \ldots X_n = j_n) = \frac{k!}{j_1! \cdots j_n!} p_1^{j_1} \cdots p_n^{j_n}, \qquad \sum_{i=1}^{n} j_i = k. \tag{58}$$

8

The factor

$$\frac{k!}{j_1! \cdots j_n!} = \binom{k}{j_1}\binom{k-j_1}{j_2}\binom{k-j_1-j_2}{j_3} \cdots \binom{k-j_1-\cdots-j_{n-1}}{j_n} \tag{59}$$

is the number of ways to partition a $k$-element set into $n$ disjoint subsets, with $j_i$ elements in the $i$th subset.

### 1.3.5 Hypergeometric

The hypergeometric distribution describes the probability of picking $j$ red balls, after drawing $k$ times without replacement from an urn containing $n$ balls, of which $m$ are red,

$$P(X = j) = \frac{\binom{m}{j}\binom{N-m}{k-j}}{\binom{N}{k}}, \tag{60}$$

where $\binom{m}{j}$ is the number of ways of choosing $j$ balls from the $m$ red ones and $\binom{N-m}{k-j}$ is the number of ways of choosing $k-j$ balls from $N-m$ non-red ones. In the limit $N \to \infty$ holding $p \equiv m/N$ fixed the hypergeometric distribution approaches the binomial distribution; that is,

$$\lim_{N\to\infty} \frac{\binom{m}{j}\binom{N-m}{k-j}}{\binom{N}{k}} = \binom{k}{j} p^j (1-p)^{k-j}. \tag{61}$$

## 1.4 Continuous probability distributions

### 1.4.1 Normal

A normal random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ with mean $\mu$ and variance $\sigma^2$ has pdf

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)} \tag{62}$$

### 1.4.2 Poisson

The Poisson distribution can be thought of as the limit of the binomial distribution $\mathrm{B}(k, p)$ where $k \to \infty$ holding the expectation $\mathbb{E}[X] = kp = \lambda$ fixed,

$$P(X = j) = \lim_{k\to\infty} \binom{k}{j} \left(\frac{\lambda}{k}\right)^j \left(1 - \frac{\lambda}{k}\right)^{k-j} \tag{63}$$

$$= \lim_{k\to\infty} \frac{k(k-1)\cdots(k-j+1)}{k^j} \frac{\lambda^j}{j!} \left(1 - \frac{\lambda}{k}\right)^k \left(1 - \frac{\lambda}{k}\right)^{-j} \tag{64}$$

$$= \lim_{k\to\infty} \frac{(k^j + \cdots)}{k^j} \frac{\lambda^j}{j!} \left(1 - \frac{\lambda}{k}\right)^k \left(1 - \frac{\lambda}{k}\right)^{-j} \tag{65}$$

$$= \frac{\lambda^j e^{-\lambda}}{j!}. \tag{66}$$

The variance is also easily found as a limit of the binomial variance

$$\text{var}(X) = \lim_{k \to \infty} kp(1-p) = \lim_{k \to \infty} \lambda \left( 1 - \frac{\lambda}{k} \right) = \lambda. \tag{67}$$

If we relabel $\mathbb{E}[X] = \lambda t$ then $\lambda$ has the interpretation of the arrival rate and then the number of arrivals in an interval of time $\tau$ is given by

$$P(k, \tau) = \frac{(\lambda \tau)^k e^{-\lambda \tau}}{k!}. \tag{68}$$

Notice for comparison if we take the limit of $\text{B}(k, p)$ as $k \to \infty$ holding $p$ fixed we obtain a normal distribution by the central limit theorem.

## 1.5   Sums of random variables

If $X$ and $Y$ are independent random variables and $W = X + Y$ then

$$P(W = w) = P(X + Y = w) \tag{69}$$
$$= \sum_x P(X = x)P(Y = w - x). \tag{70}$$

For continuous distributions this formula becomes

$$f_W(w) = \int_{-\infty}^{\infty} dx \, f_X(x) f_Y(w - x). \tag{71}$$

The moment generating function for a random variable $X$ is given by

$$M_X(t) = \mathbb{E}[e^{tX}]. \tag{72}$$

If $X$ and $Y$ and independent random variables and $W = X + Y$ then

$$M_W(t) = M_X(t)M_Y(t). \tag{73}$$

The probability generating function is given by

$$g_X(s) = M_X(\log s) = \mathbb{E}[s^X]. \tag{74}$$

**Example 1.4.** Let $X$ and $Y$ be random variables that represent the rolls of two indepedent six-sided die with faces $\{1, 2, 3, 4, 5, 6\}$. We can obtain the distribution of $X+Y$ by computing its probability generating function

$$g_{X+Y}(s) = g_X(s)g_Y(s) \tag{75}$$
$$= g_X(s)^2 \tag{76}$$
$$= \frac{s^2}{36} + \frac{s^3}{18} + \frac{s^4}{12} + \frac{s^5}{9} + \frac{5s^6}{36} + \frac{s^7}{6} + \frac{5s^8}{36} + \frac{s^9}{9} + \frac{s^{10}}{12} + \frac{s^{11}}{18} + \frac{s^{12}}{36} \tag{77}$$

10

Factoring this expression we obtain

$$g_{X+Y}(s) = \frac{1}{36}s^2(1+s)^2\left(1-s+s^2\right)^2\left(1+s+s^2\right)^2 \tag{78}$$

$$= \frac{1}{4}(s^2+2s^3+s^4)\frac{1}{9}(1+2s^2+3s^4+2s^6+s^8) \tag{79}$$

$$= g_{X_4}(s)g_{X_9}(s) \tag{80}$$

$$= g_{X_4+X_9} \tag{81}$$

This shows that the distribution of $X + Y$ is the same as the distribution of $X_4 + X_9$ where $X_4$ and $X_9$ represent the rolls of two independent die with faces $\{2,3,3,4\}$ and $\{0,2,2,4,4,4,6,6,8\}$, respectively.

### 1.5.1 Generating arbitrarily distributed random variables

Starting with a random variable $Y$ uniformly distributed on $[0,1]$ we can generate a random variable $X$ with arbitrary probability distribution function $f(x)$ as follows. Let

$$F(x) = \int_{-\infty}^{x} du\, f(u) \tag{82}$$

be the cumulative distribution function for $X$. If $F(x)$ is continuous then the random variable $U \equiv F(X)$ is uniformly distributed on $[0,1]$ because

$$P(U \le u) = P(F(X) \le u) = P(X \le F^{-1}(u)) = F\left(F^{-1}(u)\right) = u. \tag{83}$$

The random variable $U$ is called the probability integral transform of $X$. We therefore define $X$ by the equation $F(X) = U$ or

$$X = F^{-1}(U). \tag{84}$$

In the case of the exponential distribution $f(x) = \lambda e^{-\lambda x}$ we have $F(x) = 1 - e^{-\lambda x}$ and thus $X = -\frac{1}{\lambda}\log(1-Y)$.

## 1.6 Correlation

Suppose we want to simulate $n$ random variables with a given covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$. This means we want a column vector of random variables $\vec{X} = (X_1, \ldots, X_n)^{\mathrm{T}}$ such that,

$$\Sigma = \mathbb{E}\left[(\vec{X} - \mathbb{E}[\vec{X}])(\vec{X} - \mathbb{E}[\vec{X}])^{\mathrm{T}}\right]. \tag{85}$$

Since the covariance matrix is positive semi-definite it can be diagonalized as $\Sigma = ADA^{\mathrm{T}}$ where $D$ is a diagonal matrix of eigenvalues and $A$ is an orthogonal matrix whose columns are the eigenvectors. If we take the columns of $A$ be be orthonormal then the eigenvalues in $D$ are the variances in the basis defined by the orthonormal column vectors; that is,

$$D = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2). \tag{86}$$

We now construct an $n$-component column vector of $n$ independent random variables $\vec{Z} = (Z_1, \ldots, Z_n)^{\mathrm{T}}$ with $\mathbb{E}[\vec{Z}] = (\sigma_1^2, \ldots, \sigma_n^2)$ and $\mathbb{E}[\vec{Z}] = \vec{0}$. Then $D$ is obviously the covariance matrix for $\vec{Z}$,

$$D = \mathbb{E}[\vec{Z}\vec{Z}^{\mathrm{T}}]. \tag{87}$$

Left-multiplying this expression by $A$ and right-multiplying by $A^{\mathrm{T}}$ we obtain the identity,

$$\Sigma = \mathbb{E}[A\vec{Z}(A\vec{Z})^{\mathrm{T}}]. \tag{88}$$

Thus we have found a vector or random variables $A\vec{Z}$ with vanishing mean and covariance matrix given by $\Sigma$. It follows that the vector of random variables satisfying the original requirement is given by,

$$\vec{X} = A\vec{Z} + \mathbb{E}[\vec{X}]. \tag{89}$$

In the special case where $\vec{Z}$ is a vector of independent Gaussian RVs, the resulting RVs $\vec{X}$ will also be Gaussian.

### 1.6.1 Copulas

Suppose that $(X_1, \ldots X_m)$ are continuous random variables with marginal distributions $F_1(x_1)$ up to $F_m(x_m)$ and joint distribution function $F(x_1, \ldots, x_m)$. Define the random variables $(U_1, \ldots, U_m)$ to be the associated probability integral transforms $U_1 = F_1(X_1), \ldots, U_m = F(x_m)$. The copula of $(X_1, \ldots X_m)$ is defined as the joint distribution for $(U_1, \ldots, U_m)$,

$$C(u_1, \ldots, u_m) = P(U_1 \leq u_1, \ldots, U_m \leq u_m). \tag{90}$$

Notice that $0 \leq u_i \leq 1$ so $C$ is a probability distribution on $[0, 1]^m$ and moreover the marginal distributions $P(U_i \leq u_i)$ are all uniform on $[0, 1]$. Using the definition of the probability integral transform this equation can also be written as,

$$C(u_1, \ldots, u_m) = P(F(X_1) \leq u_1, \ldots, F(X_m) \leq u_m), \tag{91}$$

$$= P(X_1 \leq F_1^{-1}(u_1), \ldots, X_m \leq F_m^{-1}(u_m)), \tag{92}$$

$$= F(F_1^{-1}(u_1), \ldots, F_m^{-1}(u_m)). \tag{93}$$

It immediately follows that,

$$F(x_1, \ldots, x_m) = C(F_1(x_1), \ldots, F_m(x_m)). \tag{94}$$

Taking draws $(X_1, \ldots, X_m)$ from $F$ is thus equivalent to taking draws $(U_1, \ldots, U_m)$ from $C$ and solving the $m$ sets of equations,

$$X_i = F_i^{-1}(U_i), \qquad 1 \leq i \leq m. \tag{95}$$

## 1.7 Law of large numbers

### 1.7.1 Weak law of large numbers

Let $X_1, \ldots, X_n$ be independent, identically distributed random variables with mean $\mu$ and variance $\sigma^2$. Define the sample mean to be the random variable

$$M_n = \frac{S_n}{n}, \qquad S_n = X_1 + \cdots + X_n \tag{96}$$

where $\mathbb{E}[M_n] = \mu$ and $\text{var}[M_n] = \sigma^2/n$. Applying the Markov inequality to $M_n$ we obtain

$$P(|M_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon}. \tag{97}$$

Taking $n \to \infty$ we find that for all $\epsilon$,

$$\lim_{n \to \infty} P(|M_n - \mu| \geq \epsilon) = 0. \tag{98}$$

This implies that $M_n$ converges in probability to the true mean $\mu$.

### 1.7.2 Central limit theorem

Now consider the standardized sum which measures the number of standard deviations $S_n$ lies from its mean,

$$Z_n \equiv \frac{S_n - \mathbb{E}[S_n]}{\sigma_{S_n}} = \frac{S_n - n\mathbb{E}[X]}{\sqrt{n}\sigma} = \frac{S_n/n - \mu}{\sigma/\sqrt{n}} \tag{99}$$

where $\mathbb{E}[Z_n] = 0$ and $\text{var}(Z_n) = 0$. In the limit as $n \to \infty$

$$P(Z_n \leq c) \to P(Z \leq c) \tag{100}$$

where

$$f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \tag{101}$$

## 1.8 Stochastic processes

### 1.8.1 Markov process

A discrete-time Markov chain is a stochastic process $\{X_n \in S \mid 0 \leq n < \infty\}$ with the property that for all $n \geq 1$

$$P(X_n = j \mid X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(X_n = s \mid X_{n-1} = i_{n-1}). \tag{102}$$

A Markov chain can alternatively be characterized by the following equivalent condition that for all $n \geq 1$ and $m \geq 0$,

$$P(X_{n+m} = j \mid X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(X_{n+m} = j \mid X_{n-1} = i_{n-1}). \tag{103}$$

Let us show this for $m = 1$ by defining the events $A = \{X_{n+m} = j\}$, $B = \{X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\}$ and $C_k = \{X_{n-1} = k\}$. Then,

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \tag{104}$$

$$= \sum_{k \in S} \frac{P(A \cap B \cap C_k)}{P(B)},$$

$$= \sum_{k \in S} \frac{P(A \cap B \cap C_k)}{P(B \cap C_k)} \frac{P(C_k \cap B)}{P(B)},$$

$$= \sum_{k \in S} P(A \mid B, C_k) P(C_k \mid B). \tag{105}$$

If we now apply the Markov property we obtain,

$$P(A \mid B) = \sum_{k \in S} P(X_{n+1} = j \mid X_n = k, X_{n-1} = i_{n-1}) P(X_{n-1} = k \mid X_{n-1} = i_{n-1}), \quad (106)$$

where we have left the event $\{X_{n-1} = i_{n-1}\}$ in the conditioning for convenience. Using the definition of conditional probability we find that $P(X_n = k, X_{n-1} = i_{n-1})$ cancels and we are left with

$$P(A \mid B) = \sum_{k \in S} \frac{P(X_{n+1} = j, X_n = k, X_{n-1} = i_{n-1})}{P(X_{n-1} = i_{n-1})}, \quad (107)$$

$$= \frac{P(X_{n+1} = j, X_{n-1} = i_{n-1})}{P(X_{n-1} = i_{n-1})}, \quad (108)$$

$$= P(X_{n+1} = j \mid X_{n-1} = i_{n-1}). \quad (109)$$

The $n$-step transition probabilities are defined by

$$p_{ij}(n) = P(X_n = j \mid X_0 = i), \quad (110)$$

and the associated matrix is denoted $\mathbf{P}(n)$ For time-homogeneous Markov chains the $n$-step transition probabilities satisfy,

$$p_{ij}(m+n) = \sum_{k \in S} p_{ik}(m) p_{kj}(n) \iff \mathbf{P}(m+n) = \mathbf{P}(m)\mathbf{P}(n) \quad (111)$$

It follows that $\mathbf{P}(n) = \mathbf{P}(1)^n$.

### 1.8.2 Branching process

Consider a branching process with offspring probabilities $p_0$, $p_1$ and $p_2$. Denote the probability of extinction by $p$. We condition the probability of extinction on each of the possible outcomes of the first branching event; namely,

$$P(\text{extinction}) = P(\text{extinction}|0)P(0) + P(\text{extinction}|1)P(1) + P(\text{extinction}|2)P(2) \quad (112)$$

Using independence we can express the extinction probability in terms of itself as

$$p = p_0 \cdot 1 + p_1 p + p_2 p^2. \quad (113)$$

Solving the quadratic equation and using the fact that $p_0 + p_1 + p_2 = 1$ we obtain $p = p_0/p_2$.

### 1.8.3 Random walk

Let $\{X_n \mid 1 \leq n \leq \infty\}$ be a sequence iid random variables with probability distribution $P(X_n = \pm 1) = 1/2$. Let $S_0$ be a constant and $S_n = S_0 + \sum_{k=1}^{n} X_n$ for $n \geq 1$. Denote by $\phi_k$ the probability that $S_n$ reaches $A$ before $-B$ given that $S_0 = k$. Then we obtain the recursion relation

$$\phi_k = \frac{1}{2}\phi_{k-1} + \frac{1}{2}\phi_{k+1}. \quad (114)$$

The boundary conditions for this finite difference equation are $\phi_A = 1$ and $\phi_{-B} = 0$. The solution for $\phi_0$ is

$$\phi_0 = \frac{B}{A + B}. \tag{115}$$

Notice that the recursion relation can be rearranged to give $\phi_{k-1} + \phi_{k+1} - 2\phi_k = 0$ which is just the discretization of the differential equation $\partial^2 \phi = 0$.

### 1.8.4 Martingales

Let $\{X_n \mid 1 \leq n \leq \infty\}$ be a sequence of random variables. A discrete-time Martingale is a sequence of random variables $M_n = f(X_1, \ldots, X_n)$ $(0 \leq n \leq \infty)$ with the property that

$$\mathbb{E}[M_n | X_1, \ldots, X_{n-1}] = M_{n-1} \tag{116}$$

or equivalently

$$\mathbb{E}[M_n - M_{n-1} | X_1, \ldots, X_{n-1}] = 0. \tag{117}$$

A Martingale in continuous time is defined as a stochastic process satisfying $\mathbb{E}_t[X(T)] = X(t)$ for $T > t$. Note that this implies that the expected increments are zero

$$\mathbb{E}_t[X(T) - X(t)] = 0. \tag{118}$$

Martingales have the property that their expectation value is constant; that is if $0 \leq s \leq t$ then $\mathbb{E}[X(t)] = \mathbb{E}[X(s)]$. This follows from the tower law of condition expectations,

$$\mathbb{E}[X(t)] = \mathbb{E}\big[\mathbb{E}[X(t) \mid \mathcal{F}_s]\big] = \mathbb{E}[X(s)]. \tag{119}$$

An important example of Martingale in continuous time is a Brownian motion. A Brownian motion is a continuous-time stochastic process which has continuous sample paths and satisfies

- $w(0) = 0$

- The increments have distribution $w(T) - w(t) \sim \mathcal{N}(0, T - t)$ for $T > t$

Notice that the second criterion implies that the distribution of increments $w(T) - w(t)$ (for $T > t$) depends only on the length $T - t$ of the time interval and is hence independent of the previous history. A Brownian motion is a Martingale because

$$\mathbb{E}_t[w(T)] = \mathbb{E}_t[w(T) - w(t) + w(t)] \tag{120}$$
$$= \mathbb{E}_t[w(T) - w(t)] + \mathbb{E}_t[w(t)] \tag{121}$$
$$= \mathbb{E}[w(T) - w(t)] + w(t) \tag{122}$$
$$= w(t). \tag{123}$$

Moreover, since $\mathbb{E}[dw(t)] = 0$, we have $\mathbb{E}[dw(t)^2] = \text{var}[dw(t)] = dt$.

The following Mathematica code generates and plots sample path for a Brownian motion

```
T = 10;
n = 1000;
dt = T/n;
dw = RandomVariate[NormalDistribution[0, Sqrt[dt]], n];
BrownianPath = Accumulate[Prepend[dw, 0]];
ListLinePlot[BrownianPath, Frame -> True]
```
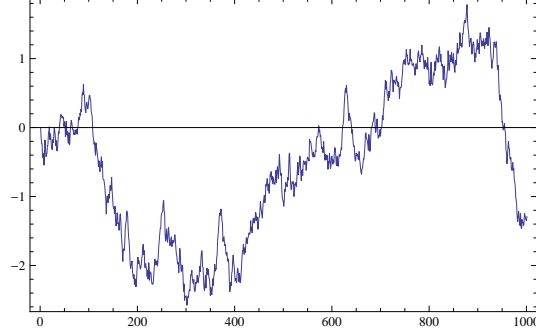


$$(124)$$

Another important example of a Martingale is the geometric Brownian motion which is a stochastic process given by $e^{-\sigma^2/2t+\sigma w(t)}$.

### 1.8.5    Ito calculus

A stochastic process $X$ is an Ito process if

$$dX(t) = \mu(t)dt + \sigma(t)dw(t) \tag{125}$$

where $\mu(t)$ and $\sigma(t)$ are stochastic processes. An Ito process is called a diffusion if $\mu(X(t), t)$ and $\sigma(X(t), t)$. In the special case where $\mu$ and $\sigma$ are constants this equation can be integrated to

$$X(t) = X(0) + \mu t + \sigma w(t). \tag{126}$$

Thus

$$X(T) - X(t) \sim \mathcal{N}\left(X(0) + \mu(T-t), \sigma^2(T-t)\right). \tag{127}$$

Ito's lemma states that if $X(t)$ is an Ito process with drift $\mu(t)$ and volatility $\sigma(t)$ and $Y(t) = f(X(t), t)$ then

$$dY(t) = f_X(X(t), t)dX(t) + f_t(X(t), t)dt + \frac{1}{2}f_{XX}(X(t), t)\sigma(t)^2 dt \tag{128}$$

$$= \left[ f_X(X(t), t)\mu(t) + f_t(X(t), t) + \frac{1}{2}f_{XX}(X(t), t)\sigma(t)^2 \right] dt + f_X(X(t), t)\sigma(t)dw(t).$$

The intuition for the second derivative term follows from considering the squared differential

$$dX(t)^2 = \mu(t)^2 dt^2 + 2\mu(t)\sigma(t)dw(t)dt + \sigma(t)^2 dw(t)^2. \tag{129}$$

We then argue that the first and second terms are negligible compared to the third because of the identity $\mathbb{E}[dw(t)^2] = dt$.

16

Ito's lemma can be used to solve stochastic differential equations by changing variables. Let us rescale the drift and volatility,

$$dX(t) = X(t)\mu(t)dt + X(t)\sigma(t)dw(t) \tag{130}$$

where $\mu(t)$ and $\sigma(t)$ now denote the relative drift and volatility, respectively. Consider the change of variables $Y(t) = f(X(t)) = \log X(t)$. Then $\mu(t) \to X(t)\mu(t)$, $\sigma(t) \to X(t)\sigma(t)$, $f_X = 1/X$ and $f_{XX} = -1/X^2$ so by Ito's lemma

$$dY(t) = \left[\mu(t) - \frac{1}{2}\sigma(t)^2\right] dt + \sigma(t)dw(t) \tag{131}$$

so

$$\log X(t) = \log X(0) + \int_0^t \left[\mu(s) - \frac{1}{2}\sigma(s)^2\right] ds + \int_0^t \sigma(s)dw(s) \tag{132}$$

Now assume that $\mu(t)$ and $\sigma(t)$ are time-dependent but not stochastic. Computing the expectation value and variance we find

$$\mathbb{E}[\log X(t)] = \log X(0) + \int_0^t \left[\mu(s) - \frac{1}{2}\sigma(s)^2\right] ds \tag{133}$$

$$\text{var}[\log X(t)] = \text{var}\left[\int_0^t \sigma(s)dw(s)\right] \tag{134}$$

$$= \int_0^t \sigma(s)^2 \, \text{var}[dw(s)] \tag{135}$$

$$= \int_0^t \sigma(s)^2 ds \tag{136}$$

and thus

$$\log X(t) \sim \mathcal{N}\left(\log X(0) + \int_0^t \left[\mu(s) - \frac{1}{2}\sigma(s)^2\right] ds, \int_0^t \sigma(s)^2 ds\right). \tag{137}$$

Taking $\mu$ and $\sigma$ to be constants (geometric Brownian motion) we obtain

$$X(t) = X(0)e^{\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma w(t)} \tag{138}$$

Hence

$$\log X(t) \sim \mathcal{N}\left(\log X(0) + \left(\mu - \frac{1}{2}\sigma^2\right) t, \sigma^2 t\right) \tag{139}$$

The following Mathematica code generates and plots a sample path for a geometric Brownian motion

```
T=10;n=1000;dt=T/n;mu=0.3;sigma=0.4;Y0=2;
dw=RandomVariate[NormalDistribution[0,Sqrt[dt]],n];
dY=(mu-sigma^2/2)dt+sigma dw;
YPath=Accumulate[Prepend[dY,Y0]];
XPath=Exp[YPath];
ListLinePlot[XPath,Frame->True]
```

$$(140)$$
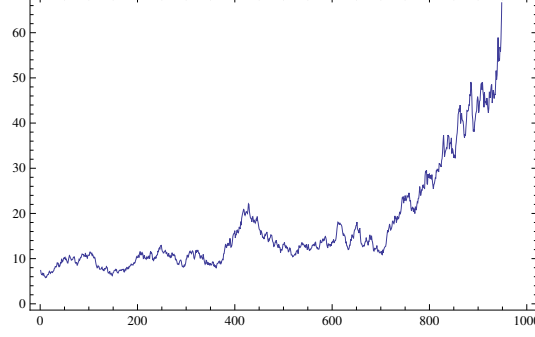
Suppose that $X$ is an Ito process under some probability measure $P$,

$$dX(t) = X(t)\mu(t)dt + X(t)\sigma(t)dw(t) \tag{141}$$

where $w$ is a Brownian motion under $P$. Given another probability measure $Q$, Girsanov's theorem states that $X$ is an Ito process under $Q$ with the same volatility,

$$dX(t) = X(t)\hat{\mu}(t)dt + X(t)\sigma(t)d\hat{w}(t) \tag{142}$$

where $\hat{w}$ is a Brownian motion under $Q$.

**Example 1.5.** (Ornstein-Uhlenbeck process) Consider the SDE

$$dr(t) = \kappa(\theta(t) - r(t))dt + \sigma dw(t). \tag{143}$$

We can rewrite this as

$$dr(t) + \kappa r(t)dt = \kappa\theta(t)dt + \sigma dw(t) \tag{144}$$
$$e^{-\kappa t}d(e^{\kappa t}r(t)) = \kappa\theta(t)dt + \sigma dw(t) \tag{145}$$
$$d(e^{\kappa t}r(t)) = \kappa e^{\kappa t}\theta(t)dt + \sigma e^{\kappa t}dw(t) \tag{146}$$

Integrating both sides gives

$$r(t) = e^{-\kappa t}r(0) + \kappa \int_0^t e^{-\kappa(t-s)}\theta(s)ds + \sigma \int_0^t e^{-\kappa(t-s)}dw(s) \tag{147}$$

It follows that

$$r(t) \sim \mathcal{N}\left(e^{-\kappa t}r(0) + \kappa \int_0^t e^{-\kappa(t-s)}\theta(s)ds, \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa t})\right) \tag{148}$$

If $\theta(s) = \theta$ then

$$r(t) = e^{-\kappa t}r(0) + \theta(1 - e^{-\kappa t}) + \sigma \int_0^t e^{-\kappa(t-s)}dw(s) \tag{149}$$

and the process is mean reverting to $\theta$.

**Example 1.6.** Consider the stochastic process

$$X(t) = \int_0^t ds\, w(s). \tag{150}$$

18

This process is an Ito process with Brownian drift term $dX(t) = w(t)dt$. $X(t)$ is a normal random variable. Let us calculate its mean and variance. If we integrate by parts we find

$$X(t) = t\,w(t) - \int_0^t s\,dw(s) = \int_0^t (t-s)dw(s) \tag{151}$$

where we have used $w(0) = 0$. Thus $\mathbb{E}[X(t)] = 0$ and $\mathrm{var}[X(t)] = \int_0^t (t-s)^2 ds = t^3/3$. We can also compute the variance of $X(t)$ using the fact that it has vanishing expectation value,

$$\mathrm{var}[X(t)] = \mathbb{E}\left[X(t)^2\right] \tag{152}$$

$$= \int_0^t ds_1 \int_0^t ds_2\,\mathbb{E}[X(s_1)X(s_2)] \tag{153}$$

$$= \int_0^t ds_1 \int_0^t ds_2\,\min(s_1, s_2) \tag{154}$$

$$= t^3/3. \tag{155}$$

Here we have used the fact that $X(t)$ has independent stationary increments so

$$\mathbb{E}[X(s_1)X(s_2)] = \begin{cases} \mathbb{E}\left[X(s_1)^2\right] = s_1, & s_2 > s_1 \\ \mathbb{E}\left[X(s_2)^2\right] = s_2, & s_1 > s_2. \end{cases} \tag{156}$$

### 1.8.6   Semimartingales

A Poisson process $N(t)$ with intensity $\lambda$ is a continuous-time stochastic process such that

- $N(0) = 0$

- $dN(t)$ is a Bernoulli random variable for all $t$

- $\mathbb{E}_t[N(T) - N(t)] = \lambda(T - t)$ for all $T \geq t$.

A compound Poisson (jump) process with intensity $\lambda$ and jump size distribution $\mathcal{D}$ is given by

$$J(t) = \sum_{i=1}^{N(t)} J_i \tag{157}$$

where $J_i$ are independent, identically distributed random variables under $\mathcal{D}$ which are independent of $N$.

A semimartingale is the sum of an Ito process and a jump process

$$dX(t) = \mu(t)dt + \sigma(t)dw(t) + dJ(t). \tag{158}$$

The following Mathematica code generates a sample path for a the process $dX(t) = dw(t) + dJ(t)$ where $dJ(t)$ is a jump process with intensity $\lambda = 1$ and $\mathcal{D} = \mathcal{N}(0, 3)$.

```
T = 10; n = 2000; dt = T/n; lambda = 1; sigma = 2;
dw = RandomVariate[NormalDistribution[0, Sqrt[dt]], n];
dJ = Table[
   If[RandomVariate[BernoulliDistribution[lambda dt]] == 0, 0,
    RandomVariate[NormalDistribution[0, sigma]]], {i, 1, n}];
dX = dw + dJ;
BrownianJumpPath = Accumulate[Prepend[dX, 0]];
ListLinePlot[BrownianJumpPath, Frame -> True]
```
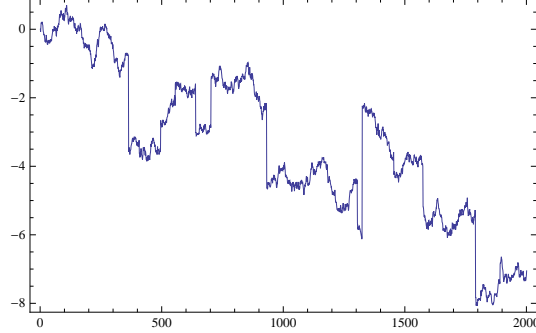


$$(159)$$

Ito's lemma can be generalized to semimartingales. If $X(t)$ is a semimartingale and $Y(t) = f(X(t), t)$ then

$$dY(t) = \left[ f_X(X(t-), t)\mu(t) + f_t(X(t-), t) + \frac{1}{2} f_{XX}(X(t-), t)\sigma(t)^2 \right] dt +$$
$$+ f_X(X(t-), t)\sigma(t)dw(t) + f(X(t), t) - f(X(t-), t) \tag{160}$$

## 1.9 Feynman-Kac formula

Consider the following parabolic PDE ($x \in \mathbb{R}$ and $t < T$)

$$\frac{\partial u}{\partial t} + \frac{1}{2}\sigma(x, t)^2 \frac{\partial^2 u}{\partial x^2} + \mu(x, t)\frac{\partial u}{\partial x} - V(x, t)u(x, t) = 0, \qquad u(x, T) = f(x). \tag{161}$$

Now consider the diffusion process

$$dX(t) = \mu(X(t), t)dt + \sigma(X(t), t)dw(t) \tag{162}$$

where $w$ is a Brownian motion under $Q$. The Feynman-Kac theorem relates the solution of the PDE to a conditional expectation value

$$u(x, t) = \mathbb{E}^Q \left( e^{-\int_t^T ds\, V(X(s), s)} f(X(T)) \,\middle|\, X(t) = x \right). \tag{163}$$

The Feynman-Kac formula has a simple multi-dimensional generalization to PDEs dependent on $p$-dimensional spatial variables,

$$\frac{\partial u}{\partial t} + \sum_{i=1}^{p} \mu_i(x, t)\frac{\partial u}{\partial x_i} + \frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{p} s_{ij}(x, t)\frac{\partial^2 u}{\partial x_i \partial x_j} - V(x, t)u(x, t) = 0, \tag{164}$$

where $s_{ii}(x,t) \geq 0$ and $s_{ij}(x,t) = s_{ji}(x,t)$. The solution to this equation for $t < T$ subject to the terminal condition $u(x,T) = f(x)$ is given by the expectation

$$u(x,t) = \mathbb{E}^Q \left( e^{-\int_t^T ds\, V(X(s),s)} f(X(T)) \middle| X(t) = x \right) \tag{165}$$

where $X_i$ are Ito processes given by

$$dX_i(t) = \mu_i(X(t),t)dt + \vec{\sigma}_i(X(t),t) \cdot d\vec{w}(t) \tag{166}$$

where $\vec{w}$ is a $d$-dimensional Brownian motion and $s_{ij} = \vec{\sigma}_i \cdot \vec{\sigma}_j$.

- In the special case where $\sigma(x,t) = \sigma$ and $\mu(x,t) = 0$ the PDE reduces to the backward heat equation

- The expectation value can be interpreted as an integration with respect to the Wiener measure

- The probabilistic interpretation of absence of mixed derivative terms is that the stochastic increments are independent

# 2 Optimal control

## 2.1 Liquidation problem

Consider the problem of selling $X$ shares of an asset during a time interval $[0,T]$. We assume that $n_j$ shares are sold during the interval $(t_{j-1}, t_j]$ where $t_j = j\tau$, $\tau = T/N$ and $j \in \{1, \ldots, N\}$ such that $X = \sum_{j=1}^{N} n_j$. The price received $\hat{S}_j$ from the sale of $n_j$ shares during the interval $(t_{j-1}, t_j]$ is assumed to be displaced from the prevailing market price $S_{j-1}$ by a temporary price impact which we model as function dependent only on $n_j$,

$$\hat{S}_j = S_{j-1} - h(n_j). \tag{167}$$

The price process is modeled as a discrete-time random process which suffers a permanent impact $-g(n_j)$ dependent on the number of trades executed during that interval,

$$S_j = S_{j-1} + \sigma \tau^{1/2} z_j - g(n_j), \tag{168}$$

where $(z_1, \ldots, z_N)$ is a vector of independent and identically distributed $\mathcal{N}(0,1)$ random variables. The recursive formula for the price process can be used to express it in terms of today's spot price $S_0$ as follows,

$$S_k = S_0 + \sum_{j=1}^{k} \left[ \sigma \tau^{1/2} z_j - g(n_j) \right]. \tag{169}$$

Hence the total revenue generated from the trading strategy is given by,

$$R = \sum_{k=1}^{N} \hat{S}_k n_k, \tag{170}$$

$$= S_0 \sum_{k=1}^{N} n_k + \sigma\tau^{1/2} \sum_{k=1}^{N}\sum_{j=1}^{k-1} z_j n_k - \sum_{k=1}^{N}\sum_{j=1}^{k-1} g(n_j)n_k - \sum_{k=1}^{N} h(n_k)n_k, \tag{171}$$

$$= S_0 X + \sigma\tau^{1/2} \sum_{k=1}^{N} z_k x_k - \sum_{k=1}^{N} g(n_k)x_k - \sum_{k=1}^{N} h(n_k)n_k, \tag{172}$$

where we have defined the inventory process $x_k$ which satisfies $x_0 = X$ and $x_N = 0$,

$$x_k = X - \sum_{j=1}^{k} n_k. \tag{173}$$

The expected value and variance of the total cost of trading $C = S_0 X - R$ is given by

$$\mathbb{E}[C] = \sum_{k=1}^{N} \left[ g(n_k)x_k + h(n_k)n_k \right], \tag{174}$$

$$= \sum_{k=1}^{N} \left[ g(x_{k-1} - x_k)x_k + h(x_{k-1} - x_k)(x_{k-1} - x_k) \right], \tag{175}$$

$$\mathrm{var}(C) = \sigma^2\tau \sum_{k=1}^{N} x_k^2. \tag{176}$$

Assuming linear price impact functions $g(x) = (c/\tau)x$ and $h(x) = (\eta/\tau)x$ we obtain,

$$C(x) = \frac{cX^2}{2} + \sum_{k=1}^{N} \tau \frac{1}{2}m \left( \frac{x_{k-1} - x_k}{\tau} \right)^2 - \sum_{k=1}^{N} \sigma\tau^{1/2} z_k x_k, \tag{177}$$

where $m = 2\eta - c$. Let us define a mean-variance objective function as follows,

$$S[x_1, \ldots, x_{N-1}] = \mathbb{E}[C] + \lambda\,\mathrm{var}(C) - \frac{cX^2}{2}, \tag{178}$$

$$= \sum_{k=1}^{N} \tau \left[ \frac{1}{2}m \left( \frac{x_{k-1} - x_k}{\tau} \right)^2 + \frac{1}{2}\omega^2 x_k^2 \right], \tag{179}$$

where we have defined $\omega^2 = 2\lambda\sigma^2$. Optimizing the objective function by setting $\partial S/\partial x_i = 0$ we obtain,

$$\frac{x_{i+1} - 2x_i + x_{i-1}}{\tau^2} = \omega^2 x_i. \tag{180}$$

In the limit $\tau \to 0$ the objective function takes the form of an action functional for a particle of mass $m$ in an inverted harmonic potential $V(x) = -\frac{1}{2}\omega^2 x^2$,

$$S[x] = \int_0^T dt \left[ \frac{1}{2}m\dot{x}^2 + \frac{1}{2}\omega^2 x^2 \right]. \tag{181}$$

The solution of the corresponding Euler-Lagrange equation $m\ddot{x} = \omega^2 x$ subject to the boundary conditions $x(0) = X$ and $x(T) = 0$ is thus,

$$x(t) = X \frac{\sinh\left[\frac{\omega}{\sqrt{m}}(T - t)\right]}{\sinh\left[\frac{\omega}{\sqrt{m}}T\right]}. \tag{182}$$

Setting $t = k\tau$ we obtain the approximate solution of the finite-difference equation found by Almgren and Chriss [1].

### 2.1.1 Dynamic strategies

In this section we review [2] which considers the liquidation problem with temporary and permanent price impact functions of the form,

$$\hat{S}_j = S_{j-1} - \eta_j n_j, \tag{183}$$
$$S_j = S_{j-1} + \tau\sigma_j z_j - cn_j \tag{184}$$

where $\{\eta_j\}_{j=0}^N$ and $\{\sigma_j\}_{j\in\mathbb{N}}$ are Markov chains with transition probabilities,

$$p_{k-1}^{(v,w)} = \mathbb{Q}[(\sigma_k, \eta_k) = w \mid (\sigma_{k-1}, \eta_{k-1}) = v], \quad k \in \{1, \ldots, N\}. \tag{185}$$

The total cost of trading is now given by

$$C(x) = \frac{cX^2}{2} + \sum_{k=1}^N \left[\left(\eta_k - \frac{c}{2}\right)(x_{k-1} - x_k)^2 - \sigma_k \tau^{1/2} z_k x_k\right]. \tag{186}$$

In contrast to Almgren and Chriss, the risk-aversion parameter is taken to be zero so the optimization problem is,

$$x^* = \underset{x\in\mathcal{A}}{\operatorname{argmin}} \mathbb{E}[Q(x) \mid (\sigma_0, \eta_0) = v], \tag{187}$$

where for convenience we wave defined $Q(x) = C(x) - cX^2/2$ and $\mathcal{A}$ denotes the set of admissible strategies $(x_0, \ldots, x_N)$ with the boundary conditions $x_0 = X$ and $x_N = 0$. Let us define

$$J_n^{(v)}(z) = \min_{x\in\mathcal{A}_n(z)} \mathbb{E}[Q_n(x) \mid (\sigma_n, \eta_n) = v], \tag{188}$$

where $\mathcal{A}_n(z)$ denotes the set of admissible substrategies $(x_n, \ldots, x_N)$ satisfying the boundary conditions $x_n = z$ and $x_N = 0$, and we have defined

$$Q_n(x) = \sum_{i=n+1}^N (x_{i-1} - x_i)^2 \tilde{\eta}_i, \qquad \tilde{\eta}_i = \eta_i - \frac{c}{2}. \tag{189}$$

The inductive hypothesis is,

$$J_n^{(v)}(x_n) = x_n^2 a_n^{(v)}, \quad n \in \{0, \ldots, N - 1\}. \tag{190}$$

Let us check this in the case $n = N - 1$. Using the fact that $x_N = 0$ we obtain $Q_{N-1}(x) = (x_{N-1}^2)\tilde{\eta}_N$ and thus

$$J_{N-1}^{(v)}(x_{N-1}) = \min_{\mathcal{A}_{N-1}(x_{N-1})} (x_{N-1})^2 \mathbb{E}[\tilde{\eta}_N \mid (\sigma_{N-1}, \eta_{N-1}) = v], \tag{191}$$

$$= (x_{N-1})^2 \mathbb{E}[\tilde{\eta}_N \mid (\sigma_{N-1}, \eta_{N-1}) = v]. \tag{192}$$

We therefore obtain,

$$a_{N-1}^{(v)} = \mathbb{E}[\tilde{\eta}_N \mid (\sigma_{N-1}, \eta_{N-1}) = v]. \tag{193}$$

Now,

$$Q_{n-1}(x) = Q_n + (x_{n-1} - x_n)^2 \tilde{\eta}_n \tag{194}$$

so,

$$J_{n-1}^{(v)}(x_{n-1}) = \min_{\mathcal{A}_{n-1}(x_{n-1})} \mathbb{E}[Q_{n-1} \mid (\sigma_{n-1}, \eta_{n-1}) = v],$$

$$= \min_{\mathcal{A}_{n-1}(x_{n-1})} \left\{ \mathbb{E}[Q_n \mid (\sigma_{n-1}, \eta_{n-1}) = v] + (x_{n-1} - x_n)^2 \mathbb{E}[\tilde{\eta}_n \mid (\sigma_{n-1}, \eta_{n-1}) = v] \right\},$$

$$= \min_{x_n} \left\{ \min_{\mathcal{A}_n(x)} \mathbb{E}[Q_n \mid (\sigma_{n-1}, \eta_{n-1}) = v] + (x_{n-1} - x_n)^2 \mathbb{E}[\tilde{\eta}_n \mid (\sigma_{n-1}, \eta_{n-1}) = v] \right\}. \tag{195}$$

Using the tower rule for conditional expectations we obtain,

$$\min_{\mathcal{A}_n(x)} \mathbb{E}[Q_n \mid (\sigma_{n-1}, \eta_{n-1}) = v] = \min_{\mathcal{A}_n(x_n)} \mathbb{E}\big[\mathbb{E}[Q_n \mid (\sigma_n, \eta_n)] \mid (\sigma_{n-1}, \eta_{n-1}) = v\big],$$

$$= \min_{\mathcal{A}_n(x_n)} \sum_w \mathbb{E}[Q_n \mid (\sigma_n, \eta_n) = w] \mathbb{Q}[(\sigma_n, \eta_n) = w \mid (\sigma_{n-1}, \eta_{n-1}) = v],$$

$$= \sum_w \Big[ \min_{\mathcal{A}_n(x_n)} \mathbb{E}[Q_n \mid (\sigma_n, \eta_n) = w] \Big] \mathbb{Q}[(\sigma_n, \eta_n) = w \mid (\sigma_{n-1}, \eta_{n-1}) = v],$$

$$= x_n^2 \sum_w p_{n-1}^{(v,w)} a_n^{(w)}, \tag{196}$$

where we have used the inductive hypothesis $\min_{\mathcal{A}_n(x_n)} \mathbb{E}[Q_n \mid (\sigma_n, \eta_n) = w] = x_n^2 a_n^{(w)}$. Substituting back we obtain,

$$J_{n-1}^{(v)}(x_{n-1}) = \min_{x_n} \left\{ x_n^2 \sum_w p_{n-1}^{(v,w)} a_n^{(w)} + (x_{n-1} - x_n)^2 \mathbb{E}[\tilde{\eta}_n \mid (\sigma_{n-1}, \eta_{n-1}) = v] \right\}. \tag{197}$$

So $x^*$ satisfies the recursion,

$$x_n^* = x_{n-1}^* \frac{\mathbb{E}[\tilde{\eta}_n \mid (\sigma_{n-1}, \eta_{n-1}) = v]}{\mathbb{E}[\tilde{\eta}_n \mid (\sigma_{n-1}, \eta_{n-1}) = v] + \sum_w p_{n-1}^{(v,w)} a_n^{(w)}}. \tag{198}$$

Substituting back into the value function we see that the inductive hypothesis is satisfied by the optimal solution.

## 2.2 Infinite horizon

In this section we consider optimal trading of a single asset with an infinite time horion [4]. Let $S_n$ denote the price of an asset at time $t_n$ and let $r$ denote the risk-free rate. Suppose the excess returns $r_n = S_n - (1+r)S_{n-1}$ from time $t_{n-1}$ to time $t_n$ are described in terms of a $K$-dimensional vector of stochastic factors $\vec{f}_n$ by the following linear hypothesis,

$$r_n = \vec{\theta} \cdot \vec{f}_n + u_n, \tag{199}$$

where $u_n$ is a random variable with mean zero and variance $\sigma^2$. The factors are assumed to evolve according to a mean-reverting process,

$$\vec{f}_n = \vec{f}_{n-1} + \kappa(\vec{m}_n - \vec{f}_{n-1}) + \vec{\xi}_n, \tag{200}$$

where $\vec{\xi}_n$ is sequence of iid mean zero random vectors with covariance matrix $\Omega$, $\vec{m}_n$ is a vector-valued stochastic process representing the mean-reversion values and $\kappa \in \mathbb{R}^{K \times K}$ is a matrix which describes the speed of mean reversion of the stochastic factor vector $\vec{f}_n$. Let $x_n$ denote the amount of holdings in the asset at time $t_n$. The cost of trading $x_n - x_{n-1}$ shares over the interval $(t_{n-1}, t_n]$ is given by,

$$C_n = (x_n - x_{n-1})\hat{S}_n, \tag{201}$$

where $\hat{S}_n$ is the transacted cost which differs from the spot price $S_{n-1}$ by a term proportional to the number of trades executed $x_n - x_{n-1}$, multiplied by a stochastic liquidity $\eta_n$,

$$\hat{S}_n = S_{n-1} + \eta_n(x_n - x_{n-1}). \tag{202}$$

The stochastic process $(\vec{m}_n, \eta_n)$ is taken to be a time-homogeneous Markov chain with transition probabilities,

$$p^{(v,w)} = \mathbb{Q}[(\vec{m}_k, \eta_k) = w \mid (\vec{m}_{k-1}, \eta_{k-1}) = v]. \tag{203}$$

The optimal trading strategy is such that it maximizes the expected returns minus transaction costs. Following Markowitz, it is instructive to penalize the risk of the position by introducing a risk-aversion coefficient $\lambda$. The required strategy as a function of the initial state $v$ is thus,

$$x^* = \underset{x \in \mathcal{A}}{\operatorname{argmax}} \, \mathbb{E}\big[Q(x) \mid (\vec{m}_0, \eta_0) = v\big], \tag{204}$$

where

$$Q(x) = \sum_{k \geq 1} r_k x_k - \eta_k(x_k - x_{k-1})^2 - \frac{1}{2}x_k^2 \gamma \sigma^2. \tag{205}$$

and $\mathcal{A}$ denotes the set of strategies with $x_0 = X$ and $f_0 = f$. Let us define,

$$J_n^{(v)}(z, f) = \max_{x \in \mathcal{A}_n(z,f)} \mathbb{E}\big[Q_n(x) \mid (\vec{m}_n, \eta_n) = v\big], \qquad n \in \{0, 1, \ldots, \infty\}, \tag{206}$$

where

$$Q_n(x) = \sum_{k \geq n+1} r_k x_k - \eta_k(x_k - x_{k-1})^2 - \frac{1}{2}x_k^2 \gamma \sigma^2. \tag{207}$$

We are interested in the case $n = 0$ but since the problem is time-homogeneous we choose a time-homogeneous ansatz. Following [3], we choose the following quadratic ansatz,

$$J^{(v)}(z, \vec{f}) = -\frac{1}{2}a_v z^2 + b_v z + \vec{f} \cdot A_v \vec{f} + \vec{c}_v \cdot \vec{f} + \vec{d}_v \cdot \vec{f} z + e_v. \tag{208}$$

Proceeding inductively we find,

$$J_{n-1}^{(v)}(x_{n-1}, \vec{f}_{n-1}) = \max_{x \in \mathcal{A}_{n-1}(x_{n-1}, f_{n-1})} \mathbb{E}\left[Q_{n-1}(x) \mid (\vec{m}_{n-1}, \eta_{n-1}) = v\right], \tag{209}$$

$$= \max_{x \in \mathcal{A}_{n-1}(x_{n-1}, f_{n-1})} \left\{ \mathbb{E}\left[Q_n(x) \mid (\vec{m}_{n-1}, \eta_{n-1}) = v\right] + \right.$$

$$\left. + \mathbb{E}\left[r_n \mid (\vec{m}_{n-1}, \eta_{n-1}) = v\right] x_n - \frac{1}{2}x_n^2 \gamma \sigma^2 - (x_n - x_{n-1})^2 \mathbb{E}\left[\eta_n \mid (\vec{m}_{n-1}, \eta_{n-1}) = v\right] \right\}.$$

Using the tower rule for conditional expectations we obtain,

$$\mathbb{E}\left[Q_n(x) \mid (\vec{m}_{n-1}, \eta_{n-1}) = v\right] = \mathbb{E}\left[Q_n(x) \mid (\vec{m}_{n-1}, \eta_{n-1}) = v\right], \tag{210}$$

$$= \mathbb{E}\left[\mathbb{E}[Q_n \mid (\vec{m}_n, \eta_n)] \mid (\vec{m}_{n-1}, \eta_{n-1}) = v\right], \tag{211}$$

$$= \sum_w p^{(v,w)} \mathbb{E}[Q_n \mid (\vec{m}_n, \eta_n) = w]. \tag{212}$$

Now recall that in our optimization problem $\vec{f}_{n-1}$ is known. On the other hand, $\vec{f}_n$ is uncertain because it depends on the random variables $\vec{\xi}_n$. Therefore we condition on the value of $\vec{\xi}_n$ as follows,

$$\mathbb{E}\left[Q_n(x) \mid (\vec{m}_{n-1}, \eta_{n-1}) = w\right] = \sum_w p^{(v,w)} \int d\vec{\xi} \rho(\vec{\xi}) \mathbb{E}\left[Q_n \mid (\vec{m}_n, \eta_n) = w, \vec{\xi}_n = \vec{\xi}\right]. \tag{213}$$

Since $\vec{\xi}_n$ is known in the above expectation, and using the fact that

$$\mathbb{E}[u_n \mid (\vec{m}_{n-1}, \eta_{n-1}) = v] = 0, \tag{214}$$

$$\mathbb{E}[\xi_n \mid (\vec{m}_{n-1}, \eta_{n-1}) = v] = 0, \tag{215}$$

we obtain,

$$J_{n-1}^{(v)}(x_{n-1}, f_{n-1}) = \max_{x_n} \left\{ \sum_w p^{(v,w)} \int d\vec{\xi} \rho(\vec{\xi}) \left. J^{(w)}(x_n, \vec{f}_n)\right|_{\vec{f}_n = \vec{f}_{n-1} + \kappa(\vec{m}_{(w)} - \vec{f}_{n-1}) + \vec{\xi}} + \right. \tag{216}$$

$$\left. + \vec{\theta} \cdot \left[(1 - \kappa)\vec{f}_{n-1} + \sum_w p^{(v,w)} \kappa \vec{m}_{(w)}\right] x_n - \frac{1}{2}x_n^2 \gamma \sigma^2 - (x_n - x_{n-1})^2 \sum_w p^{(v,w)} \eta_{(w)} \right\}.$$

Integrating over the random vector $\vec{\xi}$ we obtain,

$$\int d\vec{\xi} \rho(\vec{\xi}) \left. J^{(w)}(x_n, \vec{f}_n)\right|_{\vec{f}_n = \vec{f}_{n-1} + \kappa(\vec{m}_{(w)} - \vec{f}_{n-1}) + \vec{\xi}} = -\frac{1}{2}a_w x_n^2 + b_w x_n + \vec{d}_w \cdot \left[\vec{f}_{n-1} + \kappa(\vec{m}_{(w)} - \vec{f}_{n-1})\right] x_n$$

$$+ \left[\vec{f}_{n-1} + \kappa(\vec{m}_{(w)} - \vec{f}_{n-1})\right] \cdot A_w \left[\vec{f}_{n-1} + \kappa(\vec{m}_{(w)} - \vec{f}_{n-1})\right] + \text{tr}(A_w \Omega) + \vec{c}_w \left[\vec{f}_{n-1} + \kappa(\vec{m}_{(w)} - \vec{f}_{n-1})\right] + e_w.$$

Substituting back and collecting terms in $x_n$ gives,

$$J_{n-1}^{(v)}(x_{n-1}, f_{n-1}) = \max_{x_n} \sum_w p^{(v,w)} \left\{ -\frac{1}{2}(\gamma\sigma^2 + a_w + 2\eta_{(w)})x_n^2 - x_{n-1}^2\eta_{(w)} + \right.$$

$$+ x_n\left(\vec{\theta}\cdot\left[(1-\kappa)\vec{f}_{n-1} + \kappa\vec{m}_{(w)}\right]x_n + 2x_{n-1}\eta_{(w)} + b_w + \vec{d}_w\cdot\left[(1-\kappa)\vec{f}_{n-1} + \kappa\vec{m}_{(w)}\right]\right) +$$

$$+ \left[(1-\kappa)\vec{f}_{n-1} + \kappa\vec{m}_{(w)}\right]\cdot A_w\left[(1-\kappa)\vec{f}_{n-1} + \kappa\vec{m}_{(w)}\right] + \text{tr}(A_w\Omega) +$$

$$\left. + \vec{c}_w\left[(1-\kappa)\vec{f}_{n-1} + \kappa\vec{m}_{(w)}\right] + e_w \right\}, \tag{217}$$

$$= \max_{x_n} \sum_w p^{(v,w)} \left\{ -\frac{1}{2}(\gamma\sigma^2 + a_w + 2\eta_{(w)})x_n^2 - x_{n-1}^2\eta_{(w)} + \right.$$

$$+ x_n\left(\left[(1-\kappa)^{\text{T}}(\vec{\theta} + \vec{d}_w)\right]\cdot\vec{f}_{n-1} + 2x_{n-1}\eta_{(w)} + b_w + (\vec{\theta} + \vec{d}_w)\cdot\kappa\vec{m}_{(w)}\right) +$$

$$+ \vec{f}_{n-1}\cdot(1-\kappa)^{\text{T}}A_w(1-\kappa)\vec{f}_{n-1} +$$

$$+ \left[(1-\kappa)^{\text{T}}(A_w + A_w^{\text{T}})\kappa\vec{m}_{(w)} + (1-\kappa)^{\text{T}}\vec{c}_w\right]\cdot\vec{f}_{n-1} +$$

$$\left. + \vec{m}_{(w)}^{\text{T}}\kappa^{\text{T}}A_w\kappa\vec{m}_{(w)} + \vec{c}_w\cdot\kappa\vec{m}_{(w)} + \text{tr}(A_w\Omega) + e_w. \tag{218}$$

Partially differentiating the argument of the maximization with respect to $x_n$ we find that the optimal solution $x^*$ satisfies the recursion,

$$x_n^* = q_v\left[2\eta^{(v)}x_{n-1}^* + B_v + \vec{V}_v\cdot\vec{f}_{n-1}\right], \tag{219}$$

where we have defined

$$q_v = \frac{1}{\gamma\sigma^2 + a^v + 2\eta^{(v)}}, \tag{220}$$

$$\eta^{(v)} = \sum_w p^{(v,w)}\eta_{(w)}, \tag{221}$$

$$a^v = \sum_w p^{(v,w)}a_w, \tag{222}$$

$$B_v = \sum_w p^{(v,w)}\left[b_w + (\vec{\theta} + \vec{d}_w)\cdot\kappa\vec{m}_{(w)}\right], \tag{223}$$

$$\vec{V}_v = (1-\kappa)^{\text{T}}\sum_w p^{(v,w)}(\vec{\theta} + \vec{d}_w). \tag{224}$$

Then the terms in the objective function which involve $x$ can be written in a surprisingly simple form,

$$-(x_{n-1}^*)^2\eta^{(v)} + q_v^{-1}(x_n^*)^2 - \frac{1}{2}q_v^{-1}(x_n^*)^2 = \frac{1}{2}q_v\left[(4(\eta^{(v)})^2 - 2\eta^{(v)}q_v^{-1})(x_{n-1}^*)^2 + B_v^2 + \vec{f}_{n-1}\cdot\vec{V}_v\vec{V}_v^{\text{T}}\vec{f}_{n-1} + \right.$$

$$\left. + 4\eta^{(v)}B_vx_{n-1}^* + 4\eta^{(v)}\vec{V}_v\cdot\vec{f}_{n-1}x_{n-1}^* + 2B_v\vec{V}_v\cdot\vec{f}_{n-1}\right]. \tag{225}$$

So the coefficients in the quadratic ansatz satisfy the following conditions,

$$a_v = -4q_v(\eta^{(v)})^2 + 2\eta^{(v)}, \tag{226}$$

$$b_v = 2q_v\eta^{(v)}B_v, \tag{227}$$

$$A_v = \sum_w p^{(v,w)}(1-\kappa)^{\mathrm{T}}A_w(1-\kappa) + \frac{1}{2}q_v\vec{V}_v\vec{V}_v^{\mathrm{T}}, \tag{228}$$

$$\vec{c}_v = (1-\kappa)^{\mathrm{T}}\sum_w p^{(v,w)}\big[(A_w + A_w^{\mathrm{T}})\kappa\vec{m}_{(w)} + \vec{c}_w\big] + q_vB_v\vec{V}_v, \tag{229}$$

$$\vec{d}_v = 2q_v\eta^{(v)}\vec{V}_v, \tag{230}$$

$$e_v = \sum_w p^{(v,w)}\Big[\vec{m}_{(w)}^{\mathrm{T}}\kappa^{\mathrm{T}}A_w\kappa\vec{m}_{(w)} + \vec{c}_w \cdot \kappa\vec{m}_{(w)} + \mathrm{tr}(A_w\Omega) + e_w\Big] + \frac{1}{2}q_vB_v^2. \tag{231}$$

In addition we have the constraint ensuring concavity of the objective function,

$$\gamma\sigma^2 + a^v + 2\eta^{(v)} > 0. \tag{232}$$

If transaction costs are zero then $\eta_{(w)} = 0$ for all $w$ and hence $a_v = b_v = \vec{d}_v = 0$. Thus

$$x_n^* = \frac{1}{\gamma\sigma^2}\vec{\theta} \cdot \Big[\kappa\vec{m}^{(v)} + (1-\kappa)\vec{f}_{n-1}\Big]. \tag{233}$$

Now consider the case when transaction costs are constant and nonzero so that $\eta_{(w)} = \eta$ for all $w$ and for some constant $\eta$. The coefficients $a_v$ are the solution of the following system of quadratic equations,

$$a_v a^v + a_v(\gamma\sigma^2 + 2\eta^{(v)}) - 2a^v\eta^{(v)} = 2\eta^{(v)}\gamma\sigma^2. \tag{234}$$

Setting $\eta^{(v)} = \eta$ this becomes,

$$a_v a^v + a_v(\gamma\sigma^2 + 2\eta) - 2a^v\eta = 2\eta\gamma\sigma^2. \tag{235}$$

Interestingly, there exists a solution to this system of the form $a_v = a$ where $a$ satisfies the quadratic,

$$a^2 + \gamma\sigma^2 a - 2\eta\gamma\sigma^2 = 0. \tag{236}$$

The solutions are

$$a = \frac{\gamma\sigma^2}{2}\left[-1 \pm \sqrt{1 + \frac{8\eta}{\gamma\sigma^2}}\right]. \tag{237}$$

Substituting back into the concavity constraint we find,

$$\frac{\gamma\sigma^2}{2}\left[1 + \frac{4\eta}{\gamma\sigma^2} \pm \sqrt{1 + \frac{8\eta}{\gamma\sigma^2}}\right] > 0. \tag{238}$$

It follows that if $\eta > 0$ then the constraint is satisfied by both roots of the quadratic. We have,

$$q = \frac{1}{\gamma\sigma^2 + a + 2\eta}. \tag{239}$$

$$\vec{d} = 2q\eta\vec{V}, \tag{240}$$

$$= 2q\eta(1-\kappa)^{\mathrm{T}}(\vec{\theta} + \vec{d}), \implies \tag{241}$$

$$\vec{d} = 2q\eta\big[1 - 2q\eta(1-\kappa)^{\mathrm{T}}\big]^{-1}(1-\kappa)^{\mathrm{T}}\vec{\theta}. \tag{242}$$

We conclude that

$$\vec{V} = q\big[1 - 2q\eta(1-\kappa)^{\mathrm{T}}\big]^{-1}(1-\kappa)^{\mathrm{T}}\vec{\theta}. \tag{243}$$

The coefficient $b$ satisfies

$$b = 2q\eta\big[b + (\vec{\theta} + \vec{d})^{\mathrm{T}}(\vec{\theta} + \vec{d})\big] \implies \tag{244}$$

$$b = \frac{2q\eta}{1 - 2q\eta}(\vec{\theta} + \vec{d})^{\mathrm{T}}\kappa\vec{m}, \tag{245}$$

where $\vec{d}$ is given by the above formula. Thus the strategy is,

$$x_n^* = q\big[2\eta x_{n-1}^* + B + \vec{V} \cdot \vec{f}\big], \tag{246}$$

where

$$B = b + (\vec{\theta} + \vec{d})^{\mathrm{T}}\kappa\vec{m}, \tag{247}$$

$$= \frac{(\vec{\theta} + \vec{d})^{\mathrm{T}}\kappa\vec{m}}{1 - 2q\eta}. \tag{248}$$

where $b$ and $\vec{d}$ are as above.

# 3 Portfolio theory

## 3.1 Markowitz portfolio

Consider $n$ risky assets with expected returns $\mathbb{E}[R_i] = \mu_i$ and let $\Sigma$ denote the matrix of return covariances

$$\sigma_{ij} = \mathbb{E}\big[\big(R_i - \mathbb{E}[R_i]\big)\big(R_j - \mathbb{E}[R_j]\big)\big]. \tag{249}$$

where $i, j \in \{1, \ldots, n\}$. Assume, in addition, that there exists a single riskless asset $R_0$ with non-random return $r$. The total expected return of the portfolio $R = \sum_{i=0}^{n} w_i R_i$ is

$$\mathbb{E}[R] = r + \sum_{i=1}^{n} w_i(\mu_i - r) = r + (\mu - r\mathbf{1})^{\mathrm{T}}\boldsymbol{w}, \tag{250}$$

where we have used the constraint $w_0 = 1 - \sum_{i=1}^{n} w_i$ to express the expected return in terms of the risky allocation vector $\boldsymbol{w} \in \mathbb{R}^n$. The total variance of the portfolio as a function of the risky allocations is,

$$\mathrm{var}(R) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \sigma_{ij} = \boldsymbol{w}^{\mathrm{T}}\Sigma\boldsymbol{w}. \tag{251}$$

We seek to minimize the following objective function,

$$J(\boldsymbol{w}, \lambda, \mu_R) = \frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\Sigma\boldsymbol{w} - \lambda\big[r + (\mu - r\mathbf{1})^{\mathrm{T}}\boldsymbol{w} - \mu_R\big]. \tag{252}$$

The first term is one-half the variance of the portfolio and the second term is proportional to returns in excess of the amount $\mu_R$. The factor of $-\lambda$ is included to encourage excess returns in the minimization problem.

For fixed $\lambda$, the portfolio which minimizes the Lagrangian is given by,

$$\frac{\partial}{\partial\boldsymbol{w}}J(\boldsymbol{w}, \lambda, \mu_R) = 0 \implies \boldsymbol{w} = \lambda\Sigma^{-1}(\mu - r\mathbf{1}). \tag{253}$$

Left-multiplying (253) by $\mathbf{1}^{\mathrm{T}}$ we find the optimal allocation of the riskless asset in terms of $\lambda$,

$$w_0 = 1 - \lambda\mathbf{1}^{\mathrm{T}}\Sigma^{-1}(\mu - r\mathbf{1}). \tag{254}$$

In the case $\lambda = 0$ all of the resources are allocated in the riskless asset and for $\lambda$ sufficiently large we have $w_0 < 0$ which corresponds to borrowing money at the risk-free rate in order to finance a risky portfolio. It follows that $1/\lambda$ should be interpreted as a risk-aversion parameter. Substituting the allocations (253) we find,

$$\mathbb{E}[R] = r + \lambda(\mu - r\mathbf{1})^{\mathrm{T}}\Sigma^{-1}(\mu - r\mathbf{1}), \tag{255}$$
$$\mathrm{var}(R) = \lambda^2(\mu - r\mathbf{1})^{\mathrm{T}}\Sigma^{-1}(\mu - r\mathbf{1}). \tag{256}$$

Since the expected return increases without bound as $\lambda \to \infty$, it is more instructive to consider the so-called Sharp ratio which measures the returns in excess of the risk-free rate, normalized by the standard deviation of the portfolio which is then independent of $\lambda$,

$$\mathrm{SR} = \frac{\mathbb{E}[R] - r}{\sqrt{\mathrm{var}(R)}} = \sqrt{(\mu - r\mathbf{1})^{\mathrm{T}}\Sigma^{-1}(\mu - r\mathbf{1})}. \tag{257}$$

It is also noteworthy that the above formulas imply a trivial relation between the expectation and variance in terms of the risk-aversion parameter; namely,

$$\lambda = \frac{\mathrm{var}(R)}{\mathbb{E}[R] - r}. \tag{258}$$

Using the above formula to eliminate $\lambda$ from (253) we obtain the famous formula linking excess returns of a particular asset to its covariance with the total return,

$$\mu - r\mathbf{1} = \frac{\Sigma\boldsymbol{w}}{\mathrm{var}(R)}\big(\mathbb{E}[R] - r\big) \tag{259}$$

In component notation this says,

$$\mu_i - r = \beta_i\big(\mathbb{E}[R] - r\big), \qquad \beta_i = \frac{\mathrm{cov}(R_i, R)}{\mathrm{var}(R)}. \tag{260}$$

The *Markowitz portfolio* is defined as the allocation vector which minimizes the total variance for fixed expected total return $\mu_R$,

$$\min_{\boldsymbol{w}} \ \boldsymbol{w}^{\mathrm{T}}\Sigma\boldsymbol{w}, \tag{261}$$

$$\text{s.t.} \ \ r + (\mu - r\mathbf{1})^{\mathrm{T}}\boldsymbol{w} - \mu_R = 0. \tag{262}$$

By constrained optimization theory it satisfies,

$$\frac{\partial}{\partial \boldsymbol{w}} J(\boldsymbol{w}, \lambda, \mu_R) = 0, \qquad \frac{\partial}{\partial \lambda} J(\boldsymbol{w}, \lambda, \mu_R) = 0, \tag{263}$$

where the second equation just gives the total expectation constraint $\mu_R = r + (\mu - r\mathbf{1})^{\mathrm{T}}\boldsymbol{w}$. Imposing the total expectation constraint in (253) we can solve for the Lagrange multiplier in terms of the allocations,

$$\lambda = \frac{\mu_R - r}{(\mu - r\mathbf{1})^{\mathrm{T}}\Sigma^{-1}(\mu - r\mathbf{1})}. \tag{264}$$

The expressions for the total expected return and variance are now given by,

$$\mathbb{E}[R] = \mu_R, \tag{265}$$

$$\mathrm{var}(R) = \frac{(\mu_R - r)^2}{(\mu - r\mathbf{1})^{\mathrm{T}}\Sigma^{-1}(\mu - r\mathbf{1})}. \tag{266}$$

Hence we obtain a linear relationship between excess total return and portfolio standard deviation,

$$\mu_R = r + \sigma_R\sqrt{(\mu - r\mathbf{1})^{\mathrm{T}}\Sigma^{-1}(\mu - r\mathbf{1})} \tag{267}$$

Now assume that there is no riskless asset so the expected return is $\mathbb{E}[R] = \mu^{\mathrm{T}}\boldsymbol{w}$. The optimization problem corresponding to the Markowitz portfolio now involves an additional endowment constraint,

$$\min_{\boldsymbol{w}} \ \boldsymbol{w}^{\mathrm{T}}\Sigma\boldsymbol{w}, \tag{268}$$

$$\text{s.t.} \ \ \mu^{\mathrm{T}}\boldsymbol{w} = \mu_R, \tag{269}$$

$$\text{and} \ \ \mathbf{1}^{\mathrm{T}}\boldsymbol{w} = 1. \tag{270}$$

The objective function is now,

$$J(\boldsymbol{w}, \lambda_1, \lambda_2, \mu_R) = \frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\Sigma\boldsymbol{w} - \lambda_1(\mu^{\mathrm{T}}\boldsymbol{w} - \mu_R) - \lambda_2(\mathbf{1}^{\mathrm{T}}\boldsymbol{w} - 1). \tag{271}$$

For fixed $\lambda_1$ and $\lambda_2$, the value of the allocation vector that minimizes this objective is given by

$$\boldsymbol{w} = \lambda_1\Sigma^{-1}\mu + \lambda_2\Sigma^{-1}\mathbf{1}. \tag{272}$$

Applying the total expected return and endowment constraints gives

$$\mu_R = \lambda_1 a + \lambda_2 b, \tag{273}$$

$$1 = \lambda_1 b + \lambda_2 c. \tag{274}$$

where

$$a = \mu^{\mathrm{T}} \Sigma^{-1} \mu, \tag{275}$$

$$b = \mathbf{1}^{\mathrm{T}} \Sigma^{-1} \mu, \tag{276}$$

$$c = \mathbf{1}^{\mathrm{T}} \Sigma^{-1} \mathbf{1}. \tag{277}$$

Solving for the Lagrange multipliers we obtain,

$$\lambda_1 = \frac{b - c\mu_R}{b^2 - ac}, \tag{278}$$

$$\lambda_2 = \frac{b\mu_R - a}{b^2 - ac}. \tag{279}$$

The variance and expected return are now given by,

$$\mathbb{E}[R] = \mu_R \tag{280}$$

$$\mathrm{var}(R) = \lambda_1^2 a + \lambda_2^2 c + 2\lambda_1 \lambda_2 b, \tag{281}$$

$$= \frac{c\mu_R^2 - 2b\mu_R + a}{ac - b^2}, \tag{282}$$

$$= \frac{\mu_R^2 - 2(b/c)\mu_R + a/c}{a - b^2/c}, \tag{283}$$

$$= \frac{\left(\mu_R - b/c\right)^2 + a/c - (b/c)^2}{a - b^2/c}, \tag{284}$$

$$\sigma_R^2 = \left(\frac{\mu_R - b/c}{\sqrt{a - b^2/c}}\right)^2 + \frac{a/c - (b/c)^2}{a - b^2/c}, \tag{285}$$

which is the equation for a hyperbola. This should be compared with the linear relationship obtained with the riskless asset; namely,

$$\mu_R = r + \sigma_R \sqrt{a - 2br + cr^2}. \tag{286}$$

Solving these equations simulatneously we find the point of tangency,

$$\sigma_R = \frac{\sqrt{a - 2br + cr^2}}{b - cr}, \tag{287}$$

$$\mu_R = \frac{a - br}{b - cr}. \tag{288}$$

Setting $\mu_R = (a - br)/(b - cr) + \epsilon$ we obtain

$$w_0 = -\frac{b - cr}{a - 2br + cr^2}\epsilon, \tag{289}$$

$$= -\frac{\mathbf{1}^{\mathrm{T}} \Sigma^{-1}(\mu - r\mathbf{1})}{(\mu - r\mathbf{1})^{\mathrm{T}} \Sigma^{-1}(\mu - r\mathbf{1})}\epsilon. \tag{290}$$
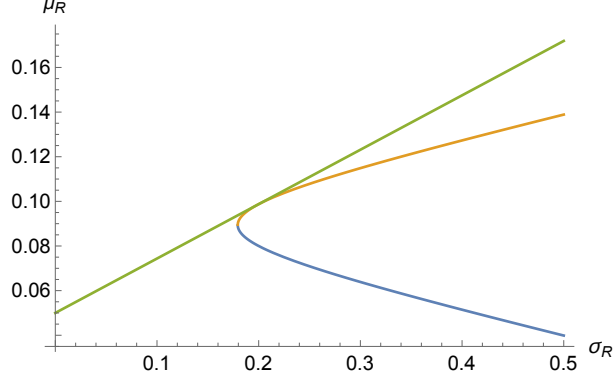
Figure 1: The set of efficient portfolios with and without a risky asset intersect at a single point in the mean-variance space $(\mu_R, \sigma_R)$.

## 3.2 Universal portfolios

Let $\vec{x}_t \in (0, \infty)^n$ represent the relative price change of $n$ assets between times $t$ and $t+1$,

$$x_{t,i} = \frac{\text{Price}_{t+1}(i)}{\text{Price}_t(i)}. \tag{291}$$

Consider a a strategy specified by a constant allocation vector $\vec{b} \in B_n$, where

$$B_n = \{\vec{b} \in \mathbb{R}^n \mid b_i \geq 0, \ \vec{b} \cdot \vec{1} = 1\}. \tag{292}$$

Let's say we have an initial wealth equal to unity, which we invest at time $t = 1$ according to $\vec{b}$. At the end of the first trading period this portfolio achieves a wealth given by,

$$S_1 = \vec{b} \cdot \vec{x}_1. \tag{293}$$

We now rebalance the portfolio such that the wealth $S_1$ is again divided amongst the assets according to $\vec{b}$. At the end of the next trading period the wealth is,

$$S_2 = S_1(\vec{b} \cdot \vec{x}_1) = (\vec{b} \cdot \vec{x}_1)(\vec{b} \cdot \vec{x}_2). \tag{294}$$

Generalizing, we define the wealth achieved by a constantly rebalanced portofolio after the completion of $k$ investment opportunities to be,

$$S_k(\vec{b}) = \prod_{t=1}^{k} \vec{b} \cdot \vec{x}_t, \tag{295}$$

and we define

$$S_T^* = \max_{\vec{b} \in B_n} S_T(\vec{b}), \tag{296}$$

where we have suppressed the dependence of these quantities on the stock sequence $(\vec{x}_1, \cdots, \vec{x}_T)$. The goal is to compare $S_T^*$ to the wealth achieved by an adapted strategies of the form,

$$\hat{S}_T = \prod_{t=1}^{T} \vec{b}_t \cdot \vec{x}_t. \tag{297}$$

33

In the case of the universal portfolio the allocations are given by,

$$\vec{b}_t = \frac{\displaystyle\int_{B_n} d^n\vec{b}\, S_{t-1}(\vec{b})\, \vec{b}}{\displaystyle\int_{B_n} d^n\vec{b}\, S_{t-1}(\vec{b})}, \quad 1 \le t \le T. \tag{298}$$

Substituting $t = 1$ we find the explicit expression $\vec{b}_1 = \frac{1}{m}\vec{1}$. Computing the inner product $\vec{b}_t \cdot \vec{x}_t$ we find the convenient identity,

$$\vec{b}_t \cdot \vec{x}_t = \frac{\displaystyle\int_{B_n} d^n\vec{b}\, S_t(\vec{b})}{\displaystyle\int_{B_n} d^n\vec{b}\, S_{t-1}(\vec{b})}, \quad 1 \le t \le T. \tag{299}$$

It follows from the above identity that $\hat{S}_T$ is simply the average of $S_T(\vec{b})$ over the simplex $B_n$,

$$\hat{S}_T = \frac{\displaystyle\int_{B_n} d^n\vec{b}\, S_T(\vec{b})}{\displaystyle\int_{B_n} d^n\vec{b}}. \tag{300}$$

Setting $T = 3$, for example, we find

$$\hat{S}_3 = \frac{\displaystyle\int_{B_n} d^n\vec{b}\, \vec{b} \cdot \vec{x}_1}{\displaystyle\int_{B_n} d^n\vec{b}} \cdot \frac{\displaystyle\int_{B_n} d^n\vec{b}\, S_1(\vec{b})\, \vec{b} \cdot \vec{x}_2}{\displaystyle\int_{B_n} d^n\vec{b}\, S_1(\vec{b})} \cdot \frac{\displaystyle\int_{B_n} d^n\vec{b}\, S_2(\vec{b})\, \vec{b} \cdot \vec{x}_3}{\displaystyle\int_{B_n} d^n\vec{b}\, S_2(\vec{b})} = \frac{\displaystyle\int_{B_n} d^n\vec{b}\, S_3(\vec{b})}{\displaystyle\int_{B_n} d^n\vec{b}}. \tag{301}$$

The integral over $B_n$ can be performed by mapping it to $\mathbb{S}^{n-1}$. For example in the case of $B_4$, we have

$$b_1 = \cos^2 \theta_1, \tag{302}$$
$$b_2 = \sin^2 \theta_1 \cos^2 \theta_2, \tag{303}$$
$$b_3 = \sin^2 \theta_1 \sin^2 \theta_2 \cos^2 \theta_3, \tag{304}$$
$$b_4 = \sin^2 \theta_1 \sin^2 \theta_2 \sin^2 \theta_3 \tag{305}$$

where $0 \le \theta_1, \theta_2, \theta_3 \le \pi/2$. In the two-asset case, the metric on the simplex is

$$ds^2 = 8 \sin^2 \theta_1 \cos^2 \theta_1 d\theta_1 \implies \int_{B_1} db = \int_0^{\pi/2} ds = \sqrt{2}. \tag{306}$$

In the three-asset case we have,

$$\sqrt{g} = \sqrt{12 \cos^2 \theta_1 \sin^6 \theta_1 \sin^2(2\theta_2)} \implies \int_{B_2} d^2\vec{b} = \int_0^{\pi/2} d\theta_1 \int_0^{\pi/2} d\theta_2 \sqrt{g} = \frac{\sqrt{3}}{2}. \tag{307}$$

# 4 Fundamental theorem of asset pricing

## 4.1 Single-period models

Consider a stock with share price $S_t$ per share where $S_0$ is known but $S_1$ is uncertain. Let the money-market be an asset with share price $e^{tr}$ with $t = 0, 1$ and no uncertainty.

### 4.1.1 Forward contracts

The buyer agrees to pay the seller an amount $F$ in exchange for one share of stock at time $t = 1$. The forward price $F$ is decided at time $t = 0$. Naive expectation pricing would lead to price $F = \mathbb{E}[S_1]$ where the expectation is with respect to the real-word probabilities. In fact, there will be an arbitrage opportunity unless $F = S_0 e^r$.

*Proof.* Consider the following trading strategies. (Cash-and-carry arbitrage) At $t = 0$ the seller borrows an amount $S_0$ from the money-market to buy 1 share and enters a forward contract with the buyer. At $t = 1$ the stock is sold to the buyer at the forward price $F$ and the seller settles their accounts with the money market, making a profit $F - S_0 e^r$. (Reverse cash-and-carry arbitrage) At $t = 0$ buyer shorts 1 share of stock, invests the proceeds $S_0$ in the money-market and enters the forward contract with the seller $S$. At $t = 1$, the buyer buys the stock from the seller at the agreed forward price $F$ and the short position can now be settled. The amount of profit to the buyer is $S_0 e^r - F$.
   Notice that if the forward contract is undervalued ($F < S_0 e^r$) then buyer has an arbitrage opportunity while if $F > S_0 e^r$ then the seller has an arbitrage opportunity. Therefore absence of arbitrage implies $F = S_0 e^r$. □

### 4.1.2 European call

Consider a European call with strike price $K$. This allows the buyer of the option to buy the underlying stock at price $K$. Let $V_t$ denote the value of the option and time $t = 0, 1$. Then $V_1 = \max(0, S_1 - K)$. Let $r$ denote the risk-free rate of return. Consider a two-market scenario where $S_1(\omega_1) \leq K \leq S_1(\omega_2)$ and $S_1(\omega_1) \leq S_0 e^r \leq S_1(\omega_2)$. There will be an arbitrage opportunity unless

$$V_0 = (d_2 - K)\frac{S_0 e^r - d_1}{d_2 - d_1}e^{-r} \tag{308}$$

where $d_i = S_1(\omega_i)$.

*Proof.* (i) At $t = 0$ shorts $a$ shares of the underlying stock and the buy the option for the price $V_0$. Invest the remainder $aS_0 - V_0$ is invested in the money-market. At $t = 1$ the option is exercised iff $\omega_2$ is realized. The pay-off at time $t = 1$ is given by

$$e^r(aS_0 - V_0) - aS_1(\omega) + V_1(\omega). \tag{309}$$

Suppose $\omega_1$ is realized so $V_1(\omega_1) = 0$. Then the profit is

$$\varphi(\omega_1) = e^r(aS_0 - V_0) - ad_1 = a(S_0 e^r - d_1) - e^r V_0 \tag{310}$$

The pay-off will be positive provided

$$V_0 < a(S_0 e^r - d_1)e^{-r} \equiv V \tag{311}$$

If $\omega_2$ is realized then the pay-off is

$$\varphi(\omega_2) = e^r(aS_0 - V_0) - ad_2 + (d_2 - K) \tag{312}$$
$$> e^r(aS_0 - V) - ad_2 + (d_2 - K) \tag{313}$$
$$\varphi(\omega_2) > ad_1 - ad_2 + (d_2 - K) \tag{314}$$

So for $\varphi(\omega_2) > 0$ we want

$$a = \frac{d_2 - K}{d_2 - d_1}. \tag{315}$$

(ii) Now consider the seller of the option. If $\omega_2$ is realized they are out of the money $(-d_2 + K)$. If, however, they borrow $aS_0 - V_0$ from the money market and buy $a$ shares of stock then the pay-offs are

$$\varphi(\omega_2) = -(S_0 a - V_0)e^r + ad_2 + (-d_2 + K) > 0 \tag{316}$$
$$\varphi(\omega_1) = -(S_0 a - V_0)e^r + ad_1 > 0. \tag{317}$$

Note that this is the reverse of the (i). If $V_0 > S_0 a$ there is no need to borrow from the money-market. We simply purchase $a$ shares using the proceeds of the call option and invest the remainder $V_0 - aS_0$. The pay-offs are then

$$\varphi(\omega_2) = (V_0 - S_0 a)e^r + ad_2 + (-d_2 + K) > 0 \tag{318}$$
$$\varphi(\omega_1) = (V_0 - S_0 a)e^r + ad_1 > 0. \tag{319}$$

$\square$

Notice that $V_0$ can be written in terms of a probability measure as

$$V_0 = e^{-r} \sum_{\omega \in \Omega} P(\omega)V_1(\omega) = e^{-r}E(V_1) \tag{320}$$

where $V_1(\omega_2) = d_2 - K$, $V_1(\omega_1) = 0$ and

$$P(\omega_2) = \frac{S_0 e^r - d_1}{d_2 - d_1}, \quad P(\omega_1) = 1 - P(\omega_2). \tag{321}$$

Calculating the expectation of $S_1$ with respect to this measure we find $E(S_1) = S_0 e^r$ and thus

$$S_0 = e^{-r}E(S_1). \tag{322}$$

A simpler explanation: Consider a portfolio which is long $a$ shares and short a call. Then after one period the portfolio has value $ad_2 - (d_2 - K)$ if the share price goes up and has value $ad_1$ if the share price goes down. If the portfolio has zero risk then these must be equal and thus

$$a = \frac{d_2 - K}{d_2 - d_1}. \tag{323}$$

36

Since this portfolio is riskless its return must be equal to the interest earned at the risk free rate on the initial value of the portfolio; that is

$$(S_0 a - V_0)e^r = a d_1 \implies V_0 = (d_2 - K)\frac{S_0 e^r - d_1}{d_2 - d_1}e^{-r} \tag{324}$$

A replicating portfolio argument: Since a call is bullish the replicating portfolio must buy some stock (say $\Delta$ shares) and borrow some money $B$,

$$\Pi_t = \Delta S_t - B e^{rt}. \tag{325}$$

We choose $\Delta$ and $B$ such that the final pay-off of the portfolio matches the final pay-off of the call. We then obtain two equations corresponding to the two market outcomes

$$\Delta d_2 - B e^r = d_2 - K \tag{326}$$
$$\Delta d_1 - B e^r = 0 \tag{327}$$

which implies

$$\Delta = \frac{d_2 - K}{d_2 - d_1} \tag{328}$$
$$= \frac{c_2 - c_1}{d_2 - d_1} \tag{329}$$
$$B = e^{-r} d_1 \Delta \tag{330}$$

The value of the call should be equal to the initial cost of the replicating portfolio; namely

$$\Pi_0 = \Delta S_0 - B = (d_2 - K)\frac{S_0 e^r - d_1}{d_2 - d_1}e^{-r}. \tag{331}$$

## 4.2   Existence and uniquness of risk-neutral measure

Consider a single-period market with $m$ market scenarios $\{\omega_i\}_{i=1}^m$ and $n + 1$ traded assets with share prices $\{S_t^j\}_{j=0}^n$ where $t \in \{0, 1\}$. The first asset is is assumed to be riskless with share price at time $t$ given by $S_t^0 = e^{tr}$. A portfolio is defined by a vector in $\mathbb{R}^{n+1}$. The value of the portfolio at time $t = 0$ is known and the value of at time $t = 1$ depends on which market scenario is realized,

$$V_0(\theta) = \sum_{j=0}^n \theta_j S_0^j, \tag{332}$$

$$V_1(\theta; \omega_i) = \sum_{j=0}^n \theta_j S_1^j(\omega_i). \tag{333}$$

An arbitrage is a portfolio $\theta$ which satisfies either of the following properties,

$$V_0(\theta) \le 0, \quad V_1(\theta; \omega_i) > 0, \quad \forall i \in \{1, \dots, m\}, \tag{334}$$
$$V_0(\theta) < 0, \quad V_1(\theta; \omega_i) \ge 0, \quad \forall i \in \{1, \dots, m\}. \tag{335}$$

A risk-neutral measure is defined as a probability distribution function $\pi$ with the property that

$$S_0^j = e^{-r} \sum_{i=1}^{m} \pi(\omega_i) S_1^j(\omega_i), \quad \forall j \in \{0, \ldots, n\}. \tag{336}$$

**Theorem 4.1.** *There exists a risk-neutral measure $\iff$ there does not exist arbitrage.*

*Proof.* ($\implies$) Suppose there exists a risk-neutral measure. Then summing over the $j$ index in (336) with an arbitrary portfolio $\theta_j$ we obtain

$$V_0(\theta) = e^{-r} \sum_{i=1}^{m} \pi(\omega_i) V_1(\theta; \omega_i). \tag{337}$$

It is easy to see that the condition for an arbitrage is incompatible with this equation because $\pi(\omega_i) \geq 0$. ($\impliedby$) We begin by noting that the equation

$$S_0^0 = e^{-r} \sum_{i=1}^{m} \pi(\omega_i) S_1^0(\omega_i) \tag{338}$$

holds for any PDF because $S_t^0 = e^{tr}$. We therefore define a vector in $\mathbb{R}^n$ by the equation

$$\vec{y} = e^{-r} \sum_{i=1}^{m} \pi(\omega_i) \vec{S}_1(\omega_i). \tag{339}$$

where

$$\vec{S}_1(\omega_i) = \left( S_1^1(\omega_i), \ldots, S_1^n(\omega_i) \right). \tag{340}$$

If we define

$$\xi = \{ \vec{y} \mid \pi \text{ is a PDF} \} \tag{341}$$

then $\xi$ is the convex hull of the $m$ points $\{ \vec{S}_1(\omega_1), \ldots, \vec{S}_1(\omega_m) \} \subset \mathbb{R}^n$. The existence of a risk-neutral measure is thus equivalent to the statement that $\vec{S}_0 \in \xi$. We assume for a contradiction that there does not exist a risk-neutral measure; that is, $\vec{S}_0 \notin \xi$. The separating hyperplane theorem ensures that there exists a hyperplane in $\mathbb{R}^n$ separating the convex set $\xi$ and from $\vec{S}_0 \notin \xi$. Let us parametrize the separating hyperplane by

$$H = \{ \vec{x} \in \mathbb{R}^n \mid \vec{\theta} \cdot \vec{x} = \alpha \}. \tag{342}$$

The statement that $H$ separates $\vec{S}_0$ from $\xi$ is equivalent to $\vec{\theta} \cdot \vec{S}_0 < \alpha$ and $\vec{\theta} \cdot \vec{y} \geq \alpha$ for all $\vec{y} \in \xi$, which implies $\vec{\theta} \cdot \vec{S}_0 < \vec{\theta} \cdot \vec{y}$ for all $\vec{y} \in \xi$. Since each $\vec{y}$ corresponds to a choice of PDF, we choose the $\vec{y}$ corresponding to

$$\pi(\omega_i) = \begin{cases} 1, & i = k \\ 0, & \text{o.w.} \end{cases}. \tag{343}$$

It follows that for all $k \in \{1, \ldots, m\}$ we have

$$\vec{\theta} \cdot \vec{S}_0 < e^{-r} \vec{\theta} \cdot \vec{S}_1(\omega_k). \tag{344}$$

Thus

$$\sum_{j=1}^{n} \theta_j S_0^j < e^{-r} \sum_{j=1}^{n} \theta_j S_1^j(\omega_i), \quad \forall i \in \{1, \ldots, m\}. \tag{345}$$

It follows that we can find $\theta_0 \in \mathbb{R}$ such that adding to both sides we obtain

$$\sum_{j=0}^{n} \theta_j S_0^j < 0 < e^{-r} \sum_{j=0}^{n} \theta_j S_1^j(\omega_i), \quad \forall i \in \{1, \ldots, m\}. \tag{346}$$

$\square$

**Example 4.1.** Consider an economy with two traded assets and two states of the world $\omega_1$ and $\omega_2$. As usual one of the assets is assumed to be riskless with price process $B_t = e^{rt}$. The risk-neutral probabilities are defined by

$$\begin{bmatrix} B_0 \\ S_0 \end{bmatrix} = e^{-r} \begin{bmatrix} B_1(\omega_1) & B_1(\omega_2) \\ S_1(\omega_1) & S_1(\omega_2) \end{bmatrix} \begin{bmatrix} \pi(\omega_1) \\ \pi(\omega_2) \end{bmatrix}. \tag{347}$$

Now consider a call option whose pay-off is given by $V_1(\omega) = (S_1(\omega) - K)^+$ where $d_1 < K < d_2$. The value at time zero is given by the discounted expectation

$$V_0 = e^{-r} \big[ V_1(\omega_1)\pi(\omega_1) + V_1(\omega_2)\pi(\omega_2) \big] = (d_2 - K)\frac{S_0 e^r - d_1}{d_2 - d_1} e^{-r}. \tag{348}$$

**Example 4.2.** Now consider an economy with two tradeable assets and three states of the world. The risk-neutral probabilities are defined by

$$\begin{bmatrix} B_0 \\ S_0 \end{bmatrix} = e^{-r} \begin{bmatrix} B_1(\omega_1) & B_1(\omega_2) & B_1(\omega_3) \\ S_1(\omega_1) & S_1(\omega_2) & S_1(\omega_3) \end{bmatrix} \begin{bmatrix} \pi(\omega_1) \\ \pi(\omega_2) \\ \pi(\omega_3) \end{bmatrix}. \tag{349}$$

Consider a call option with strike $d_1 < K < d_3$ and assume $d_1 < d_2 < d_3$. Then

$$V_0 = e^{-r} \big[ (d_2 - K)_+ p_2 + (d_3 - K)p_3 \big] \tag{350}$$

where $p_2$ and $p_3$ are constrained to satisfy

$$S_0 e^r = p_1 d_1 + p_2 d_2 + p_3 d_3, \quad p_1 + p_3 + p_3 = 1, \quad p_1 > 0, \quad p_2 > 0, \quad p_3 > 0. \tag{351}$$

We shall now assume that the single-period market above is arbitrage-free and thus that there exists a unique risk-neutral probability measure. We define a derivative security as any asset whose value at time $t = 1$ is a random variable.

**Theorem 4.2.** *The risk neutral measure is unique $\iff$ the set $\{S_1^j\}_{j=0}^{n}$ spans the space of random variables.*

*Proof.* ($\impliedby$) We will show that the risk-neutral measure is unique by explicitly constructing it. Let $1 \leq i \leq m$ be arbitrary and let $V^{(i)}$ denote the random variable which is equal to 1 if

$\omega_i$ occurs and 0 otherwise; that is, $V^{(i)}(\omega_k) = \delta_{ik}$. By the hypothesis there exists a portfolio vector $\theta^{(i)}$ such that

$$V^{(i)}(\omega_k) = \sum_{j=0}^{n} \theta_j^{(i)} S_1^j(\omega_k). \tag{352}$$

If we assume that $V^{(i)}$ represents the value of an asset $V_t^{(i)}$ at $t = 1$, then the value of the asset at time zero is given by

$$V_0^{(i)} = \sum_{j=0}^{n} \theta_j^{(i)} S_0^j. \tag{353}$$

On the other hand, the absence of arbitrage implies that a risk-neutral probability measure $\pi$ exists and thus we can express the value of the portfolio at time zero as,

$$V_0^{(i)} = e^{-r} \sum_{k=1}^{m} \pi(\omega_k) V_1^{(i)}(\omega_k) = e^{-r} \pi(\omega_i). \tag{354}$$

It follows that the unique risk-neutral measure is given by

$$\pi(\omega_i) = e^r \sum_{j=0}^{n} \theta_j^{(i)} S_0^j \quad \forall i \in \{1, \ldots, m\}. \tag{355}$$

( $\implies$ ) We show the contrapositive statement; that is, if there exists a random variable $V$ which cannot be replicated by the tradeable assets then the risk-neutral measure is not unique. It is convenient to define the following vectors in $\mathbb{R}^m$,

$$\mathbf{V} = (V(\omega_1), \ldots, V(\omega_m)), \tag{356}$$
$$\mathbf{S}^j = (S_1^j(\omega_1), \ldots, S_1^j(\omega_m)) \quad \forall j \in \{0, \ldots, n\}. \tag{357}$$

If $\mathbf{V}$ is not a linear combination of the $n+1$ vectors $\{\mathbf{S}^j\}_{j=0}^{n}$ then it follows that $\{\mathbf{S}^j\}_{j=0}^{n}$ does not span $\mathbb{R}^m$. There thus exists a vector $\mathbf{v} \in \mathbb{R}^m$ which is orthogonal to $\mathbf{S}^j$ for all $0 \le j \le n$. In particular, $\mathbf{v} \cdot \mathbf{S}^0 = 0$ implies

$$\sum_{i=1}^{m} v_i = 0. \tag{358}$$

Suppose we are given a risk-neutral probability measure $\pi$. We will show that there exists $\epsilon > 0$ such that the following equation defines a second risk neutral measure,

$$\pi^*(\omega_i) = \pi(\omega_i) + \epsilon v_i. \tag{359}$$

It is clear that $\sum_{i=1}^{m} \pi^*(\omega_i) = 1$ and moreover (since $\mathbf{v} \cdot \mathbf{S}^j$ for all $0 \le j \le n$),

$$\sum_{i=1}^{m} \pi^*(\omega_i) S_1^j(\omega_i) = \sum_{i=1}^{m} \pi(\omega_i) S_1^j(\omega_i) \quad \forall j \in \{0, \ldots, n\}. \tag{360}$$

The only other requirement for $\pi^*$ to define a risk-neutral measure is that $\pi^*(\omega_i) > 0$ for all $1 \le i \le m$. This can always be achieved by choosing $\epsilon$ sufficiently small since $\pi(\omega_i) > 0$ for all $1 \le i \le m$. $\qquad\square$

## 4.3 Continuous-time models

A numeraire is any strictly positive self-financing portfolio. The most common numeraires are the price $P(t, T)$ of a non-defaultable zero-coupon bond with maturity $T$ and face value equal to 1, $N(t) = P(t, T)$. The other is the value of a money-market account with instantaneous risk-free rate $r(t)$, $N(t) = B(t) \equiv B(0)e^{\int_0^t ds\, r(s)}$.

The relative gain process with respect to a numeraire $N(t)$ for a dividend-paying asset $S(t)$ is

$$\text{RGP}(t) = \frac{S(t)}{N(t)} + \int_0^t \frac{dD(s)}{N(s)} \tag{361}$$

where $D(t) = \int_0^t ds\, D(s)$ is the cumulative dividend paid by the asset from time 0 to $t$. The fundamental theorem states that in the absence of arbitrage there exists a probability measure $Q_N$ with respect to which the relative gain processes are martingales, where $N$ is any numeraire process,

$$\text{RGP}(t) = \mathbb{E}_t^{Q_N}[\text{RGP}(T)]. \tag{362}$$

For non-dividend-paying assets this reduces to

$$\frac{S(t)}{N(t)} = \mathbb{E}_t^{Q_N}\left[\frac{S(T)}{N(T)}\right]. \tag{363}$$

The probability measures associated with non-defaultable ZCBs and the money-market account are called the $T$-forward and risk-neutral probability measures, respectively.

As an example consider a non-dividend paying asset with pay-off $V(T)$ (for example a European option). Using the money-market account as the numeraire we obtain the price at time $t$

$$V(t) = B(t)\mathbb{E}_t^{Q_B}\left[\frac{V(T)}{B(T)}\right] \tag{364}$$

$$= \mathbb{E}_t^{Q_B}\left[\frac{B(t)}{B(T)}V(T)\right] \tag{365}$$

$$= \mathbb{E}_t^{Q_B}\left[e^{-\int_t^T ds\, r(s)}V(T)\right]. \tag{366}$$

If the interest rate is not stochastic then the integral can be pulled outside of the expectation value. For stochastic interest rates, however, this is not possible and it is more convenient to use the $T$-forward measure which prices the asset as (using $P(T, T) = 1$)

$$V(t) = P(t, T)\mathbb{E}_t^{Q_T}[V(T)] \tag{367}$$

The continuously compounded zero-coupon rate $r(t, T)$ for maturity $T$ at time $t$ is defined by

$$P(t, T) = e^{-r(t,T)(T-t)}. \tag{368}$$

If interest rates are constant then absence of arbitrage implies $r(t, T) = r$ and thus $Q_B = Q_T$. In general the price of a zero-coupon bond can be related to the instantaneous interest rate using the FTAP with the money-market as the numeraire (a ZCB does not pay dividends)

$$P(t, T) = B(t)\mathbb{E}_t^{Q_B}\left[\frac{P(T, T)}{B(T)}\right] = B(t)\mathbb{E}_t^{Q_B}\left[\frac{1}{B(T)}\right] = \mathbb{E}_t^{Q_B}\left[e^{-\int_t^T ds\, r(s)}\right]. \tag{369}$$

# 5 Equity derivatives

## 5.1 Forwards and futures prices

To price a forward contract we view it as a derivative with pay-off function for the seller given by

$$V(T) = S(T) - K \tag{370}$$

where $K$ is the delivery price and $T$ is the delivery date for the forward contract. In other words the seller pays $S(T)$ and receives $K$ from the buyer. Notice that for any $t < T$ the value of the forward is zero since the forward contract can be entered at zero cost at any time. It follows from the FTAP that

$$0 = \mathbb{E}_t^{Q_N} \left[ \frac{S(T) - G(t,T)}{N(T)} \right] \tag{371}$$

where we have denoted the forward price at time $t$ for delivery at $T$ by $G(t,T)$. Using the $T$-forward measure we obtain $G(t,T) = \mathbb{E}_t^{Q_T}[S(T)]$. Since the forward price converges to the spot price at maturity $G(T,T) = S(T)$ we also have $G(t,T) = \mathbb{E}_t^{Q_T}[G(T,T)]$. This shows that the forward price is a Martingale under the $T$-forward measure. If $S$ does not pay dividends, then the forward price is determined by

$$G(t,T) = \mathbb{E}_t^{Q_T} \left[ \frac{S(T)}{P(T,T)} \right] = \frac{S(t)}{P(t,T)} \tag{372}$$

If $S$ pays dividends, then we introduce a non-dividend-paying asset $V$ which is defined by re-investing all dividends. The value of such an asset is given by

$$V(t) = S(t) e^{\int_0^t ds\, \delta(s)}. \tag{373}$$

Since $V$ does not pay dividends, then by the above argument the corresponding forward price must be

$$G_V(t,T) = \mathbb{E}_t^{Q_T}[V(T)] = \mathbb{E}_t^{Q_T} \left[ \frac{V(T)}{P(T,T)} \right] = \frac{V(t)}{P(t,T)} = \frac{S(t) e^{\int_0^t ds\, \delta(s)}}{P(t,T)} \tag{374}$$

If the dividend rate is not stochastic then

$$G_V(t,T) = \mathbb{E}_t^{Q_T} \left[ S(T) e^{\int_0^T ds\, \delta(s)} \right] \tag{375}$$

$$= \mathbb{E}_t^{Q_T}[S(T)] \, e^{\int_0^T ds\, \delta(s)} \tag{376}$$

$$= G(t,T) e^{\int_0^T ds\, \delta(s)} \tag{377}$$

and thus

$$G(t,T) = G_V(t,T) e^{\int_T^0 ds\, \delta(s)} = \frac{S(t)}{P(t,T)} e^{-\int_t^T ds\, \delta(s)}. \tag{378}$$

Like a forward contract, a futures contract is an agreement which can be entered at any time $t < T$ at zero cost. During the interval $[t,T]$, the buyer collects the gains and loses given by the changes in the futures price

$$\int_t^T dF(s,T) = F(T,T) - F(t,T). \tag{379}$$

Since the futures price converges to the spot price at maturity $F(T, T) = S(T)$, the futures contract has the same pay-off as a forward contract. Unlike the forward contract, however, there is no exchange of the underlying. The futures price is a Martingale under the risk-neutral probability measure

$$F(t, T) = \mathbb{E}_t^{Q_B} [F(T, T)].$$ (380)

*Proof.* Consider the following trading strategy. At time $t$ invest $F(t, T)$ in the money-market and acquire $n(t)$ futures contracts at no cost. After a time $\Delta t$ has elapsed, the money-market account has grown to $F(t, T)e^{r(t)\Delta t}$ and the futures position yields a pay-off $n(t)[F(t+\Delta t, T) - F(t, T)]$ which is invested. If we choose $n = e^{r(t)\Delta t}$ then the account balance at time $t + \Delta t$ is $F(t, T)e^{r(t)\Delta t} + n(t)[F(t + \Delta t, T) - F(t, T)] = e^{r(t)\Delta t}F(t + \Delta t, T)$. Now increase the future position to $n(t+\Delta t)n(t) = e^{[r(t+\Delta t)+r(t)]\Delta t}$ contracts. The value of the money-market account at time $t + 2\Delta t$ is

$$e^{r(t+\Delta t)\Delta t}[e^{r(t)\Delta t}F(t + \Delta t, T)] = e^{[r(t)+r(t+\Delta t)]\Delta t}F(t + \Delta t, T)$$ (381)

and after investing the pay-off from the new futures position yields a total

$$e^{[r(t)+r(t+\Delta t)]\Delta t}F(t+\Delta t, T)+n(t+\Delta t)n(t)[F(t+2\Delta t, T)-F(t+\Delta t, T)] = e^{[r(t)+r(t+\Delta t)]\Delta t}F(t+2\Delta t, T)$$ (382)

Continuing this process until the contract maturity date $T$ and taking the continuous limit we obtain the following pay-off at maturity

$$e^{\int_t^T r(s)ds}F(T, T) = e^{\int_t^T r(s)ds}S(T).$$ (383)

By the FTAP the value of this pay-off at $t < T$ is given by (using the risk-neutral measure

$$V(t) = B(t)\mathbb{E}_t^{Q_B}\left[\frac{e^{\int_t^T r(s)ds}S(T)}{B(T)}\right]$$ (384)

$$= \mathbb{E}_t^{Q_B}\left[\frac{B(t)}{B(T)}e^{\int_t^T r(s)ds}S(T)\right]$$ (385)

$$= \mathbb{E}_t^{Q_B}[S(T)]$$ (386)

On the other hand, our investment at time $t$ to yield this return was $F(t, T)$. So we obtain

$$F(t, T) = \mathbb{E}_t^{Q_B}[S(T)] = \mathbb{E}_t^{Q_B}[F(T, T)]$$ (387)

$\square$

## 5.2 Bounds on puts and calls

### 5.2.1 Put-call parity

Consider a European call and put on an underlying asset $S$, which have identical strikes $K$ and expirations $T$. Denote their time-$t$ prices by $c(t)$ and $p(t)$. A portfolio consisting of a

call and $K$ unit-face ZCBs with expiration $T$ has an identical pay-off to a portfolio consisting of a put and one unit of the underlying; namely[1]

$$\max(S(T) - K, 0) + K = \max(K - S(T), 0) + S(T) = \max(S(T), K) \quad (388)$$

Applying the FTAP to the first portfolio with the $T$-forward measure we obtain

$$P(t,T)\mathbb{E}_t^{Q_T}\left[\max(S(T) - K, 0) + K\right] = c(t) + KP(t,T) \quad (389)$$

and similarly for the second portfolio

$$P(t,T)\mathbb{E}_t^{Q_T}\left[\max(K - S(T), 0) + S(T)\right] = p(t) + P(t,T)G(t,T) \quad (390)$$

where we have used $S(T) = G(T,T)$ and $G(t,T)$ is a martingale under the $T$-forward measure. We therefore have the identity

$$c(t) + P(t,T)K = p(t) + P(t,T)G(t,T) \quad (391)$$

or

$$c(t) - p(t) = P(t,T)(G(t,T) - K). \quad (392)$$

Recall that for non-dividend-paying stock we have $G(t,T) = S(t)/P(t,T)$ and thus

$$c(t) - p(t) = S(t) - P(t,T)K. \quad (393)$$

We see that if the present value of the strike is equal to the prevailing stock price then puts and calls have the same value. These options are said to be at the money (ATM). On the other hand if the stock price exceeds $P(t,T)K$ then call is said to be in the money (ITM) and the put is out of the money (OTM). Thus the call is more expensive than the put. The opposite holds true if the stock price falls below $P(t,T)K$.

### 5.2.2 Lower bounds for calls

Consider a portfolio consisting of a European call and $K$ unit-face ZCBs. The pay-off of this portfolio is

$$\max(0, S(T) - K) + K = \max(K, S(T)). \quad (394)$$

Notice that the pay-off from this portfolio is at least as big as a portfolio consisting of just one unit of the share; that is,

$$\max(0, S(T) - K) + K \geq S(T). \quad (395)$$

Applying the fundamental theorem of asset pricing we obtain

$$P(t,T)\mathbb{E}_t^{Q_T}[\max(0, S(T) - K) + K] \geq P(t,T)\mathbb{E}_t^{Q_T}[S(T)] \quad (396)$$

If the stock does not pay dividends then

$$\mathbb{E}_t^{Q_T}[S(T)] = \mathbb{E}_t^{Q_T}\left[\frac{S(T)}{P(T,T)}\right] = \frac{S(t)}{P(t,T)} \quad (397)$$

---

[1]The first portfolio can be thought of as purchasing a call and lending money.

and thus

$$c(t) + P(t,T)K \geq S(t) \tag{398}$$

or

$$c(t) \geq S(t) - P(t,T)K. \tag{399}$$

If the put-call parity relation is violated, then there is an arbitrage opportunity that can be established using a "buy low, sell high" strategy. It consists of purchasing the call and shorting the stock and investing the proceeds. If $S(T) > K$ then the stock is purchased through the call at price $K$ leading to a final pay-off which is positive by assumption

$$\frac{S(t) - c(t)}{P(t,T)} - K > 0. \tag{400}$$

On the other hand if $S(T)$ falls below $K$ then the call option expires out of the money and the stock is bought at the market price leading to pay-off which is also positive

$$\frac{S(t) - c(t)}{P(t,T)} - S(T) > \frac{S(t) - c(t)}{P(t,T)} - K > 0. \tag{401}$$

The price of an American call $C(t)$ is at least the price of the corresponding European call and thus

$$C(t) \geq S(t) - P(t,T)K. \tag{402}$$

We can now understand why early exercise of an American call option on a non-dividend paying asset is never favorable. The pay-off for an American call from immediate exercise at time $t$ is $S(t) - K$. On the other hand we have just seen that the value of the American call satisfies

$$C(t) \geq S(t) - P(t,T)K \geq S(t) - K, \tag{403}$$

and hence the option is worth more unexercised. If one believes that the stock price will drop below the strike, rendering the call worthless, then it is still not favorable early exercise. One should simply sell the option at the price $C(t)$.

### 5.2.3 Lower bounds for puts

Consider a European put and one unit of the share (no dividends). The pay-off is $\max(S(T), K)$. This is at least the pay-off of $K$ ZCBs; namely $\max(S(T), K) \geq K$. Applying the FTAP we obtain

$$p(t) + S(t) \geq KP(t,T) \implies p(t) \geq KP(t,T) - S(t) \tag{404}$$

Obviously for an American call we have $P(t) \geq p(t) \geq KP(t,T) - S(t)$ but in fact $P(t) \geq K - S(t) \geq KP(t,T) - S(t)$. This is because if $P(t) < K - S(t)$ then we can buy the put and exercise it immediately yielding a positive cash flow.

### 5.2.4   Convexity of puts and calls with respect to strike

The price of a European call (or put) is a convex function of the strike price; that is

$$\frac{V(K+\delta)+V(K-\delta)}{2} > V(K). \tag{405}$$

This result can be proved by showing that the converse leads to an arbitrage. The arbitrage position is a "buy low, sell high" strategy of buying a call of strike $K+\delta$, buying a call of strike $K-\delta$ and writing two calls at strike $K$, all of the same expiration $T$. We know this strategy can be financed because by assumption we have

$$2V(K) \geq V(K+\delta)+V(K-\delta). \tag{406}$$

One can show that the four possible market scenarios: $S(T) > K+\delta$, $K < S(T) < K+\delta$, $K-\delta < S(T) < K$ and $S(T) < K-\delta$ lead to a non-negative pay-off. The same strategy also works for put options.

Note that just two option quotes of the same maturity are sufficient to detect an arbitrage opportunity. This is because the prevailing prices for vanilla calls and puts with strike $K = 0$ are fixed to be $c(t) = S(t)$ and $p(t) = 0$, respectively. An arbitrage position can be constructed by buying the undervalued option and shorting the requisite number of overvalued options.

## 5.3   Risk-neutral pricing

Let us assume that the underlying asset is modeled by an Ito process,

$$dS(t) = S(t)\mu(t)dt + S(t)\sigma(t)dw(t) \tag{407}$$

Assume that the underlying pays dividends at a rate $\delta(t)$ and that the risk-free interest rate is given by $r(t)$. Then $D(t) = \int_0^t S(s)\delta(s)ds$ and $B(t) = B(0)e^{\int_0^t r(s)ds}$. Taking the differential of the relative gain process using Leibniz's rule gives

$$d\,\mathrm{RGP}(t) = d\left[\frac{S(t)}{N(t)} + \int_0^t \frac{dD(s)}{N(s)}\right] \tag{408}$$

$$= d\left(\frac{S(t)}{B(t)}\right) + \frac{S(t)}{B(t)}\delta(t)dt \tag{409}$$

$$= \frac{1}{B(t)}\left[dS(t) + (\delta(t)-r(t))S(t)dt\right] \tag{410}$$

$$= \frac{S(t)}{B(t)}(\mu(t)+\delta(t)-r(t))dt + \frac{S(t)}{B(t)}\sigma(t)dw(t) \tag{411}$$

By Girsanov's theorem the underlying asset is an Ito process with respect to the risk-neutral probability measure

$$dS(t) = S(t)\hat{\mu}(t)dt + S(t)\sigma(t)d\hat{w}(t) \tag{412}$$

The differential of the relative gain process with respect to the risk-neutral measure is

$$d\,\mathrm{RGP}(t) = \frac{S(t)}{B(t)}(\hat{\mu}(t)-r(t)+\delta(t))dt + \frac{S(t)}{B(t)}\sigma(t)d\hat{w}(t). \tag{413}$$

By the fundamental theorem of asset pricing the relative gain process is a Martingale with respect to $Q_B$ and hence $\hat{\mu}(t) = r(t) - \delta(t)$ and consequently the asset price follows,

$$dS(t) = S(t)[r(t) - \delta(t)]dt + S(t)\sigma(t)d\hat{w}(t). \tag{414}$$

## 5.4   Black-Scholes model

If we assume $r(t) = r$, $\delta(t) = \delta$ and $\sigma(t) = \sigma$ then the SDE corresponds to geometric Brownian motion which has the exact solution

$$S(t) = S(0)e^{\left(r-\delta-\frac{1}{2}\sigma^2\right)t+\sigma\hat{w}(t)} \tag{415}$$

so

$$S(T) = S(t)e^{\left(r-\delta-\frac{1}{2}\sigma^2\right)(T-t)+\sigma(\hat{w}(T)-\hat{w}(t))} \tag{416}$$

and thus

$$\log S(T) \sim \mathcal{N}\left(\log S(t) + \left(r - \delta - \frac{1}{2}\sigma^2\right)(T-t), \sigma^2(T-t)\right) \tag{417}$$

where we have used $\hat{w}(T) - \hat{w}(t) \sim \mathcal{N}(0, T-t)$. It is now possible to compute the time-$t$ value of a European derivative with pay-off $V(T) = \varphi(S(T))$ at time $T$ using the FTAP

$$V(t) = e^{-r(T-t)}\mathbb{E}_t^{Q_B}[V(T)]. \tag{418}$$

Notice that we have applied the FTAP in its non-dividend-paying form because a European derivative does not produce any cash-flow until the maturity date $T$.

European calls and puts can be priced analytically in the Black-Scholes model. It is useful to express the pay-off of a European call as

$$\max(S(T) - K, 0) = S(T)1_{\{S(T)\geq K\}} - K1_{\{S(T)\geq K\}}. \tag{419}$$

The pricing formula for a call follows from the fact that if $\log X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ then

$$\mathbb{E}[1_{X\geq K}] = P(X \geq K) = N\left(\frac{\mu_X - \log K}{\sigma_X}\right) \tag{420}$$

$$\mathbb{E}[X1_{X\geq K}] = e^{\mu_X+\sigma_X^2/2}N\left(\frac{\mu_X + \sigma_X^2 - \log K}{\sigma_X}\right) \tag{421}$$

where

$$N(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} dz\, e^{-z^2/2} \tag{422}$$

is the cumulative distribution function for the standard normal distribution. Substituting $\mu_X = \log\left(S(t)e^{(r-\delta)(T-t)}\right) - \frac{1}{2}\sigma^2(T-t)$ and $\sigma_X = \sigma\sqrt{T-t}$ we find

$$c(t) = S(t)e^{-\delta(T-t)}N(y + \sigma\sqrt{T-t}) - Ke^{-r(T-t)}N(y) \tag{423}$$

where

$$y = \frac{\log\left(\frac{S(t)e^{(r-\delta)(T-t)}}{K}\right) - \frac{1}{2}\sigma^2(T-t)}{\sigma\sqrt{T-t}}. \tag{424}$$

The price of a European put can be obtained in a similar fashion, or alternatively by using the put-call parity relation

$$c(t) - p(t) = e^{-r(T-t)}(S(t)e^{(r-\delta)(T-t)} - K) \tag{425}$$

which gives

$$p(t) = Ke^{-r(T-t)}N(-y) - S(t)e^{-\delta(T-t)}N(-y - \sigma\sqrt{T-t}). \tag{426}$$

where we have used $N(x) + N(-x) = 1$.

Let us consider the various limits of the Black-Scholes pricing formula. If we let $K \to 0$ then the pay-off for a European call option is simply the value of the underlying at the maturity date of the call. If the underlying does not pay dividends we expect $c(t) \to S(t)$ and indeed this happens because $y \to \infty$ as $K \to 0$. If the underlying pays dividends then $c(t) \to S(t)e^{-\delta(T-t)}$ because the call option does not entitle the holder to the dividend payments. Similarly a call with an infinite strike will never be exercised and is thus worthless so $c(t) \to 0$ as $K \to \infty$ which can be seen from the fact that $y \to -\infty$ as $K \to \infty$. For a European put option $p(t) \to 0$ as $K \to 0$ because a put with vanishing strike entitles the holder to nothing. As $K \to \infty$ we have $p(t) \simeq Ke^{-r(T-t)} - S(t)e^{-\delta(T-t)}$ because a put with infinite strike is guaranteed to be exercised at maturity. As $T \to \infty$ the price of a European call non a non-dividend paying asset approaches $S(t)$ while for a put $p(t) \to 0$ as $T \to \infty$.
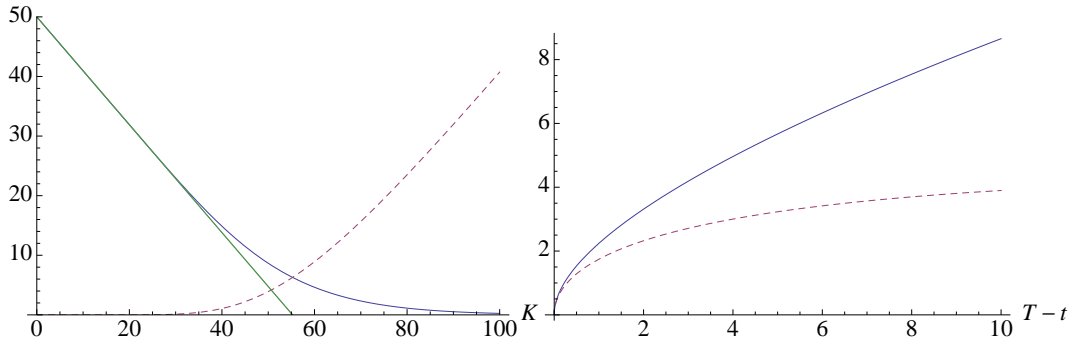


Figure 2: Left: Call price (solid) and put price (dashed) as a function of strike for $S(t) = 50$, $t - T = 10$, $r = 0.01$, $\sigma = 0.1$ and $\delta = 0$. The green lines shows the put-call parity relation. Right: Time-dependence of call and put price with fixed strike $K = 50$.

### 5.4.1 Black-Scholes PDE

Now consider the price process $V(t)$ of a general European-style derivative of the underlying $S(t)$. The discounted price process $\hat{V}(t) = e^{-rt}V(t)$ is a Martingale with respect to the risk-neutral probability measure,

$$\hat{V}(t) = \mathbb{E}_t^{Q_B}[\hat{V}(T)]. \tag{427}$$

Taking the differential of $\hat{V}(t)$ and applying Ito's lemma we find

$$d\hat{V}(t) = e^{-rt}\left[dV(t) - rV(t)\right]. \tag{428}$$

where

$$dV(t) = \left[\frac{\partial V}{\partial t} + \frac{\partial V}{\partial S}S(t)(r-\delta) + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}S(t)^2\sigma^2\right]dt + \frac{\partial V}{\partial S}S(t)\sigma d\hat{w}(t). \tag{429}$$

The Martingale condition then implies the following partial differential equation,

$$\frac{\partial V(S,t)}{\partial t} + \frac{\partial V(S,t)}{\partial S}S(r-\delta) + \frac{1}{2}\frac{\partial^2 V(t,S)}{\partial S^2}S^2\sigma^2 - rV(S,t) = 0. \tag{430}$$

A slicker way to obtain the Black-Scholes PDE is to apply the Feynman-Kac formula to $V(t) = e^{-r(T-t)}\mathbb{E}_t^{Q_B}[V(T)]$. More generally, if we assume deterministic discounting then $V(t) = e^{-\int_t^T ds\, r(s)}\mathbb{E}_t^{Q_B}[V(T)]$. If we further assume deterministic volatility and dividend rate then the Feynman-Kac formula gives

$$\frac{\partial V(S,t)}{\partial t} + \frac{\partial V(S,t)}{\partial S}S\big(r(t) - \delta(t)\big) + \frac{1}{2}\frac{\partial^2 V(S,t)}{\partial S^2}S^2\sigma(t)^2 - r(t)V(S,t) = 0. \tag{431}$$

### 5.4.2 Relationship with the heat equation

The Black-Scholes differential equation can be converted to the heat equation. We first perform a field redefinition to the discounted price process $u = e^{-rt}V$ which eliminates the $rV$ term,

$$\frac{\partial u}{\partial t} + \frac{\partial u}{\partial S}S(r-\delta) + \frac{1}{2}\frac{\partial^2 u}{\partial S^2}S^2\sigma^2 = 0. \tag{432}$$

Now perform the coordinate transformation $\tilde{x} = \log S$ and $\tilde{\tau} = T - t$ which brings it to the form of arithmetic, rather than geometric Brownian motion,

$$\frac{\partial u}{\partial \tilde{\tau}} - \left(r - \delta - \frac{\sigma^2}{2}\right)\frac{\partial u}{\partial \tilde{x}} - \frac{\sigma^2}{2}\frac{\partial^2 u}{\partial \tilde{x}^2} = 0. \tag{433}$$

Finally, the drift term can be removed by performing the following Galilean boost,

$$x = \tilde{x} + (r - \delta - \sigma^2/2)\tilde{\tau} \tag{434}$$
$$\tau = \tilde{\tau}. \tag{435}$$

By the chain rule we have

$$\frac{\partial}{\partial \tilde{x}} = \frac{\partial x}{\partial \tilde{x}}\frac{\partial}{\partial x} + \frac{\partial \tau}{\partial \tilde{x}}\frac{\partial}{\partial \tau} = \frac{\partial}{\partial x} \tag{436}$$
$$\frac{\partial}{\partial \tilde{\tau}} = \frac{\partial x}{\partial \tilde{\tau}}\frac{\partial}{\partial x} + \frac{\partial \tau}{\partial \tilde{\tau}}\frac{\partial}{\partial \tau} = (r - \delta - \sigma^2/2)\frac{\partial}{\partial x} + \frac{\partial}{\partial \tau} \tag{437}$$

Then

$$\frac{\partial u}{\partial \tau} = \frac{1}{2}\sigma^2\frac{\partial^2 u}{\partial x^2}. \tag{438}$$

The solution of the heat equation can be expressed as an integral transform of the initial condition

$$u(x, \tau) = \frac{1}{\sigma\sqrt{2\pi\tau}} \int_{-\infty}^{\infty} ds \, u(x, 0) \exp\left[-\frac{(x-s)^2}{2\sigma^2\tau}\right]. \tag{439}$$

For example, if $u(x, 0) = \delta(x)$ then

$$u(x, \tau) = \frac{1}{\sigma\sqrt{2\pi\tau}} \exp\left(-\frac{x^2}{2\sigma^2\tau}\right). \tag{440}$$

At any fixed $\tau$ this describes a normal distribution in $x$ with mean zero and variance $\sigma^2\tau$. Let us consider a European call option. The boundary condition for the time direction is the final condition $V(S, T) = \max(S - K, 0)$. In terms of the new variables this becomes an initial condition $u(x, 0) = \max(e^x - K, 0)$ In the stock price direction we impose two boundary conditions valid for all $t$, $V(0, t) = 0$ and $V(S, t) \to S$ as $S \to \infty$.

### 5.4.3 Replicating portfolio argument

Another perspective on the Black-Scholes PDE is provided by considering a self-financing portfolio $\Pi(t)$ which is a linear combination of the asset $S(t)$ and the money market account $B(t)$. We will assume that the underlying asset does not pay dividends. Since $S(t)$ and $V(t)$ have the same stochastic driver, this guarantees that we can construct $\Pi(t)$ to replicate $V(t)$.

$$\Pi(t) = x(t)B(t) + y(t)S(t). \tag{441}$$

A process is self-financing if the gains and losses of the portfolio are entirely due to gains and losses on the underlying asset; that is,

$$d\Pi(t) = x(t)dB(t) + y(t)dS(t) \tag{442}$$
$$= [rx(t)B(t) + y(t)S(t)r] \, dt + \sigma y(t)S(t)d\hat{w}(t). \tag{443}$$

Matching coefficients of differentials with $dV(t)$ (setting $\delta = 0$ for simplicity) we obtain

$$y(t) = \frac{\partial V}{\partial S}, \qquad rx(t)B(t) = \frac{\partial V}{\partial t} + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}S(t)^2\sigma^2. \tag{444}$$

Substituting $x(t)$ and $y(t)$ back into the expression for $\Pi(t)$ and using the fact that $\Pi(t) = V(t)$ for all $t$ we obtain the Black-Scholes equation,

$$\frac{\partial V}{\partial t} + \frac{\partial V}{\partial S}Sr + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}S^2\sigma^2 - rV = 0. \tag{445}$$

### 5.4.4 Hedging argument

Suppose that a Financial institution writes a call option valued at $V(t)$ on an underlying asset with value $S(t)$ at time $t$. In order to minimize exposure they buy $\Delta(t)$ shares of the

underlying. The value of the firm's portfolio at time $t$ is $\Pi(t) = -V(t) + \Delta(t)S(t)$. The change in value of the portfolio after an instant of time $dt$ is given by

$$d\Pi(t) = -dV(t) + \Delta(t)dS(t) + \Delta(t)S(t)\delta dt \tag{446}$$

$$= \left[\Delta(t) - \frac{\partial V}{\partial S}\right]dS(t) - \left[\frac{\partial V}{\partial t} + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}S(t)^2\sigma^2 - \Delta(t)S(t)\delta\right]dt \tag{447}$$

where we have made the Black-Scholes assumption that $\sigma(t) = \sigma$. If we assume $\Delta(t) = \partial V/\partial S$ then we notice that the growth in the portfolio is deterministic and so by absence of arbitrage we must have $d\Pi(t) = r\Pi(t)dt$. This implies

$$\frac{\partial V}{\partial t} + \frac{\partial V}{\partial S}S(r - \delta) + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}S^2\sigma^2 - rV = 0. \tag{448}$$

which is the Black-Scholes equation. Another way to think about it is that the firm can finance the stock and call position by initially borrowing an amount at time $t$ given by $V(t) - \Delta(t)S(t) = -\Pi(t)$. Thus the initial value of the portfolio taking into account borrowing is zero. In an instant of time $dt$ the change in the value of the portfolio taking into account borrowing is $d\Pi(t) - \Pi(t)rdt$ which can be evaluated to give

$$d\Pi(t) - \Pi(t)rdt = -\left[\frac{\partial V}{\partial t} + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}S(t)^2\sigma^2 + (r - \delta)\Delta S(t) - rV(t)\right]dt + \left[\Delta(t) - \frac{\partial V}{\partial S}\right]dS(t). \tag{449}$$

We see that this vanishes if both $\Delta(t) = \partial V/\partial S$ and the Black-Scholes equation is satisfied. The firm can make a profit by charging an additional premium on top of the Black-Scholes no-arbitrage call price.

The Black-Scholes delta can be computed either by using the Chain rule, or alternatively by noticing that the Black-Scholes call price is homogeneous in the stock price and the strike; that is,

$$c(\alpha S, \alpha K) = \alpha \, c(S, K). \tag{450}$$

It follows from Euler's theorem that

$$c(S, K) = S\frac{\partial c}{\partial S} + K\frac{\partial c}{\partial K}. \tag{451}$$

Comparing with the Black-Scholes call price we conclude that

$$\Delta_{\text{BS}} = e^{-\delta(T-t)}N(y + \sigma\sqrt{T - t}). \tag{452}$$

### 5.4.5 Discrete-time hedging

Suppose a call option on a non-dividend paying stock is sold at time $t_0$ and delta hedged at a discrete set of times $t_j = t_0 + j\Delta t$. Let $S_i = S(t_i)$ and $\Delta_i = \Delta_{\text{BS}}(t_i)$. At time $t_0$ we buy $\Delta_0$ shares so that stock position is $\Delta_0 S_0$. This position is financed by the sale of the call and borrowing from the money market. The borrowing position at time $t_0$ is $B_0 \equiv c(t_0) - \Delta_0 S_0$ so that the net position is

$$\underbrace{-c(t_0)}_{\text{sell call}} + \underbrace{\Delta_0 S_0}_{\text{buy stock}} + \underbrace{c(t_0) - \Delta_0 S_0}_{\text{amount borrowed}} = 0. \tag{453}$$

At time $t_1$ the share holdings has changed in value to $\Delta_0 S_1$ and we change the number of holdings to $\Delta_1$ shares. This is accomplished by selling the $\Delta_0$ shares at the prevailing price $S_1$ and purchasing $\Delta_1$ shares at the same price $S_1$. It is clear that this position can be financed by borrowing an amount $(\Delta_0 - \Delta_1)S_1$ from the money market. The borrowing position is now

$$B_1 = B_0 e^{r\Delta t} + (\Delta_0 - \Delta_1)S_1. \tag{454}$$

Similarly at time $t_j$ the number of shares is chosen to be $\Delta_j$. Iterating we find that the borrowing position required to maintain the delta hedge is given by

$$B_j = B_{j-1} e^{r\Delta t} + (\Delta_{j-1} - \Delta_j)S_j \tag{455}$$

where $B_0 = c(t_0) - \Delta_0 S_0$.

### 5.4.6  Path-dependent derivatives

A path-dependent derivative $V$ has a pay-off that depends on the price of the underlying asset $S$ at a set of times $t \le t_1 \le \cdots \le t_n \le T$,

$$V(T) = \varphi(S(t_1), \ldots S(t_n)). \tag{456}$$

For the risk-neutral probability measure with constant interest $r$, the FTAP gives

$$V(t) = e^{-r(T-t)} \mathbb{E}_t^{Q_B} \left[ \varphi(S(t_1), \ldots S(t_n)) \right] \tag{457}$$

To compute this expectation value by Monte Carlo simulation one generates a sample path with increments given by $t_1 - t, t_2 - t_1, \ldots$. From the sample path one obtains a value of the pay-off $\varphi(S(t_1), \ldots S(t_n))$. Repeating this process many times and averaging the resulting set of pay-offs returns an approximation to the expectation value.

### 5.4.7  Volatility from time series

Recall that in the Black-Scholes model we have the following distribution of the stock price increments

$$\log S(t + \Delta t) - \log S(t) \sim \mathcal{N}\left( \left( r - \delta - \frac{1}{2}\sigma^2 \right) \Delta t, \sigma^2 \Delta t \right). \tag{458}$$

Suppose we have $n + 1$ observations of the stock price $S_0, S_1, \ldots S_n$ measured at fixed time intervals $\Delta t$. The unbiased estimate of the variance (second central moment) of logarithmic price differences is

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (u_i - \bar{u})^2, \qquad u_i = \log S_i - \log S_{i-1}. \tag{459}$$

We then have the following estimate for the volatility in terms of $s$,

$$\sigma = \frac{s}{\sqrt{\Delta t}}. \tag{460}$$

The assumption of log-normality can be tested by computing higher central moments. For example the third central moment or skewness should vanish for a normal distribution. The unbiased estimate of the skewness is calculate by

$$S = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} (u_i - \bar{u})^3. \tag{461}$$

Similarly the kurtosis (fourth central moment divided by square of second central moment) should equal 3 for a normal distribution. The formula for the unbiased estimate of the kurtosis is more complicated.

## 5.5  Black model

Assume that that the forward price of the underlying asset $S$ follows an Ito process with constant relative volatility

$$dG(t,T) = \mu(t)G(t,T)dt + G(t,T)\sigma dw(t) \tag{462}$$

By Girsanov's theorem, $G(t,T)$ is an Ito process with respect to the $T$-forward measure with the same volatility. Moreover $G(t,T)$ is a martingale with respect to the $T$-forward measure so it must have zero drift and thus

$$dG(t,T) = G(t,T)\sigma d\hat{w}(t). \tag{463}$$

Recall that this SDE has solution

$$G(t,T) = G(0,T)e^{-\frac{1}{2}\sigma^2 t + \sigma \hat{w}(t)}. \tag{464}$$

Since the forward price converges to the spot price at maturity we have

$$S(T) = G(T,T) = G(0,T)e^{-\frac{1}{2}\sigma^2 T + \sigma \hat{w}(T)} = G(t,T)e^{-\frac{1}{2}\sigma^2 (T-t) + \sigma[\hat{w}(T) - \hat{w}(t)]} \tag{465}$$

which implies

$$\log S(T) \sim \mathcal{N}\left(\log G(t,T) - \frac{1}{2}\sigma^2(T-t), \sigma^2(T-t)\right) \tag{466}$$

Thus we know the distribution of $S(T)$ without any assumption about interest rates or dividends provided that we know the current forward prices $G(t,T)$. The price of a European derivative is computed by the FTAP as in the Black-Scholes model, except now with respect to the $T$-forward measure

$$V(t) = P(t,T)\mathbb{E}_t^{Q_T}[\varphi(G(T,T))]. \tag{467}$$

European call and put options can be explicitly priced in the Black model

$$c(t) = P(t,T)\mathbb{E}_t^{Q_T}\left[(G(T,T) - K)^+\right] \tag{468}$$

$$= P(t,T)\left[G(t,T)N(y + \sigma\sqrt{T-t}) - KN(y)\right] \tag{469}$$

$$p(t) = P(t,T)\mathbb{E}_t^{Q_T}\left[(K - G(T,T))^+\right] \tag{470}$$

$$= P(t,T)\left[KN(-y) - G(t,T)N(-y - \sigma\sqrt{T-t})\right] \tag{471}$$

where

$$y = \frac{\log\left(\frac{G(t,T)}{K}\right) - \frac{1}{2}\sigma^2(T-t)}{\sigma\sqrt{T-t}}. \tag{472}$$

Note that if the interest rate is constant and the stock pays dividends at a constant rate $r$ then recall that

$$G(t,T) = S(t)e^{(r-\delta)(T-t)} \tag{473}$$

and thus we recover the Black-Scholes result. The pricing formulas for the Black model can be obtained from the Black-Scholes pricing formulas by setting $\delta = r$, replacing $r$ by the continuously compounded zero-coupon rate $r(t,T)$ and replacing $S(t)$ by forward price $G(t,T)$.

### 5.5.1 Black equation

It is possible to represent the Black pricing formula for European derivatives as the solution of a partial differential equation. In the Black model the price at time zero of a European derivative with pay-off $\varphi(S(T))$ is given by the following discounted expectation

$$V(0) = P(0,T)\mathbb{E}^{Q_T}\left[\varphi(G(T,T))\right] \tag{474}$$

where we have used the fact that the forward price converges to the spot price at maturity $G(T,T) = S(T)$. Knowledge of the forward price dynamics under the $T$-forward measure and of the yield curve at time zero allows the above quantity to be estimated by Monte Carlo simulation. In particular, since the forward price is a Martingale under the $T$-forward measure, the Black specification of geometric Brownian motion corresponds to the following SDE

$$dG(t,T) = \sigma G(t,T)d\hat{w}(t) \tag{475}$$

where $\hat{w}$ is a Brownian motion under $Q_T$. If we define the instantaneous forward rate by

$$f(t,s) = -\frac{\partial}{\partial s}\log P(t,s) \tag{476}$$

then

$$V(0) = \mathbb{E}^{Q_T}\left[e^{-\int_0^T ds\, f(0,s)}\varphi(G(T,T))\right]. \tag{477}$$

Comparing with the Feynman-Kac formula we see that that $V(0)$ can be obtained by backward evolving the terminal condition $V(G,T) = \varphi(G)$ to time zero under the following PDE,

$$\frac{\partial V(G,t)}{\partial t} + \frac{1}{2}\sigma^2 G^2 \frac{\partial^2 V(G,t)}{\partial G^2} - f(0,t)V(G,t) = 0. \tag{478}$$

### 5.5.2 Volatility surfaces

The Black model expresses the price of European vanilla options as a function of the volatility parameter $\sigma$ and four market observables; the forward price $G(t,T)$ of the underlying asset, the price $P(t,T)$ of a non-defaultable ZCB with unit face value, the strike price $K$, and the time to expiration $T-t$ of the option. This implies a one-to-one map between prices of

European vanillas and volatility parameters. A plot of the Black-implied volatility parameter versus strike price and time to expiration is called a volatility surface.

In order to determine the Black-implied volatilities one needs to know the forward prices $G(t, T)$ as a function of the expiration $T$. The forward contract can be priced by looking at European put and call options with the same strike $K$ expiration equal to $T$. By put-call parity relation we then have

$$G(t, T) = K + \frac{c(t) - p(t)}{P(t, T)}. \tag{479}$$

The ZCB prices $P(t, T)$ can be found from the continuously compounded zero-coupon rates $P(t, T) = e^{r(t,T)(T-t)}$ which are determined by the LIBOR zero coupon yield curve. This method leads to slight variability in the forward prices from options with different strikes but same expiration. The best estimate for the forward price is determined by choosing the strike corresponding to the most near-the-money option. A call is in the money if $G(t, T) > K$ and a put is in the money if $K > G(t, T)$. It follows from the put-call parity formula that the most near-the-money options have the smallest value of $|c(t) - p(t)|$.

Having computed the forward prices we can determine the black-implied volatilities. Typically out-of-the-money options are used for these calculations because they are the most heavily traded options. For example if the traded option is a call then we will select it if $G(t, T) - K < 0$ and similarly a put is selected if $G(t, T) - K > 0$.

If the options are American then put-call parity does not apply. If we assume that the stock does not pay dividends then the forward price is given by

$$G(t, T) = \frac{S(t)}{P(t, T)} = S(t)e^{r(t,T)(T-t)}. \tag{480}$$

Since it is never favorable to exercise early an American call option on a non-dividend paying asset, the price of an American call is the same as the price of the corresponding European call. Therefore one simply uses the call options (both OTM and ITM) to calculate Black-implied volatilities, treating them as though they were European.

### 5.5.3 Volatility skews and smiles

The Black-implied volatilities of short-maturity options display a convex dependence on the strike price $K$ referred to as a volatility smile. As time to expiration increases, this behavior gives way a monotonically decreasing dependence on $K$. The shape of the volatility surface reflects the risk-neutral distribution of $\log S(T)$. In particular, excess kurtosis results in symmetric volatility smiles while negative skewness increases the value of low-strike and decreases the value of high-strike options leading to a volatility skew.

The fact that the Black-implied volatilities of short-maturity options rises sharply for low and high $K$ indicates that the log-normality assumption fails to capture the possibility of extreme stock price movements. In other words, the Black model is trying to compensate for the absence of a fat-tailed probability distribution by increasing the standard deviation for these strikes.

The negative skewness can be attributed to several factors including risk aversion; namely, that volatility increases when markets are falling. It can also be understood as a consequence

of the leverage effect. The total value of a company is given by the sum of its equity and debt $V = E + D$, this implies

$$\frac{dE}{E} = \frac{V}{E}\frac{dV}{V} - \frac{D}{E}\frac{dD}{D}. \tag{481}$$

Under the simplifying assumption that the debt is risk free, the change in the company's value is entirely due to the change in equity and thus

$$\sigma_E \approx \frac{V}{E}\sigma_V. \tag{482}$$

If the price process for the value $V$ follows geometric Brownian motion then $\sigma_V = $ cst and it follows that $\sigma_E$ is negatively correlated with equity price.

## 5.6   Local volatility models

A forward local volatility model assumes that the forward price on the underlying asset follows a diffusion process,

$$dG(t,T) = G(t,T)\mu(t)dt + G(t,T)\sigma(G(t,T),t)dw(t). \tag{483}$$

Changing to the $T$-forward measure using Girsanov's theorem and using the fact that $G(t,T)$ is a martingale with respect to the $T$-forward measure we obtain

$$dG(t,T) = \sigma(G(t,T),t)d\hat{w}(t) \tag{484}$$

where $\hat{w}$ is a Brownian motion under $Q_T$. The CEV model postulates that $\sigma(G,t) = \alpha G^{-\beta}$ where $\alpha, \beta \geq 0$. We have

$$dG(t,T) = \alpha G(t,T)^{1-\beta}d\hat{w}(t). \tag{485}$$

Note that for $\beta > 1$ the volatility diverges as $G(t,T) \to 0$ which leads to slow convergence of the Euler discretization. Convergence can be improved by changing to a variable which has constant volatility. The appropriate change of variables is $X(t) = G(t,T)^\beta/\beta$. Then by Ito's lemma we obtain

$$dX(t) = G(t,T)^{\beta-1}dG(t,T) + (\beta - 1)G(t,T)^{\beta-2}dG(t,T)^2 \tag{486}$$

$$= \frac{\alpha^2(\beta - 1)}{2\beta X(t)}dt + \alpha d\hat{w}(t). \tag{487}$$

For $\beta = 0$ we recover the Black model and thus the forward price can never vanish. For $\beta > 0$, however, we see that $X(t)$ (and thus $G(t,T)$) can vanish with positive probability.

European calls can be priced analytically in the CEV model,

$$c(t) = P(t,T)\left[G(t,T)\left(1 - \chi^2(2y; 2 + 2\nu, 2x)\right) - K\chi^2(2x; 2\nu, 2y)\right] \tag{488}$$

where

$$\nu = \frac{1}{2\beta} \tag{489}$$

$$x = \frac{2}{T}\left(\frac{\nu}{\alpha}\right)^2 G(t,T)^{2\beta} \tag{490}$$

$$y = \frac{2}{T}\left(\frac{\nu}{\alpha}\right)^2 K^{2\beta} \tag{491}$$

and $\chi^2(x; n, \lambda)$ is the cumulative distribution function for the non-central chi-squared distribution with $n$ degrees of freedom and non-centrality parameter $\lambda$.

## 5.7 Stochastic volatility models

A spot stochastic volatility model assumes that the stock volatility satisfies its own Ito process,

$$dS(t) = S(t)(r - \delta)dt + \sqrt{v(t)}S(t)d\hat{w}(t) \tag{492}$$

$$dv(t) = \alpha(S(t), v(t), t)dt + \sqrt{v(t)}\beta(S(t), v(t), t)d\hat{w}_v(t) \tag{493}$$

where $\hat{w}$ and $\hat{w}_v$ are correlated Brownian motions with correlation $\rho$,

$$\text{cov}(d\hat{w}, d\hat{w}_v) = \mathbb{E}[d\hat{w}\, d\hat{w}_v] = \rho\, dt. \tag{494}$$

For simulating sample paths it is more convenient to change variables to $X(t) = \log S(t)$. It follows from (131) that

$$dX(t) = \left[r - \delta - \frac{1}{2}v(t)\right]dt + \sqrt{v(t)}d\hat{w}(t). \tag{495}$$

It is easy to show that in stochastic volatility models a delta hedge is incapable of replicating a derivative asset. This can be seen by considering the delta hedged portfolio

$$\Pi(t) = -V(t) + \Delta(t)S(t). \tag{496}$$

The change in value in an instant of time is given by

$$
\begin{aligned}
d\Pi &= -\left[\frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial S}dS + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}dS^2 + \frac{\partial V}{\partial v}dv + \frac{1}{2}\frac{\partial^2 V}{\partial v^2}dv^2 + \frac{\partial V}{\partial S\partial v}dS\, dv\right] + \Delta dS + \Delta\, S\delta dt \\
&= -\left[\frac{\partial V}{\partial t} - \Delta\, S\delta + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}S^2 v + \frac{1}{2}\frac{\partial^2 V}{\partial v^2}\beta^2 v + \frac{\partial^2 V}{\partial S\partial v}S\beta v\rho\right]dt + \left[\Delta - \frac{\partial V}{\partial S}\right]dS - \frac{\partial V}{\partial v}dv
\end{aligned}
\tag{497}
$$

Hedging can be achieved by introducing a second derivative asset $V_1$ which depends on $v$,

$$\Pi(t) = -V(t) + \Delta(t)S(t) + \Delta_1(t)V_1(t). \tag{498}$$

In an instant of time the change in the value of the portfolio is

$$
\begin{aligned}
d\Pi = &-\left[\frac{\partial V}{\partial t} - \Delta\, S\delta + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}S^2 v + \frac{1}{2}\frac{\partial^2 V}{\partial v^2}\beta^2 v + \frac{\partial^2 V}{\partial S\partial v}S\beta v\rho\right]dt + \\
&+ \Delta_1\left[\frac{\partial V_1}{\partial t} + \frac{1}{2}\frac{\partial^2 V_1}{\partial S^2}S^2 v + \frac{1}{2}\frac{\partial^2 V_1}{\partial v^2}\beta^2 v + \frac{\partial^2 V_1}{\partial S\partial v}S\beta v\rho\right]dt \\
&+ \left[\Delta + \Delta_1\frac{\partial V_1}{\partial S} - \frac{\partial V}{\partial S}\right]dS + \left[\Delta_1\frac{\partial V_1}{\partial v} - \frac{\partial V}{\partial v}\right]dv
\end{aligned}
\tag{499}
$$

It follows that we should choose

$$\Delta_1 = \frac{\frac{\partial V}{\partial v}}{\frac{\partial V_1}{\partial v}}, \qquad \Delta = \frac{\partial V}{\partial S} - \Delta_1 \frac{\partial V_1}{\partial S} \tag{500}$$

Making these substitutions and setting $d\Pi = \Pi r dt$ we obtain

$$\frac{\frac{\partial V}{\partial t} + \frac{\partial V}{\partial S}S(r-\delta) + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}S^2 v + \frac{1}{2}\frac{\partial^2 V}{\partial v^2}\beta^2 v + \frac{\partial^2 V}{\partial S \partial v}S\beta v\rho - rV}{\frac{\partial V}{\partial v}} = (V \longleftrightarrow V_1) \tag{501}$$

Since the left-hand side only depends on $V$ and the right-hand side depends only on $V_1$, they must equal the same function $f(S, v, t)$ and thus

$$\frac{\partial V}{\partial t} + \frac{\partial V}{\partial S}S(r-\delta) + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}S^2 v + \frac{1}{2}\frac{\partial^2 V}{\partial v^2}\beta^2 v + \frac{\partial^2 V}{\partial S \partial v}S\beta v\rho - rV = f\frac{\partial V}{\partial v}. \tag{502}$$

We can learn about the function $f$ by considering a delta-hedged portfolio

$$\Pi_1(t) = -V(t) + \frac{\partial V}{\partial S}S(t). \tag{503}$$

We have

$$d\Pi_1 - r\Pi_1 dt = -\left[\frac{\partial V}{\partial t} + \frac{\partial V}{\partial S}S(r-\delta) + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}S^2 v + \frac{1}{2}\frac{\partial^2 V}{\partial v^2}\beta^2 v + \frac{\partial^2 V}{\partial S \partial v}S\beta\rho - rV\right]dt - \frac{\partial V}{\partial v}dv$$

$$= \frac{\partial V}{\partial v}(f - \alpha)\,dt - \frac{\partial V}{\partial v}\sqrt{v}\beta\,d\hat{w}_v \tag{504}$$

It can be argued that working in the risk-neutral probability measure $\alpha$ should be chosen to cancel $f$ which is equivalent to setting $f = \alpha$ in the pricing equation

$$\frac{\partial V}{\partial t} + \frac{\partial V}{\partial S}S(r-\delta) + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}S^2 v + \frac{1}{2}\frac{\partial^2 V}{\partial v^2}\beta^2 v + \frac{\partial^2 V}{\partial S \partial v}S\beta v\rho - rV = \alpha\frac{\partial V}{\partial v}. \tag{505}$$

### 5.7.1 Pricing with the characteristic function

The price of a European call can be expressed as an integral transform of the characteristic function for log-stock returns,

$$c(t) = S(t)e^{-\delta(T-t)}\Pi_1 - Ke^{-r(T-t)}\Pi_2 \tag{506}$$

$$\Pi_1 = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty d\omega\,\mathrm{Re}\left[\frac{e^{-i\omega \log K}\phi(\omega - i)}{i\omega\phi(-i)}\right] \tag{507}$$

$$\Pi_2 = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty d\omega\,\mathrm{Re}\left[\frac{e^{-i\omega \log K}\phi(\omega)}{i\omega}\right] \tag{508}$$

$$\phi(\omega) = \mathbb{E}_t^{Q_B}[e^{i\omega \log S(T)}], \tag{509}$$

The most common stochastic volatility model is the Heston model in which the variance rate is assumed to be a mean-reverting process with constant volatility of volatility,

$$\alpha(S(t), v(t), t) = \kappa[\bar{v} - v(t)], \qquad \beta(S(t), v(t), t) = \sigma_v. \tag{510}$$

Analytical expressions for the characteristic function are available in Heston model, which allows the integrals defining the call price to be performed numerically,

$$\phi(u) = e^{iu[\log S(t)+(r-\delta)(T-t)]-\frac{1}{\sigma_v^2}[\bar{v}\kappa(a(T-t)+2\log\beta)+v_0\gamma]} \tag{511}$$

$$\beta = \frac{1 - ge^{-d(T-t)}}{1 - g} \tag{512}$$

$$\gamma = \frac{a(1 - e^{-d(T-t)})}{1 - ge^{-d(T-t)}} \tag{513}$$

$$d = \sqrt{(i\rho\sigma_v u - \kappa)^2 + \sigma_v^2(iu + u^2)} \tag{514}$$

$$g = a/b \tag{515}$$

$$a = i\rho\sigma_v u - \kappa + d \tag{516}$$

$$b = i\rho\sigma_v u - \kappa - d \tag{517}$$

## 5.8  Jump-diffusion models

The Merton model assumes that the stock price evolves according geometric Brownian motion with jumps,

$$dS(t) = \hat{\mu}(t)S(t-)dt + S(t-)\sigma d\hat{w}(t) + S(t-)dJ(t), \tag{518}$$

where $dJ(t)$ is a jump process with intensity $\lambda$ and with jump size distributions given by

$$\log(1 + J_i) \sim \mathcal{N}(\mu_J, \sigma_J^2). \tag{519}$$

It can be shown that

$$\hat{\mu}(t) = r - \delta - \lambda\left(e^{\mu_J + \frac{1}{2}\sigma_J^2} - 1\right). \tag{520}$$

The affine stochastic volatility jump (ASVJ) model is the Heston specification with jumps,

$$dS(t) = \hat{\mu}(t)S(t-)dt + \sqrt{v(t)}S(t-)d\hat{w}(t) + S(t-)dJ(t), \tag{521}$$

$$dv(t) = \kappa[\bar{v} - v(t)]dt + \sigma_v\sqrt{v(t)}d\hat{w}_\sigma(t). \tag{522}$$

## 5.9  American derivatives

The Black-Scholes PDE expressed in terms of the coordinate $x = \log\big(S(t)/S(0)\big)$ is given by

$$\frac{\partial V}{\partial t} + \left(r - \delta - \frac{1}{2}\sigma^2\right)\frac{\partial V}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2 V}{\partial x^2} - rV = 0. \tag{523}$$

We will work on a discretized grid with coordinates $t_n = n\Delta t$ with $0 \le n \le N$ ($T = N\Delta t$) and $x_j = j\Delta x$ with $-M \le j \le M$ and will denote $S_j = S(0)e^{x_j}$ and $V_n^j = V(S_j, t_n)$. Since we are solving a final-value problem, it is convenient to approximate the time derivative using the so-called forward difference operator, which relates the current value at time $t_n$ to the value one time step in the future $t_{n+1}$,

$$\frac{\partial V}{\partial t} \simeq \frac{V_{n+1}^j - V_n^j}{\Delta t}. \tag{524}$$

In the $x$-direction, we choose to approximate the derivatives by symmetric difference operators as follows

$$\frac{\partial V}{\partial x} \simeq \frac{V_{n+1}^{j+1} - V_{n+1}^{j-1}}{2\Delta x}, \tag{525}$$

$$\frac{\partial^2 V}{\partial x^2} \simeq \frac{V_{n+1}^{j+1} + V_{n+1}^{j-1} - 2V_{n+1}^{j}}{(\Delta x)^2}. \tag{526}$$

The partial derivatives with respect to $x$ are evaluated at time $t_{n+1}$, which is convenient because it means that the resulting linear system expresses $V_n^j$ in terms of information purely at time $t_{n+1}$. If we evaluate the discount term at time $t_n$, then substituting these approximations into (523) and solving the linear system for $V_n^j$ we obtain

$$V_n^j = \mathrm{CV}(S_j, t_n), \tag{527}$$

where

$$\mathrm{CV}(S_j, t_n) = \frac{1}{1 + r\Delta t}\left[p_{\mathrm{U}} V_{n+1}^{j+1} + p_{\mathrm{M}} V_{n+1}^{j} + p_{\mathrm{D}} V_{n+1}^{j-1}\right] \tag{528}$$

is the continuation (or holding) value and the 'probabilities' are given by

$$p_{\mathrm{U}} = \frac{\sigma^2 \Delta t}{2\Delta x^2} + \frac{\mu \Delta t}{2\Delta x}, \qquad p_{\mathrm{D}} = \frac{\sigma^2 \Delta t}{2\Delta x^2} - \frac{\mu \Delta t}{2\Delta x}, \qquad p_{\mathrm{M}} = 1 - \frac{\sigma^2 \Delta t}{\Delta x^2}, \tag{529}$$

where $\mu = r - \delta - \frac{1}{2}\sigma^2$. This is called the trinomial tree scheme because the expression for $V_n^j$ has the form of a discounted expectation value. It can be shown that convergence requires

$$\Delta x > \sigma\sqrt{\Delta t}, \qquad \Delta x > |\mu|\sqrt{\Delta t}. \tag{530}$$

Since we are working on a finite grid we also need to impose boundary conditions at $x_M$ and $x_{-M}$ for each $t_n$ with $0 \le n \le N-1$. We define the endpoint values $V_n^M$ and $V_n^{-M}$ by solving the following systems

$$0 = V_n^M + V_n^{M-2} - 2V_n^{M-1}, \tag{531}$$

$$0 = V_n^{-M} + V_n^{-M+2} - 2V_n^{-M+1}. \tag{532}$$

In the continuum limit these are equivalent to imposing the vanishing of the second partial derivative with respect to $x$ at the boundaries. An American option gives the right to exercise at time $t_n$ or to continue. For American options we therefore simply replace (527) by

$$V_n^j = \max\left[\mathrm{EV}(S_j, t_n), \mathrm{CV}(S_j, t_n)\right] \tag{533}$$

where $\mathrm{EV}(S_j, t_n)$ is the intrinsic (or exercise) value of the option at time $t_n$ as a function of the stock price $S_j$.
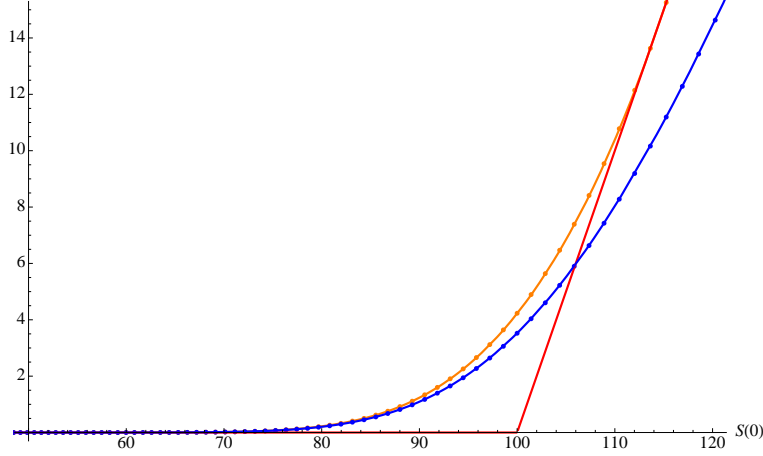
Figure 3: The present value of a European/American call (blue/orange) computed in the trinomial tree scheme. The call has expiry $T = 1$ and strike $K = 100$ on an underlying with $S(0) = 100$, dividend yield $\delta = 0.11$, volatility $\sigma = 0.16$ assuming risk-free rate $r = 0.05$. The finite-difference grid is chosen with $\Delta x = 0.142$, $M = 79$ and $\Delta t = 1/252$ so $N = T/\Delta t = 252$.

# 6 Fixed income derivatives

## 6.1 Swaps

Recall that $r(t, T)$ is the rate at which an investment of $P(t, T)$ at time $t$ accrues to yield a unit amount of currency at time $T$ assuming continuous compounding $P(t, T)e^{r(t,T)(T-t)} = 1$. Let $R(t, T)$ denote the simply compounded zero-coupon rate

$$R(t, T) \equiv \frac{1 - P(t, T)}{(T - t)P(t, T)}. \tag{534}$$

This is the rate at which an investment of $P(t, T)$ at time $t$ accrues to yield a unit amount of currency at time $T$, assuming that the accruing is proportional to the investment time; that is,

$$[1 + R(t, T)(T - t)] P(t, T) = 1. \tag{535}$$

Suppose that a company takes out a loan of $N$ units of currency at a variable interest rate $R(t, T)$. At the end of each interval $[T_{i-1}, T_i]$ (assumed to be of length $\tau$), they must pay an amount $NR(T_{i-1}, T_i)\tau$ to the lender. A $T_{i-1} \times T_i$ forward rate agreement is a forward contract which allows the borrower to lock in a fixed interest rate $K$ during the interval $[T_{i-1}, T_i]$. The pay-off of the FRA at time $T_i$ from the borrower's perspective is $N(K - R(T_{i-1}, T_i))\tau$. In other words, the company pays a fixed rate $K$ in exchange for the variable rate $R(T_{i-1}, T_i)$. Taking into account the payment to the lender we see that the company is effectively paying a fixed rate $K$ on the notional amount $N$. The cash flows for the borrower are given by[2]

$$\underbrace{-NR(T_{i-1}, T_i)\tau}_{\text{payment to lender}} + \underbrace{N(R(T_{i-1}, T_i) - K)\tau}_{\text{FRA pay-off}} = -NK\tau \tag{536}$$

---

[2]In this section we take cash inflows to be positive.

Since $R(T_{i-1}, T_i)$ is known at the fixing date $T_{i-1}$, the pay-off of a FRA on a unit notional at the maturity date $T_i$ is equivalent to a pay-off at time $T_{i-1}$ of

$$\mathbf{FRA}(T_{i-1}, T_i, K) = P(T_{i-1}, T_i)\left[R(T_{i-1}, T_i) - K\right]\tau \tag{537}$$

This is the amount that changes hands if the contract is settled at the fixing date $T_{i-1}$. In terms of the ZCB prices we find

$$\mathbf{FRA}(T_{i-1}, T_i, K) = P(T_{i-1}, T_{i-1}) - (1 + K\tau)P(T_{i-1}, T_i) \tag{538}$$

This shows that the FRA can be replicated by a portfolio long a ZCB with maturity $T_{i-1}$ and short $1 + K\tau$ ZCBs with maturity $T_i$, so for any $t \leq T_{i-1}$

$$\mathbf{FRA}(t, T_i, K) = P(t, T_{i-1}) - (1 + K\tau)P(t, T_i). \tag{539}$$

The fair value of the fixed interest rate $K$ decided on the contract date $t = 0$ is the one that renders the value of the FRA worthless at $t = 0$; that is, $K = F(0, T_{i-1}, T_i)$ where

$$F(t, T_{i-1}, T_i) \equiv \frac{P(t, T_{i-1}) - P(t, T_i)}{\tau P(t, T_i)} \tag{540}$$

is the simply compounded forward rate agreed upon at time $t$ for borrowing in the interval $[T_{i-1}, T_i]$. It then follows that the time-$t$ price of a $[T_{i-1}, T_i]$ forward rate agreement is given by

$$\mathbf{FRA}(t, T_i, K) = P(t, T_i)\left[F(t, T_{i-1}, T_i) - K\right]\tau, \tag{541}$$

which clearly satisfies

$$\mathbf{FRA}(0, T_i, K) = 0. \tag{542}$$

On the other hand, we may use the FTAP to price the FRA at any time $t \leq T_{i-1}$ as follows

$$\mathbf{FRA}(t, T_i, K) = N(t)\mathbb{E}_t^{Q_N}\left[\frac{\mathbf{FRA}(T_{i-1}, T_i, K)}{N(T_{i-1})}\right] \tag{543}$$

Choosing $N(t) = P(t, T_i)$ we obtain

$$\mathbf{FRA}(t, T_i, K) = P(t, T_i)\mathbb{E}_t^{Q_i}\left[\frac{P(T_{i-1}, T_i)\left[R(T_{i-1}, T_i) - K\right]}{P(T_{i-1}, T_i)}\right]\tau \tag{544}$$

$$= P(t, T_i)\left\{\mathbb{E}_t^{Q_i}\left[R(T_{i-1}, T_i)\right] - K\right\}\tau. \tag{545}$$

Comparing with (541) we obtain

$$F(t, T_{i-1}, T_i) = \mathbb{E}_t^{Q_i}\left[R(T_{i-1}, T_i)\right]. \tag{546}$$

Now using the fact that $F(T_{i-1}, T_{i-1}, T_i) = R(T_{i-1}, T_i)$ we discover the Martingale property of forward rates,

$$F(t, T_{i-1}, T_i) = \mathbb{E}_t^{Q_i}\left[F(T_{i-1}, T_{i-1}, T_i)\right] \tag{547}$$

or

$$F_i(t) = \mathbb{E}_t^{Q_i}\left[F(T_{i-1})\right] \tag{548}$$

where $F_i(t) \equiv F(t, T_{i-1}, T_i)$. An interest rate swap is just a sum of forward rate agreements. Let $T_n$ denote the date of the last settlement. Then

$$\mathbf{IRS}(t, T_j, T_n, K) = \sum_{i=j+1}^{n} \mathbf{FRA}(t, T_i, K). \tag{549}$$

If $j = 0$ then the interest-rate swap is referred to as a spot IRS whereas if $j > 0$ then it is a forward-start IRS. The fixed rate for an IRS is determined by the condition that the value of the IRS at inception vanishes: $\mathbf{IRS}(0, T_j, T_n, K) = 0$. Representing the interest-rate swap in terms of zero-coupon bonds using (539) we obtain

$$\mathbf{IRS}(t, T_j, T_n, K) = P(t, T_j) - P(t, T_n) - K\tau A(t, T_j, T_n) \tag{550}$$

where

$$A(t, T_j, T_n) = \sum_{i=j+1}^{n} P(t, T_i) \tag{551}$$

is the value of an annuity paying one unit at dates $\{T_{j+1}, T_{j+2}, \ldots, T_n\}$. It follows that $K = \mathrm{FSR}(0, T_j, T_n)$ where the forward swap rate is defined as

$$\mathrm{FSR}(t, T_j, T_n) \equiv \frac{P(t, T_j) - P(t, T_n)}{\tau A(t, T_j, T_n)}. \tag{552}$$

Therefore

$$\mathbf{IRS}(t, T_j, T_n, K) = A(t, T_j, T_n)\left[\mathrm{FSR}(t, T_j, T_n) - K\right]\tau. \tag{553}$$

## 6.2  Yield curve construction

The goal of yield curve construction is to deduce the zero-coupon yield curve today $P(0, T)$ from the prevailing prices of financial instruments such as zero-coupon bonds, forward-rate agreements, and interest rate swaps.

The first step of yield curve construction is to compute the discount factors associated with the deposit accounts. These are the easiest instruments because their prices are algebraically related to the discount factor. In realistic deposits one must distinguish between today's trade date ($t = 0$) and the settlement date $t_s > 0$ at which actually money changes hands. For a deposit with trade date $t = 0$ settlement date $t_s$ and expiry date $T_E$ the discount factor $P(t_s, T_E)$ can be extracted from the quoted Libor rate $R(t_s, T_E)$ as follows

$$P(t_s, T_E) = \frac{1}{1 + R(t_s, T_E)(T_E - t_s)}. \tag{554}$$

Since we are ultimately interested in $P(0, T_E)$, however, we need to compute the additional discount factor $P(0, t_s)$ from which we compute the desired quantity; namely,

$$P(0, T_E) = P(0, t_s)P(t_s, T_E). \tag{555}$$

Fortunately, it is possible to determine $P(0, t_s)$ by compounding the so-called overnight (O/N) and tomorrow/next (T/N) rates.

The next class of instruments we consider are the forward rate agreements. Recall that break-even forward rate for borrowing in the period $[T_{i-1}, T_i]$ is related to the discount factors as follows,

$$F(0, T_{i-1}, T_i) = \frac{1}{\tau}\left[\frac{P(0, T_{i-1})}{P(0, T_i)} - 1\right]. \tag{556}$$

Rearranging we obtain the discount factor for expiration $T_i$ in terms of the Libor forward rate for borrowing in the period $[T_{i-1}, T_i]$ and the discount factor for expiration $T_{i-1}$,

$$P(0, T_i) = \frac{P(0, T_{i-1})}{1 + \tau F(0, T_{i-1}, T_i)}. \tag{557}$$

The discount factor for expiration $T_i$ can now be obtained recursively from this expression by interpolating the existing discount curve to estimate $P(0, T_{i-1})$.

## 6.3  Bonds

The price at time $t$ of a non-defaultable bond paying cash flows $C_i + \delta_{ni}$ at dates $\{T_1, \ldots, T_n\}$ is given by,

$$B(t) = \sum_{T_i > t}^{T_n} (C_i + \delta_{ni}) P(t, T_i). \tag{558}$$

The fact that the bond price suffers discontinuities at the coupon dates has motivated traders to introduce the concept of accrued interest at time $t \in (T_{i-1}, T_i]$,

$$\text{AI}(t) = C_i \frac{t - T_{i-1}}{T_i - T_{i-1}}. \tag{559}$$

The quoted prices are given by $B_{\text{clean}}(t) = B(t) - \text{AI}(t)$. The clean price is continuous but its first derivative is discontinuous.

The continuously compounded yield to maturity for a bond is defined as the unique solution to,

$$B(t) = \sum_{T_i > t}^{T_n} (C_i + \delta_{ni}) e^{-y(t)(T_i - t)}. \tag{560}$$

In the case of a non-defaultable bond we can find $y(t)$ in terms of the discount curve by solving,

$$\sum_{T_i > t}^{T_n} (C_i + \delta_{ni}) P(t, T_i) = \sum_{T_i > t}^{T_n} (C_i + \delta_{ni}) e^{-y(t)(T_i - t)}. \tag{561}$$

Realistic bond yields do not satisfy the above identity as a result of factors such as default risk. It is therefore of general interest to consider the change in price of a bond induced by shifting the yield curve. For this purpose we define the Macaulay duration as

$$\frac{dB(t)}{dy(t)} = -D_{\text{M}}(t) B(t) \tag{562}$$

or

$$D_{\mathrm{M}}(t) \equiv -\frac{1}{B(t)}\frac{dB(t)}{dy(t)} = \frac{1}{B(t)}\sum_{T_i>t}^{T_n}(C_i + \delta_{ni})(T_i - t)e^{-y(t)(T_i-t)}.$$

(563)

It is also interesting to consider the change in price induced by parallel shifts of the zero-coupon yield curve,

$$B(t;s) = \sum_{T_i>t}^{T_n}(C_i + \delta_{ni})e^{-(r(t,T_i)+s)(T_i-t)}$$

(564)

and hence we define the Fisher-Weil duration as

$$D_{\mathrm{FW}}(t) \equiv -\frac{1}{B(t)}\frac{d}{ds}B(t;s)\bigg|_{s=0} = \frac{1}{B(t)}\sum_{T_i>t}^{T_n}(C_i + \delta_{ni})(T_i - t)P(t,T_i).$$

(565)

## 6.4 Caps and floors

Interest rate caps and floors are defined as

$$\mathbf{Cap}(t,T_j,T_n,K) = \sum_{i=j+1}^{n}\mathbf{Cpt}(t,T_i,K), \quad \mathbf{Flr}(t,T_j,T_n,K) = \sum_{i=j+1}^{n}\mathbf{Flt}(t,T_i,K)$$

(566)

where $\mathbf{Cpt}(t,T_i,K)$ is the price of a call option on the interest rate $R(T_{i-1},T_i)$ with strike $K$ and expiration $T_i$ while $\mathbf{Flt}(t,T_i,K)$ is the corresponding put option. The pay-offs at the settlement dates $T_i$ are

$$\mathbf{Cpt}(T_i,T_i,K) = \max(R(T_{i-1},T_i) - K, 0)\tau$$

(567)

$$\mathbf{Flt}(T_i,T_i,K) = \max(K - R(T_{i-1},T_i), 0)\tau.$$

(568)

Notice that since $R(T_{i-1},T_i) = F(T_{i-1},T_{i-1},T_i)$ a caplet can be viewed as a call option on forward rate. Since the forward rate $F(t,T_{i-1},T_i)$ is a martingale with respect to $Q_i$, it follows that we can price a caplet by applying the fundamental theorem of asset pricing with $P(t,T_i)$ as the numeraire,

$$\mathbf{Cpt}(t,T_i,K) = P(t,T_i)\mathbb{E}_t^{Q_i}\left[(F_i(T_{i-1}) - K)^+\right]\tau.$$

(569)

Starting from the identity

$$\mathbf{Cpt}(T_i,T_i,K) + K\tau = \mathbf{Flt}(T_i,T_i,K) + F(T_{i-1},T_{i-1},T_i)\tau$$

(570)

and applying the FTAP with $P(t,T_i)$ as the numeraire we obtain the cap-floor parity relation

$$\mathbf{Cpt}(t,T_i,K) - \mathbf{Flt}(t,T_i,K) = \mathbf{FRA}(t,T_i,K).$$

(571)

If the forward rate follows an Ito process, then the martingale condition implies

$$dF_i(t) = F_i(t)\sigma_i(t)d\hat{w}_i(t)$$

(572)

where $\hat{w}_i$ is a Brownian motion under $Q_i$. In the case when $\sigma_i(t)$ is deterministic, the usual Black formulas can be applied to give

$$\mathbf{Cpt}_{\mathrm{B}}(t, T_i, K) = P(t, T_i) \left[ F_i(t) N(y_i + \bar{\sigma}_i \sqrt{T_{i-1} - t}) - K N(y_i) \right] \tau \tag{573}$$

where

$$y_i = \frac{\log\left(\frac{F_i(t)}{K}\right) - \frac{1}{2}\bar{\sigma}_i^2(T_{i-1} - t)}{\bar{\sigma}_i \sqrt{T_{i-1} - t}} \tag{574}$$

$$\bar{\sigma}_i^2 = \frac{1}{T_{i-1} - t} \int_t^{T_{i-1}} ds\, \sigma_i(s)^2 \tag{575}$$

Cap prices are reported as cap implied volatilities which are defined by matching the Black cap pricing formula with equal-volatility component caplets to the cap market price,

$$\mathbf{Cap}_{\mathrm{M}}(t, T_j, T_n, K) = \sum_{i=j+1}^{n} \mathbf{Cpt}_{\mathrm{B}}(t, T_i, K; \sigma). \tag{576}$$

An alternative way to price caplets and floorlets is to view them as options on bonds. We begin by noting that since $R(T_{i-1}, T_i)$ is known on the fixing date $T_{i-1}$, the pay-off of a caplet on the settlement date $T_i$ is equivalent to a pay-off on the fixing date of

$$\mathbf{Cpt}(T_{i-1}, T_i, K) = P(T_{i-1}, T_i)(R(T_{i-1}, T_i) - K)^+ \tau \tag{577}$$

$$= [1 - P(T_{i-1}, T_i)(1 + K\tau)]^+ \tag{578}$$

$$= (1 + K\tau)\left(\frac{1}{1 + K\tau} - P(T_{i-1}, T_i)\right)^+. \tag{579}$$

It follows that a caplet can be viewed as $1 + K\tau$ put options with strike $1/(1 + K\tau)$ and expiration $T_{i-1}$ on a ZCB with maturity $T_i$;

$$\mathbf{Cpt}(t, T_i, K) = (1 + K\tau)\mathbf{PZCB}(t, T_{i-1}, T_i, 1/(1 + K\tau)) \tag{580}$$

$$\mathbf{Flt}(t, T_i, K) = (1 + K\tau)\mathbf{CZCB}(t, T_{i-1}, T_i, 1/(1 + K\tau)) \tag{581}$$

where $\mathbf{CZCB}(t, T_i, T_j, K)$ is the price of a call option with strike $K$ and expiration $T_i$ on a ZCB with expiration $T_j$.

## 6.5   Spot-rate models

A spot-rate model is based on the idea that the prices of zero-coupon bonds can be expressed in terms of the instantaneous spot rate as follows

$$P(t, T) = \mathbb{E}_t^{Q_B}\left[ e^{-\int_t^T ds\, r(s)} \right]. \tag{582}$$

It follows that specifying the stochastic process for the instantaneous spot rate $r(t)$ is sufficient to reproduce the entire term structure of interest rates. A one-factor spot-rate model assumes that the spot rate follows a diffusion

$$dr(t) = \mu(r(t), t)dt + \sigma(r(t), t)d\hat{w}(t) \tag{583}$$

where $\hat{w}$ is a Brownian motion under $Q_B$. The assumption of a diffusion implies that $P(t,T)$ only depends on $r(t)$ and $t$. The dynamics of the process $r(t)$ should be chosen such that the initial term structure of interest rates matches the market data

$$P(0,T) = P_{\mathrm{M}}(0,T). \tag{584}$$

The pricing PDE for a zero-coupon bond can be obtained by using the fact that the discounted price process $P(t,T)/B(t)$ is a Martingale under $Q_B$,

$$\frac{P(t,T)}{B(t)} = \mathbb{E}_t^{Q_B}\left[\frac{P(T,T)}{B(T)}\right]. \tag{585}$$

Computing the differential of the left-hand side we obtain

$$d\left(\frac{P(t,T)}{B(t)}\right) = \frac{1}{B(t)}\left[dP(t,T) - P(t,T)r(t)dt\right] \tag{586}$$

where we have used Leibniz's rule

$$\frac{d}{dt}\left(\frac{1}{B(t)}\right) = \frac{d}{dt}\left(e^{-\int_0^t ds\, r(s)}\right) = -e^{-\int_0^t ds\, r(s)}\frac{d}{dt}\int_0^t ds\, r(s) = -\frac{r(t)}{B(t)}. \tag{587}$$

Applying Ito's lemma to $dP(t,T)$ using (583) and writing $P(t,T) = g(r(t),t)$ we arrive at the partial differential equation

$$\frac{\partial g}{\partial t} + \mu(r,t)\frac{\partial g}{\partial r} + \frac{1}{2}\sigma(r,t)^2\frac{\partial^2 g}{\partial r^2} - g(r,t)r = 0. \tag{588}$$

The boundary condition suitable for a zero-coupon bond is $g(r,T) = 1$. Alternatively we can apply the Feynman-Kac formula directly to (582).

### 6.5.1 Extended Vasicek model

The Hull-White (or extended Vasicek) model assumes that the spot rate follows an Ornstein-Uhlenbeck process

$$dr(t) = \kappa(\theta(t) - r(t))dt + \sigma d\hat{w}(t), \tag{589}$$

where $\hat{w}$ is a Brownian motion under $Q_B$. The solution is $r(t) = x(t) + \varphi(t)$ where

$$\varphi(t) = \kappa\int_0^t ds\, e^{-\kappa(t-s)}\theta(s) \tag{590}$$

$$x(t) = r(0)e^{-\kappa t} + \sigma\int_0^t e^{-\kappa(t-s)}d\hat{w}(s). \tag{591}$$

It is convenient to reformulate the model in terms of $\varphi(t)$ and a stochastic process $x(t)$ such that

$$r(t) = \varphi(t) + x(t), \tag{592}$$
$$dx(t) = -\kappa x(t)dt + \sigma d\hat{w}(t), \tag{593}$$
$$x(0) = r(0) \tag{594}$$

where $\hat{w}$ is a Brownian motion under $Q_B$. The price of a zero-coupon bond is given in this formulation by

$$P(t, T) = e^{-\int_t^T ds\, \varphi(s)} \mathbb{E}_t^{Q_B}\left[ e^{-\int_t^T ds\, x(s)} \right] \tag{595}$$

$$= e^{-\int_t^T ds\, \varphi(s)} e^{-a(T-t)-b(T-t)x(t)} \tag{596}$$

where

$$a(\tau) = \left[ \frac{b(\tau) - \tau}{\kappa} + \frac{b(\tau)^2}{2} \right] \frac{\sigma^2}{2\kappa} \tag{597}$$

$$b(\tau) = \frac{1}{\kappa}(1 - e^{-\kappa\tau}). \tag{598}$$

In the limit $\kappa \to 0$ we have $a(\tau) \to -(\sigma^2/6)\tau^3$ and $b(\tau) \to \tau$. Taking the differential of $P(t, T)$ we obtain

$$dP(t, T) = P(t, T)\left[ \varphi(t)dt - \frac{d}{dt}a(T-t)dt - x(t)\frac{d}{dt}b(T-t)dt - b(T-t)dx(t) + \frac{1}{2}b(T-t)^2\sigma^2 dt \right] \tag{599}$$

where the last term follows from Ito's lemma. Using the expressions for $a(t)$ and $b(\tau)$ we find that $P(t, T)$ satisfied the SDE,

$$dP(t, T) = P(t, T)r(t)dt - P(t, T)b(T-t)\sigma d\hat{w}(t). \tag{600}$$

We see that the proportional volatility $-b(T-t)\sigma$ decays monotonically to zero as $t$ approaches the maturity date $T$.

The fitting function $\varphi(t)$ is chosen so that $P(0, t)$ matches the initial term structure of interest rates $P_M(0, t)$ on the calibration date,

$$e^{-\int_0^t ds\, \varphi(s)} = e^{a(t)+b(t)r(0)} P_M(0, t), \tag{601}$$

where we have used $x(0) = r(0)$. Using this relation to eliminate the dependence on the fitting function we obtain

$$P(t, T) = A(t, T)e^{-b(T-t)x(t)} \tag{602}$$

where

$$A(t, T) = \frac{e^{a(T)+b(T)r(0)} P_M(0, T)}{e^{a(t)+b(t)r(0)} P_M(0, t)} e^{-a(T-t)}. \tag{603}$$

We have thus expressed the prices of ZCBs in terms of the initial term structure $P_M(0, T)$, the prevailing spot rate $r(0)$ and two calibration parameters; $\kappa$ and $\sigma$. The expression (602) implies that ZCB prices are log-normally distributed. This allows puts and calls on ZCBs to be priced analytically as in the Black model.

Let us consider a European call with expiration $T_i$ on a ZCB with maturity $T_j > T_i$. Since the pay-off occurs at time $T_i$ it is natural to use $P(t, T_i)$ as the numeraire. Defining $G(t) = P(t, T_j)/P(t, T_i)$ we find

$$dG(t) = -G(t)\left[ b(T_j - s) - b(T_i - s) \right] \sigma\, d\hat{w}_i(t) \tag{604}$$

where $\hat{w}_i$ is a Brownian motion under $Q_i$. The prices of puts and calls on $P(t, T_j)$ can now be obtained by applying the Black formulas,

$$\mathbf{CZCB}_{\mathrm{EV}}(t, T_i, T_j, K) = P(t, T_i)\left[G(t)N(y + \bar{\sigma}\sqrt{T_i - t}) - KN(y)\right] \tag{605}$$

$$\mathbf{PZCB}_{\mathrm{EV}}(t, T_i, T_j, K) = P(t, T_i)\left[KN(-y) - G(t)N(-y - \bar{\sigma}\sqrt{T_i - t})\right]. \tag{606}$$

where

$$y = \frac{\log\left(\frac{G(t)}{K}\right) - \frac{1}{2}\bar{\sigma}^2(T_i - t)}{\bar{\sigma}\sqrt{T_i - t}} \tag{607}$$

$$\bar{\sigma}^2 = \frac{1}{T_i - t}\int_t^{T_i} ds\, \sigma^2\left[b(T_j - s) - b(T_i - s)\right]^2, \tag{608}$$

$$= \frac{b(T_j - T_i)^2 b(2(T_i - t))}{2(T_i - t)}\sigma^2, \tag{609}$$

Thus we have expressed the price at time zero of a European call or put in terms of the initial discount curve $P(0, T)$, the prevailing level of the spot rate $r(0)$ and two calibration parameters $\sigma$ and $\kappa$, and we didn't need to make any assumptions about the arbitrary fitting function.

Now we consider pricing a caplet using Monte-Carlo simulation. We start by expressing the pay-off for a caplet at the settlement date $T_i$ in terms of the price of a zero-coupon bond,

$$\mathbf{Cpt}(T_i, T_i, K) = (R(T_{i-1}, T_i) - K)^+ \tau \tag{610}$$

$$= \left(\frac{1 - P(T_{i-1}, T_i)}{P(T_{i-1}, T_i)} - K\tau\right)^+. \tag{611}$$

The distribution of $P(T_{i-1}, T_i)$ under $Q_B$ is known to be given by (596). So it follows from the FTAP that the time $t$ price of the caplet is given by

$$\mathbf{Cpt}(t, T_i, K) = \mathbb{E}_t^{Q_B}\left[e^{-\int_t^{T_i} ds\, r(s)}\left(\frac{1 - P(T_{i-1}, T_i)}{P(T_{i-1}, T_i)} - K\tau\right)^+\right]. \tag{612}$$

This expectation can be computed by simulating $x$ under $Q_B$. It is more convenient, however, to use $P(t, T_i)$ as the numeraire so that the pricing formula becomes

$$\mathbf{Cpt}(t, T_i, K) = P(t, T_i)\mathbb{E}_t^{Q_i}\left[\left(\frac{1 - P(T_{i-1}, T_i)}{P(T_{i-1}, T_i)} - K\tau\right)^+\right]. \tag{613}$$

In order to compute this expectation we need the dynamics of $x$ under $Q_i$. Changing numeraire from $Q_B$ to $Q_i$ using Girsanov's theorem leads to the following SDE

$$dx(t) = -\left[\kappa x(t) + b(T_i - t)\sigma^2\right]dt + \sigma d\hat{w}_i(t), \tag{614}$$

$$x(0) = r(0), \tag{615}$$

where $\hat{w}_i$ is a Brownian motion under $Q_i$. The SDE has an explicit solution given by

$$x(t) = r(0)e^{-\kappa t} - \frac{\sigma^2}{2}\left[b(t)^2 + b(T_i - t)b(2t)\right] + \Delta x(t), \tag{616}$$

$$\Delta x(t) = \mathcal{N}\left(0, \frac{\sigma^2}{2}b(2t)\right). \tag{617}$$

The ZCB prices can now be simulated by drawing $\Delta x(t)$ from the distribution (617) and substituting the resulting values of $x(t)$ into (602).

### 6.5.2  Ho-Lee model

The Ho-Lee model assumes that the spot rate follows

$$dr(t) = \theta(t)dt + \sigma d\hat{w}(t) \tag{618}$$

where $\hat{w}$ is a Brownian motion under $Q_B$. This model can be obtained as a limit of the Extended Vasicek model when $\kappa \to 0$. It follows that the SDE for the bond price is given by

$$dP(t, T) = P(t, T)r(t)dt - P(t, T)(T - t)\sigma d\hat{w}(t). \tag{619}$$

It follows from the standard Black pricing formula that the price of a European call is given by

$$\mathbf{CZCB}(t, T_i, T_j, K) = P(t, T_i)\mathbb{E}_t^{Q_i}\left[\max(G(T_i) - K, 0)\right] \tag{620}$$

$$= P(t, T_i)\left[G(t)N(y + \bar{\sigma}\sqrt{T_i - t}) - KN(y)\right] \tag{621}$$

where

$$y = \frac{\log\left(\frac{G(0)}{K}\right) - \frac{1}{2}\bar{\sigma}^2(T_i - t)}{\bar{\sigma}\sqrt{T_i - t}}, \qquad \bar{\sigma} = \sigma(T_j - T_i). \tag{622}$$

### 6.5.3  Cox-Ingersoll-Ross model

A disadvantage of the Ho-Lee and extended Vasicek models is that interest rates can become negative with positive probability. This problem can be overcome by assuming that the instantaneous interest rate follows a square-root diffusion (SRD) process. A time dependent parameter can be introduced by writing $r(t) = x(x) + \varphi(t)$ where

$$dx(t) = \kappa(\theta - x(t))dt + \sigma\sqrt{x(t)}d\hat{w}(t). \tag{623}$$

This is not equivalent to assuming that $r(t)$ follows an SRD with a time-dependent $\theta$. The prices of ZCBs are given by

$$P(t, T) = A(t, T)e^{-b(T-t)x(t)} \tag{624}$$

where

$$A(t,T) = \frac{e^{a(T)+b(T)x(0)} P_{\mathrm{M}}(0,T)}{e^{a(t)+b(t)x(0)} P_{\mathrm{M}}(0,t)} e^{-a(T-t)} \tag{625}$$

$$a(\tau) = -\frac{2\kappa\theta}{\sigma^2} \left[ \frac{(\kappa+\gamma)\tau}{2} + \log\frac{2\gamma}{c(\tau)} \right] \tag{626}$$

$$b(\tau) = \frac{2(e^{\gamma\tau}-1)}{c(\tau)} \tag{627}$$

$$c(\tau) = (\kappa+\gamma)(e^{\gamma\tau}-1) + 2\gamma \tag{628}$$

$$\gamma = \sqrt{\kappa^2 + 2\sigma^2}. \tag{629}$$

This expresses the prices of ZCBs in terms of the initial term structure $P_{\mathrm{M}}(0,T)$ and four calibration parameters; $\kappa$, $\theta$ and $\sigma$ and $x(0)$.

### 6.5.4 Jamshidian trick

We can price a call option on a coupon bond by thinking of it as sum of call options on the constituent zero-coupon bonds with different strikes. The pay-off from a call option with exercise date $T_s < T_n$, written on a bond which pays coupons $c_i$ at dates $\{T_1, \ldots, T_n\}$ is given by

$$\max\left[ \sum_{T_i > T_s}^{T_n} c_i P(T_s, T_i) - K, 0 \right]. \tag{630}$$

Notice that we only include the cash flows which occur after the exercise date $T_s$ of the option. Using the Jamshidian trick we can re-write this as the pay-off for a portfolio of calls on zero-coupon bonds[3]

$$\max\left[ \sum_{T_i > T_s}^{T_n} c_i P(T_s, T_i) - K, 0 \right] = \sum_{T_i > T_s}^{T_n} c_i \max\left[ P(T_s, T_i) - K_i, 0 \right]. \tag{631}$$

The strikes $K_i$ on the zero-coupon bonds are given by

$$K_i = P(r_K; T_s, T_i) \tag{632}$$

where $r_K$ is the level of the interest rate at time $T_s$ such that

$$\sum_{T_i > T_s}^{T_n} c_i P(r_K; T_s, T_i) = K. \tag{633}$$

## 6.6 Forward rate models

In spot rate models an arbitrary fitting function was required in order to match the initial term structure of interest rates. An alternative approach is to model the term structure

---

[3]This only works for one-factor models because the bond price is a monotonic decreasing function of the interest rate.

directly and impose absence of arbitrage as an additional constraint. The continuously compounded forward rate is the rate $f(t, T, T + \tau)$ agreed upon at time $t$ for borrowing in the future in the interval $[T, T + \tau]$. Absence of arbitrage implies that the forward rate must satisfy

$$P(t, T + \tau) = P(t, T)e^{-\tau f(t,T,T+\tau)} \tag{634}$$

or

$$f(t, T, T + \tau) = \frac{\log P(t, T) - \log P(t, T + \tau)}{\tau}. \tag{635}$$

In the limit $\tau \to 0$ we obtain the instantaneous forward rate,

$$\lim_{\tau \to 0} f(t, T, T + \tau) \equiv f(t, T) = -\frac{\partial}{\partial T} \log P(t, T) \tag{636}$$

and thus

$$P(t, T) = e^{-\int_t^T ds\, f(t,s)}. \tag{637}$$

Comparing with the expression in terms of the instantaneous interest rate we find the relationship

$$f(t, t) = r(t). \tag{638}$$

The Heath-Jarrow-Morton framework assumes that $f(t, T)$ follows an Ito process

$$df(t, T) = \alpha(t, T)dt + \sum_{i=1}^{d} \sigma_i(t, T)d\hat{w}_i(t) \tag{639}$$

where $\hat{w}$ is a $d$-dimensional Brownian motion under the risk-neutral probability measure. Demanding that the term structure is free of arbitrage implies the following constraint on the drift,

$$\alpha(t, T) = \sum_{i=1}^{d} \sigma_i(t, T) \int_t^T ds\, \sigma_i(t, s). \tag{640}$$

*Proof.* Using Ito's lemma we can determine the dynamics of the zero-coupon discount curve in terms of the forward rate. Defining $Y(t, T) = -\int_t^T ds\, f(t, s)$ so that $P(t, T) = e^{Y(t)}$, an application of Ito's lemma gives

$$dP(t, T) = P(t, T)dY(t, T) + \frac{1}{2}P(t, T)dY(t, T)^2 \tag{641}$$

where summation over $i$ is implied. Applying Leibniz's rule we obtain

$$dY(t, T) = f(t, t)dt - \int_t^T ds\, [\alpha(t, s)dt + \sigma_i(t, s)d\hat{w}_i(t)] \tag{642}$$

$$= [r(t) - \bar{\alpha}(t, T)] - \bar{\sigma}_i(t, T)d\hat{w}_i(t) \tag{643}$$

where we have used $r(t) = f(t, t)$ and defined $\bar{\alpha}(t, T) = \int_t^T ds\, \alpha(t, s)$ and $\bar{\sigma}_i(t, T) = \int_t^T ds\, \sigma_i(t, s)$. Hence

$$dP(t, T) = P(t, T)\left[r(t) - \bar{\alpha}(t, T) + \frac{1}{2}\bar{\sigma}_i(t, T)\bar{\sigma}_i(t, T)\right]dt - P(t, T)\bar{\sigma}_i(t, T)d\hat{w}_i(t) \tag{644}$$

The no-arbitrage constraint is that $dP(t, T)$ must have a drift coefficient equal to $r(t)P(t, T)$ and thus

$$\bar{\alpha}(t, s) = \frac{1}{2} \sum_{i=1}^{d} \bar{\sigma}_i(t, T)^2. \tag{645}$$

Differentiating this expression with respect to $T$ gives the result. $\qquad\square$

# 7 Credit derivatives

## 7.1 Structural models

The Merton model assumes that a firm is financed by equity together with ZCBs with maturity $T$ and face value $K$. The firm is assumed to default at time $T$ if the asset value of the firm at time $T$ is insufficient to repay the outstanding debt; that is, if $V(T) < K$. The cash flow to the bond holders at time $T$ is either $V(T)$ or $K$ (whichever is smaller) and thus

$$D(T) = \min(V(T), K) = K - \max(K - V(T), 0). \tag{646}$$

The cash flow to the stakeholders in the firm is

$$E(T) = V(T) - D(T) = \max(V(T) - K, 0) \tag{647}$$

so that the standard accounting equation is satisfied

$$V(T) = D(T) + E(T). \tag{648}$$

If we assume that the asset value follows geometric Brownian motion with constant dividends, volatility and interest rates then the value of the debt at time $t$ is

$$D(t) = Ke^{-r(T-t)} + p(t) \tag{649}$$

where $p(t)$ is the price process for a European put with strike $K$ and maturity $T$. We define the credit spread by

$$D(t) = Ke^{-(r+s(t))(T-t)} \implies s(t) = -\frac{1}{T-t} \log \frac{D(t)}{K} - r. \tag{650}$$

Notice that if $K > V(t)$ then as $T - t \to 0$ the credit spread tends to infinity. This reflects the fact that it is not possible to redeem the bond if it is matured immediately. Conversely if $K < V(t)$ then there is no risk of default in the limit $T - t \to 0$ so $s(t) \to 0$. For longer maturities there is a finite probability that the asset value will grow to exceed the face value of the bond so $s(t) \to 0$. The Black-Cox model is an improvement on the Merton model in which default can be triggered by the asset value falling by below a level $H(t)$. In the case when $H(t) = Ke^{-r(T-t)}$ the debt is riskless because the bondholders receive the present value of the promised payment.
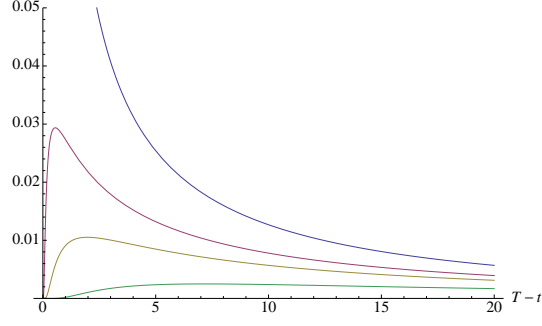
Figure 4: The credit spreads as a function of maturity for the Merton model assuming $r = 0.5\%$, $\delta = 0$ and $\sigma = 0.2$. The blue and green lines correspond to $K = 1.1V(t)$ and $K = 0.6V(t)$, respectively.

## 7.2 Defaultable bonds

The price of a zero-recovery, defaultable zero-coupon bond at time $t < T$ is given by

$$\mathcal{P}_0(t, T) = B(t)\mathbb{E}_t^{Q_B}\left[\frac{1_{\{\tau > T\}}}{B(T_n)}\right]. \tag{651}$$

In reality the most heavily traded bonds pay periodic coupons in addition to the notional value which is paid at maturity. A bullet bond is a coupon bond in which the coupon payments are a fixed fraction of the notional. Realistic bonds also pay a recovery amount in the event of default. Consider a defaultable bond with unit notional paying coupons $c_i$ at dates $\{T_1, \ldots, T_n\}$. Upon default, the bond is assumed to pay a random recovery rate $R$. The value of the bond at time $t < T_n$ is given by

$$B(t, T_n) = \sum_{T_i > t}^{T_n} c_i \mathcal{P}_0(t, T_i) + B(t)\mathbb{E}_t^{Q_B}\left[\frac{R1_{t < \tau \leq T_n}}{B(\tau)}\right]. \tag{652}$$

As the probability of default increases, the bond will tend to default immediately $\tau \to t$ and thus the last term approaches $\mathbb{E}_t^{Q_B}[R]$. It follows that the value of the bond is bounded from below by the recovery rate. If the recovery rate is independent of the time of default as well as the level of interest rates then

$$B(t, T_n) = \sum_{T_i > t}^{T_n} c_i \mathcal{P}_0(t, T_i) + \mathbb{E}_t^{Q_B}[R]\mathcal{D}(t, T_n) \tag{653}$$

where

$$\mathcal{D}(t, T_n) = B(t)\mathbb{E}_t^{Q_B}\left[\frac{1_{t < \tau \leq T_n}}{B(\tau)}\right] \tag{654}$$

is the value at time $t$ of a security that pays one unit of currency at time $\tau$ if $\tau \leq T_n$. If we moreover assume that the time of default is independent of the level of interest rates we obtain

$$\mathcal{D}(t, T_n) = \int_t^{T_n} ds\, P(t, s)\frac{\partial}{\partial s}G(t, s) \tag{655}$$

74

where

$$G(t, s) = Q_B(\tau \leq s \mid I_t) \tag{656}$$

is the default probability distribution. The survival curve is defined as

$$X(t, s) = Q_B(\tau > s \mid I_t) \tag{657}$$

## 7.3   Credit default swaps

In a single-name credit default swap the protection buyer pays quarterly coupons to the protection seller in exchange for a contingent payment if a credit event affecting the reference obligation occurs before the contract maturity. The contingent payment is equal to the difference between the par and recovery value and is called the loss given default

Consider a single-name CDS with tenor $T_n$, coupon dates $\{T_1, \ldots, T_n\}$ and unit notional. The value at time zero of the protection leg paid by the protection seller at is given by

$$V_{\text{prot}}(0) = B(0)\mathbb{E}^{Q_B}\left[\frac{(1 - R)1_{\{\tau \leq T_n\}}}{B(\tau)}\right] \tag{658}$$

where $R$ is the firm's (possibly random) recovery rate. In the case when the recovery rate is independent of the default time we have

$$V_{\text{prot}}(0) = \left(1 - \mathbb{E}^{Q_B}[R]\right)\mathcal{D}(0, T_n) \tag{659}$$

The value at time zero of the coupon stream paid by the protection buyer is given by

$$V_{\text{prem}}(0) = c\mathcal{A}(0, T_n) \tag{660}$$

where $\mathcal{A}(t, T_n)$ is the value at time $t$ of a defaultable annuity that pays 1 unit of currency at each coupon date $T_i$ provided that $\tau < T_i$ and pays $(\tau - T_{i-1})/(T_i - T_{i-1})$ if default occurs in $[T_{i-1}, T_i]$. The price of the defaultable annuity at time zero is given by

$$\mathcal{A}(0, T_n) = B(0)\sum_{i=1}^{n}\mathbb{E}^{Q_B}\left[\frac{1_{\{\tau > T_i\}}}{B(T_i)} + \frac{\tau - T_{i-1}}{T_i - T_{i-1}}\frac{1_{\{T_{i-1} \leq \tau \leq T_i\}}}{B(\tau)}\right]. \tag{661}$$

The total value of the CDS from the protection buyer's point of view is

$$V_{\text{CDS}}(t) = V_{\text{prot}}(t) - V_{\text{prem}}(t). \tag{662}$$

The value of coupon payments $c$ is determined such that the contract is worthless at inception. Defining

$$s(t) = \frac{\left(1 - \mathbb{E}^{Q_B}[R]\right)\mathcal{D}(t, T_n)}{\mathcal{A}(t, T_n)} \tag{663}$$

we find that $c = s(0)$ and

$$V_{\text{CDS}}(t) = (s(t) - c)\mathcal{A}(t, T_n). \tag{664}$$

Given a set of CDS market quotes for the par spread and an estimate for the recovery rate (typically around 40%) we can infer the probability distribution of default times $G(0, T)$.

The way this is achieved in practice is by expressing the par spreads for different maturities in terms of hazard rates, which are defined by

$$1 - G(t, T) = e^{-\int_t^T ds\, h(t,s)}. \tag{665}$$

The interpretation of $h(t, s)ds$ is the risk-neutral probability of default in the interval $[s, s+ds]$ conditional upon not defaulting before time $s$.

The market standard approach is to assume that $h(0, s)$ is piecewise constant. For example if we are given two market quotes for CDSs with maturities $T_1 < T_2$ then we choose $h(0, s) = \lambda_1$ for $0 \leq s < T_1$ such that the par spread for the $T_1$-maturity CDS matches the market quote. Having found $\lambda_1$ we then assume

$$h(0, s) = \begin{cases} \lambda_1, & 0 \leq s < T_1 \\ \lambda_2, & T_1 \leq s < T_2 \end{cases} \tag{666}$$

and determine $\lambda_2$ such that the par spread for the $T_2$-maturity CDS matches the corresponding market quote.

### 7.3.1 Approximate pricing formula

Let us assume that the term structure of interest rates observed at time $t$ is flat so that

$$P(t, s) = e^{-r_t(s-t)}. \tag{667}$$

If we moreover assume that the hazard rate $h(t, s)$ at time $t$ has a flat term structure given by $\lambda_t$ and that $R$ is a non-random variable then the value of the protection leg is given by

$$V_{\text{prot}}(t) = \lambda_t(1 - R) \int_t^{T_n} ds\, e^{-(r_t+\lambda_t)(s-t)} = \frac{\lambda_t(1 - R)(1 - e^{-(r_t+\lambda_t)(T_n-t)})}{r_t + \lambda_t}. \tag{668}$$

We see that the protection leg is bounded from above by $1 - R$ because $1 - e^{-(r_t+\lambda_t)(T_n-1)} \leq 1$ and $\lambda_t/(r_t + \lambda_t) \leq 1$. The value of the premium leg is given by

$$V_{\text{prem}}(t) = c \sum_{i=1}^n e^{-(r_t+\lambda_t)(T_i-t)}. \tag{669}$$

The value of the CDS contract is thus given by (where $c = s_0$)

$$V_{\text{CDS}}(t) = (s_t - c) \sum_{i=1}^n e^{-(r_t+\lambda_t)(T_i-t)}, \qquad s_t = \frac{\lambda_t(1 - R)(1 - e^{-(r_t+\lambda_t)(T_n-t)})}{(r_t + \lambda_t) \sum_{i=1}^n e^{-(r_t+\lambda_t)(T_i-t)}}. \tag{670}$$

### 7.4 Forward credit default swaps

A forward start CDS is an agreement to enter a credit default swap at some time $t_F > 0$ in the future, assuming the firm has not already defaulted by then. The value of the protection leg is now given by

$$V_{\text{prot}}(t) = \left(1 - \mathbb{E}_t^{Q_B}[R]\right) \int_{t_F}^T ds\, P(t, s)\frac{\partial}{\partial s}G(t, s). \tag{671}$$

76

On the premium leg we have the same cash flows as the standard CDS. However, these cash flows are now conditional upon the firm having not defaulted by time $t_F$. Consider a defaultable annuity with coupon dates $\{T_1, \ldots, T_n\}$ and a forward start annuity with start date $t_F < T_1$. By absence of arbitrage, we have

$$\mathcal{A}(t, T_n) = \mathcal{A}(t, t_F) + \mathcal{A}(t, t_F, T_n) \tag{672}$$

and thus

$$V_{\text{prem}}(t) = c\mathcal{A}(t, t_F, T_n) \tag{673}$$
$$= c\left[\mathcal{A}(t, T_n) - \mathcal{A}(t, t_F)\right]. \tag{674}$$

So the value of the forward CDS contract at time $t$ is

$$V_{\text{CDS}}(t, t_F) = (s(t, t_F) - c)\left[\mathcal{A}(t, T_n) - \mathcal{A}(t, t_F)\right]. \tag{675}$$

where

$$s(t, t_F) \equiv \frac{\left(1 - \mathbb{E}_t^{Q_B}[R]\right)\int_{t_F}^{T} ds\, P(t,s)\frac{\partial}{\partial s}G(t,s)}{\mathcal{A}(t, t_F, T_n)}. \tag{676}$$

and $c = s(0, t_F)$.

## 7.5  Default swaptions

A credit default swaption is an option to enter a credit default swap at some time $t_E > 0$ in the future. The pay-off at expiry $t_E$ of the payer swaption is given by

$$V_{\text{Pay}}(t_E) = (s(t_E) - K)^+ 1_{\{\tau > t_E\}}\mathcal{A}(t_E, T_n) \tag{677}$$

where the indicator function means that there is no pay-off if the firm defaults prior to the exercise date of the swaption. To calculate $V_{\text{Pay}}(0)$ it is convenient to use the following choice of numeraire

$$A(t) = 1_{\{\tau > t\}}\mathcal{A}(t, t_E, T_n). \tag{678}$$

By the fundamental theorem of asset pricing we obtain

$$V_{\text{Pay}}(0) = A(0)\mathbb{E}^{Q_A}\left[\frac{(s(t_E) - K)^+ 1_{\{\tau > t_E\}}\mathcal{A}(t_E, T_n)}{A(t_E)}\right] \tag{679}$$
$$= \mathcal{A}(0, t_E, T_n)\mathbb{E}^{Q_A}\left[(s(t_E) - K)^+\right] \tag{680}$$

where in the second line we have used the fact that $\mathcal{A}(t_E, t_E, T_n) = \mathcal{A}(t_E, T_n)$ and $1_{\{\tau > 0\}} = 1$. Similarly the pay-off for a receiver swaption (a put) is

$$V_{\text{Rec}}(t_E) = (K - s(t_E))^+ 1_{\{\tau > t_E\}}\mathcal{A}(t_E, T_n) \tag{681}$$
$$V_{\text{Rec}}(0) = \mathcal{A}(0, t_E, T_n)\mathbb{E}^{Q_A}\left[(K - s(t_E))^+\right] \tag{682}$$

### 7.5.1 Put-call parity

The pay-offs of a payer and receiver default swaption are related by

$$V_{\text{Pay}}(t_{\text{E}}) - V_{\text{Rec}}(t_{\text{E}}) = (s(t_{\text{E}}) - K)\, 1_{\{\tau > t_{\text{E}}\}} \mathcal{A}(t_{\text{E}}, T_n). \tag{683}$$

Applying the fundamental theorem of asset pricing using $A(t)$ as the numeraire we obtain

$$V_{\text{Pay}}(0) - V_{\text{Rec}}(0) = \mathcal{A}(0, t_{\text{E}}, T_n) \left( \mathbb{E}^{Q_A} [s(t_{\text{E}})] - K \right) \tag{684}$$
$$= \mathcal{A}(0, t_{\text{E}}, T_n) \left( s(0, t_{\text{E}}) - K \right) \tag{685}$$

where we have used the Martingale property of the spread and the fact that $s(t_{\text{E}}, t_{\text{E}}) = s(t_{\text{E}})$,

$$s(0, t_{\text{E}}) = \mathbb{E}^{Q_A} [s(t_{\text{E}}, t_{\text{E}})] = \mathbb{E}^{Q_A} [s(t_{\text{E}})] \tag{686}$$

### 7.5.2 Black model

If we assume that $s(t, t_{\text{E}})$ follows geometric Brownian motion then then the swaption can be priced using the usual Black formulas; namely,

$$V_{\text{Pay}}(0) = \mathcal{A}(0, t_{\text{E}}, T_n) \left[ s(0, t_{\text{E}}) N(y + \sigma \sqrt{T_n}) - K N(y) \right] \tag{687}$$

where

$$y = \frac{\log \left( \frac{s(0, t_{\text{E}})}{K} \right) + \frac{1}{2} \sigma^2 T_n}{\sqrt{T_n}} \tag{688}$$

## 7.6 Intensity models

Intensity models postulate that the time of default is given by the first jump time of a Poisson process with stochastic intensity $\lambda$. The survival probability is thus assumed to be given by

$$Q_B(\tau > T \mid I_t) = \mathbb{E}_t^{Q_B} \left[ e^{-\int_t^T ds\, \lambda(s)} \right]. \tag{689}$$

Typically $\lambda$ is assumed to follow a non-negative interest rate process such as the SRD model, in which case the hazard rate can be analytically calculated. Notice that by definition of the hazard rate we have the identity

$$e^{-\int_t^T ds\, h(t,s)} = \mathbb{E}_t^{Q_B} \left[ e^{-\int_t^T ds\, \lambda(s)} \right], \tag{690}$$

which implies $h(t, t) = \lambda(t)$. Let us consider the price of a zero-recovery defaultable ZCB

$$\mathcal{P}_0(t, T) = B(t) \mathbb{E}_t^{Q_B} \left[ \frac{1_{\tau > T}}{B(t)} \right]. \tag{691}$$

Assuming that the time of default is independent of the level of interest rates, this becomes simply

$$\mathcal{P}_0(t, T) = P(t, T) Q_B(\tau > T \mid I_t) \tag{692}$$

and thus the price in terms of the intensity is

$$\mathcal{P}_0(t, T) = \mathbb{E}_t^{Q_B} \left[ e^{-\int_t^T ds\,(r(s)+\lambda(s))} \right]. \tag{693}$$

It can be shown that the above expression is also valid when the interest rate and default intensity are dependent.

As a concrete example, suppose that the risk-free rate and default intensity satisfy correlated Ho-Lee processes,

$$dr(t) = \theta(t)\,dt + \sigma\,d\hat{w}_1(t) \tag{694}$$
$$d\lambda(t) = \omega(t)\,dt + \eta\,d\hat{w}_2(t) \tag{695}$$

where $\hat{w}_1$ and $\hat{w}_2$ are Brownian motions under $Q_B$ with correlation $\rho$. It is convenient to express $\hat{w}_1$ and $\hat{w}_2$ in terms of uncorrelated Brownian motions $Z_1$ and $Z_2$,

$$\hat{w} = Z_1, \tag{696}$$
$$\hat{w}_2 = \rho Z_1 + \sqrt{1 - \rho^2} Z_2. \tag{697}$$

Integrating the SDEs we obtain

$$\int_0^T ds\,[r(s) + \lambda(s)] = [r(0) + \lambda(0)]T + g(0, T) + f(0, T) + \tag{698}$$

$$+ (\eta\rho + \sigma) \int_0^T ds\, Z_1(s) + \eta\sqrt{1 - \rho^2} \int_0^T ds\, Z_2(s)$$

where $f(t, T) = \int_t^T ds \int_t^s du\,\theta(u)$ and $g(t, T) = \int_t^T ds \int_t^s du\,\omega(u)$. Since $Z_1$ and $Z_2$ are uncorrelated we find

$$-\int_0^T ds\,[r(s) + \lambda(s)] \sim \mathcal{N}\left(-[r(0) + \lambda(0)]T - g(0, T) - f(0, T), \frac{1}{3}(\sigma^2 + \eta^2 + 2\sigma\eta\rho)T^3\right) \tag{699}$$

Recall that if $X \sim \mathcal{N}(\mu, \sigma^2)$ then $\mathbb{E}\left[e^X\right] = e^{\mu + \sigma^2/2}$. This follows from

$$\int_{-\infty}^\infty dx\, f(x)e^{\mu + \sigma x} = e^{\mu + \sigma^2/2} \tag{700}$$

where $f(x)$ is the standard normal pdf. Using this result we find that the price of a zero-recovery, defaultable ZCB is given by

$$\mathcal{P}_0(t, T) = \mathbb{E}^{Q_B}\left[e^{-\int_t^T ds(r(s)+\lambda(s))}\right], \tag{701}$$

$$= \exp\left[-[r(t) + \lambda(t)](T - t) - g(t, T) - f(t, T) + \frac{1}{6}(\sigma^2 + \eta^2 + 2\sigma\eta\rho)(T - t)^3\right]. \tag{702}$$

One can show by explicit calculation that the bond price satisfies the following pricing PDE,

$$\frac{\partial g}{\partial t} + \frac{\sigma^2}{2}\frac{\partial^2 g}{\partial r^2} + \frac{\eta^2}{2}\frac{\partial^2 g}{\partial \lambda^2} + \rho\sigma\eta\frac{\partial^2 g}{\partial r\partial \lambda} + \theta(t)\frac{\partial g}{\partial r} + \omega(t)\frac{\partial g}{\partial \lambda} - (r + \lambda)g(r, \lambda, t) = 0. \tag{703}$$

Now consider a zero-coupon bond with deterministic interest rate $r = r(t)$, default intensity $\lambda = \lambda(t)$ and recovery $R = R(t)$. Clearly the default time is independent of the interest rate level, so the price of the bond at time $t < T$ is given by

$$B(t) = \mathcal{P}_0(t, T) + \int_t^T ds\, R(s) P(t, s) \frac{\partial}{\partial s} G(t, s) \tag{704}$$

$$= e^{-\int_t^T ds\,(r(s)+\lambda(s))} + \int_t^T ds\, R(s)\lambda(s) e^{-\int_t^s du\,(r(u)+\lambda(u))}. \tag{705}$$

Differentiating with respect to time $t$ and using Leibniz's rule we find that the bond price satisfies the following ODE,

$$\frac{\partial B(t)}{\partial t} - \lambda(t)\big[B(t) - R(t)\big] - r(t)B(t) = 0. \tag{706}$$

The first, second and fourth term are easily recognized as the time-derivative and effective discount term which follow from the Feynman-Kac theorem. The third term is the contribution due to the recovery. The form of these terms stay the same even when $\lambda(t)$, $r(t)$ are stochastic and when the recovery is state-dependent.

### 7.6.1 Simulating default times in intensity models

Given the hazard rates $h(0, T)$ it is a simple matter to simulate the default times. We simply use the fact that since $F(T) \equiv Q_B(\tau \leq T \mid I_0)$ is the cumulative distribution function for $\tau$, then the probability integral transform $F(\tau)$ is uniformly distributed on $[0, 1]$. Let $U$ be a random variable with uniform distribution on $[0, 1]$. For each draw of $U$ we set the default time to $\tau = F^{-1}(U)$. Said another way, the random default time is the solution of the equation

$$U = 1 - e^{-\int_0^\tau ds\, h(0,s)}. \tag{707}$$

## 7.7 Multi-name credit derivatives

### 7.7.1 Basket CDS's

A basket CDS is equivalent to a portfolio of $m$ single-name CDS's. A notional $N$ is specified and each single-name CDS is assigned a weighting $\sum_{i=1}^m N_i = N$. If, for example, the first reference obligation defaults, the protection seller makes a payment given by the loss given default multiplied by the notional for that obligation, $N_1(1 - R_1)$. The basket CDS continues to maturity with revised notional amount $N - N_1$. CDS indices are basically standardized baskets CDS's.

An $n$-th to default ($n$-2D) CDS is a basket CDS that only makes a payment after $n$ of the $m$ credits in the basket have defaulted. The spreads of $n$-2D decrease with increasing $n$ since the probability that the protection payment will be triggered decreases.

### 7.7.2 Simulation

Consider an $n$-2D CDS with tenor $T_N$ on a basket of $m$ credits. We need to simulate the default times $(\tau_1, \ldots, \tau_m)$ under the risk-neutral probability measure $Q_B$ from which we can

determine the time of $n$th default, which we denote by $\tau$. The cash flow on the protection leg is given by

$$V_{\text{prot}}(0) = B(0)\mathbb{E}^{Q_B}\left[\frac{(1-R)1_{\{\tau \leq T_N\}}}{B(\tau)}\right] \tag{708}$$

and similarly for the cash flow of the annuity. The mean of the simulated values is the Monte-Carlo estimate for $V_{\text{prot}}(0)$ and $V_{\text{prem}}(0)$ which can be used to determine the spread.

The simulation of $(\tau_1, \ldots, \tau_m)$ requires the joint probability distribution which incorporates correlations amongst the default times,

$$F(T_1, \ldots, T_m) \equiv Q_B(\tau_1 \leq T_1, \ldots, \tau_m \leq T_m \mid I_0). \tag{709}$$

The challenge is to find a joint probability distribution $F$ which is consistent with the individual probability distributions calibrated form market quotes

$$F_i(T_i) \equiv Q_B(\tau_i \leq T_i \mid I_0), \qquad 1 \leq i \leq m. \tag{710}$$

Consider the following joint probability distribution,

$$C(u_1, \ldots, u_m) = F\left(F_1^{-1}(u_1), \ldots, F_m^{-1}(u_m)\right). \tag{711}$$

By construction, this is a multivariate distribution on $[0,1]^m$ with uniform marginals.

We can now simulate correlated default times $(\tau_1, \ldots, \tau_m)$ by simply drawing random variables $(U_1, \ldots, U_m)$ with joint probability distribution $C$ and solving the following $m$ equations for $\tau_i$,

$$U_i = 1 - e^{-\int_0^{\tau_i} ds\, h(0,s)}, \qquad 1 \leq i \leq m. \tag{712}$$

# 8 FX derivatives

Let $X(t)$ be the price of a foreign currency at time $t$. The price of a forward contract entered at time $t$ for delivery of the foreign currency at time $T$ satisfies

$$0 = \mathbb{E}_t^{Q_B}\left[\frac{F(T) - X(T)}{B(T)}\right] \implies F(t) = \frac{\mathbb{E}_t^{Q_B}\left[X(T)/B(T)\right]}{\mathbb{E}_t^{Q_B}\left[1/X(T)\right]}. \tag{713}$$

If we assume that rates are constant then

$$F(t) = \mathbb{E}_t^{Q_B}\left[X(T)\right], \tag{714}$$

$$= e^{(r_d - r_f)(T-t)}X_t. \tag{715}$$

# References

[1] Almgren, Robert, and Neil Chriss. "Optimal execution of portfolio transactions." Journal of Risk 3 (2001): 5-40.

[2] Cheridito, Patrick, and Tardu Sepin. "Optimal trade execution under stochastic volatility and liquidity." Applied Mathematical Finance 21, no. 4 (2014): 342-362.

[3] Grleanu, Nicolae, and Lasse Heje Pedersen. "Dynamic trading with predictable returns and transaction costs." The Journal of Finance 68, no. 6 (2013): 2309-2340.

[4] Tardu Sepin and James Stokes, Unpublished.

# A    Discretization schemes

## A.1    Theta scheme

The trinomial tree scheme is intuitive but has poor convergence and stability properties. The theta scheme involves first replacing all $x$-derivatives by their central-difference approximations and then Taylor expanding the resulting set of coupled ordinary differential equations around an arbitrary point in $[t_n, t_{n+1}]$ parametrized by $\theta \in [0, 1]$. We can write the system of coupled ODEs in vector notation as

$$\frac{\partial \vec{V}}{\partial t} + A(t)\vec{V}(t) = 0. \tag{716}$$

Taylor expanding in $t$ around the point

$$t_{n+1}(\theta) \equiv \theta t_n + (1 - \theta)t_{n+1} \tag{717}$$

we obtain the following system of finite-difference equations for the interior points

$$\frac{\vec{V}(t_{n+1}) - V(t_n)}{\Delta t} + \theta A\big(t_{n+1}(\theta)\big)\vec{V}(t_n) + (1 - \theta)A\big(t_{n+1}(\theta)\big)\vec{V}(t_{n+1}) = 0 \tag{718}$$

which can be rearranged to give

$$\big[\mathbf{1} - \theta \Delta t A\big(t_{n+1}(\theta)\big)\big]\vec{V}(t_n) = \big[\mathbf{1} + (1 - \theta)\Delta t A\big(t_{n+1}(\theta)\big)\big]\vec{V}(t_{n+1}) \tag{719}$$

where

$$\mathbf{1} = 0 \oplus \mathrm{diag}(1, \dots, 1) \oplus 0 \tag{720}$$

In practical applications we usually replace (721) by

$$\big[\mathbf{1} - \theta \Delta t A(t_n)\big]\vec{V}(t_n) = \big[\mathbf{1} + (1 - \theta)\Delta t A(t_{n+1})\big]\vec{V}(t_{n+1}) \tag{721}$$

which does not affect the convergence properties. The special cases $\theta = 0$, $\theta = 1$ and $\theta = 1/2$ are referred to as fully explicit, fully implicit and Crank-Nicholson schemes, respectively. The fully explicit scheme is closely related to the trinomial tree scheme and has poor convergence properties. Boundary conditions must also be imposed to complete the recursive scheme.

## A.2    ADI scheme

Consider the two-dimensional backward heat equation

$$\frac{\partial u}{\partial t}(x, y, t) + \nabla^2 u(x, y, t) = 0. \tag{722}$$

82

Discretizing $x$ and $y$ we obtain the system of ODEs

$$\frac{\partial}{\partial t}u^{j_1,j_2}(t)+\frac{1}{\Delta x^2}\left[u^{j_1+1,j_2}(t)+u^{j_1-1,j_2}(t)-2u^{j_1,j_2}(t)\right]+\frac{1}{\Delta y^2}\left[u^{j_1,j_2+1}(t)+u^{j_1,j_2-1}(t)-2u^{j_1,j_2}(t)\right]=0$$
(723)

For comparison a mixed derivative term gives

$$\frac{\partial^2 u}{\partial x \partial y}=\frac{1}{2\Delta x}\left[\frac{\partial u^{j_1+1}}{\partial x}(y,t)-\frac{\partial u^{j_1-1}}{\partial x}(y,t)\right]$$
(724)

$$=\frac{1}{4\Delta x \Delta y}\left[u^{j+1,j_2+1}(t)-u^{j+1,j_2-1}(t)-u^{j_1-1,j_2+1}(t)+u^{j_1-1,j_2-1}(t)\right]$$
(725)

In general when mixed derivative terms are absent we will obtain a system of ODEs of the form

$$\frac{\partial V}{\partial t}+\left[A_1(t)\otimes\mathbf{1}+\mathbf{1}\otimes A_2(t)\right]V(t)=0$$
(726)

where $V(t)$ is now a two-component tensor with components $V(t)=\left(V^{j_1,j_2}(t)\right)$. In component notation this equation reads

$$\frac{\partial}{\partial t}V^{j_1,j_2}(t)+A_{1,j}^{j_1}\delta_k^{j_2}V^{j,k}(t)+\delta_j^{j_1}A_{2,k}^{j_1}V^{j,k}(t)=0$$
(727)

where summation over repeated indices is implied. In the theta-scheme, the finite-difference equation for the interior points is given by

$$\left[\mathbf{1}\otimes\mathbf{1}-\theta\Delta t\left(A_1(t_n)\otimes\mathbf{1}+\mathbf{1}\otimes A_2(t_n)\right)\right]V(t_n)=\left[\mathbf{1}\otimes\mathbf{1}+(1-\theta)\Delta t\left(A_1(t_{n+1})\otimes\mathbf{1}+\mathbf{1}\otimes A_2(t_{n+1})\right)\right]V(t_{n+1}).$$
(728)

In principal we just need to represent the operator on the left-hand side in matrix form and then invert to express $V(t_n)$ in terms of $V(t_{n+1})$. In practice, however, this is not computationally efficient because the matrix representation is not tri-diagonal.

The trick is to represent the operator on the left-hand side as a single tensor product. Fortunately this can be achieved up to the required level of accuracy. We begin by specializing to Crank-Nichsolson scheme ($\theta=1/2$) and notice that (up to corrections of order $\Delta t^2$)

$$\left[\mathbf{1}-\frac{1}{2}\Delta t A_1(t_n)\right]\otimes\left[\mathbf{1}-\frac{1}{2}\Delta t A_2(t_n)\right]V(t_n)=\left[\mathbf{1}+\frac{1}{2}\Delta t A_1(t_{n+1})\right]\otimes\left[\mathbf{1}+\frac{1}{2}\Delta t A_2(t_{n+1})\right]V(t_{n+1})$$
(729)

or in simplified notation,

$$\left(\mathbf{1}-M_1(t_n)\right)\otimes\left(\mathbf{1}-M_2(t_n)\right)V(t_n)=\left(\mathbf{1}+M_1(t_{n+1})\right)\otimes\left(\mathbf{1}+M_2(t_{n+1})\right)V(t_{n+1}).$$
(730)

## A.3   Callable bonds

In this section we will be concerned with pricing bonds with embedded American options. One way to think about a long position in a callable bond is as portfolio which is long a bullet bond and short a call option on the bullet bond. We therefore need to consider options on

bullet bonds. We will work in the framework of the Ho-Lee model where the spot rate follows generalized Brownian motion under the risk-neutral probability measure $Q_B$,

$$dr(t) = \theta(t)dt + \sigma d\hat{w}(t) \tag{731}$$

The zero-coupon discount curve at time $t$ is determined by

$$P(t,T) = \mathbb{E}_t^{Q_B}\left[e^{-\int_t^T ds\, r(s)}\right]. \tag{732}$$

We will ignore the problem of calibration and assume $\theta = 0$, in which case the discount curve is determined entirely in terms of the constant volatility $\sigma$ and the prevailing level of the spot rate,

$$P(t,T) = e^{-r(t)(T-t)+\frac{\sigma^2}{6}(T-t)^3}. \tag{733}$$

Consider a one-year zero-coupon bond callable at 6 months with strike $K = 0.4$. Substituting into the analytical pricing formulas we obtain[4],

$$P(0,1) - \mathbf{CZCB}(0,1/2,1,0.4) = \$\,0.344467. \tag{734}$$

An alternative approach is to view a callable bond as a spot-rate contingent claim which in the case of the Ho-Lee model satisfies the PDE (assuming constant default intensity $\lambda$),

$$\frac{\partial g}{\partial t} + \frac{1}{2}\sigma^2\frac{\partial^2 g}{\partial r^2} - (r+\lambda)g(r,t) = 0. \tag{735}$$

This equation can be discretized on a grid with coordinates $t_n = n\Delta t$ with $0 \le n \le N$ and $r_j = j\Delta r$ with $0 \le j \le M$. Denoting $g_n^j = g(r_j, t_n)$, the finite-difference equation in the trinomial tree scheme is

$$g_n^j = h_n^j, \qquad h_n^j = \frac{1}{1+r_j\Delta t}\left[pg_{n+1}^{j+1} + (1-2p)g_{n+1}^j + pg_{n+1}^{j-1}\right], \qquad p = \frac{\sigma^2\Delta t}{2\Delta r^2}. \tag{736}$$

The boundary conditions suitable for a zero-coupon bond are

$$g_N^j = 1, \qquad 0 \le j \le M \tag{737}$$

and

$$g_n^M + g_n^{M-2} - 2g_n^{M-1} = 0 = g_n^0 + g_n^2 - 2g_n^1, \qquad 0 \le n \le N-1. \tag{738}$$

In the case of optimal exercise a callable bond will be redeemed if the holding value of the bond exceeds the early redemption price $K(t)$. It follows that for each date occurring in the strike schedule we should impose the additional boundary condition,

$$g_n^j = \min\left(K(t_n), h_n^j\right), \qquad 1 \le j \le M-1. \tag{739}$$

Let us reprice the callable bond of the previous example on a grid with $N = 12000$, $M = 100$ and $\Delta r = 0.01$. Implementing the callable feature with an additional boundary condition at call date $t_{N/2+1}$, we find that the interpolated price is in excellent agreement with (734),

$$g_{N/2+1}^j = \min(K, h_{N/2+1}^j), \qquad 1 \le j \le M-1, \qquad \$\,0.344469. \tag{740}$$

[4]Unless otherwise specified, we will assume $\sigma = 0.16$ and $r(0) = 0.3$ for all numerical examples.

### A.3.1 Coupon bonds

The price of a one year bullet bond with unit face value paying semi-annual coupons of $c = 30\%$ is given by

$$cP(0, 1/2) + (1 + c)P(0, 1) = \$\, 1.22553. \tag{741}$$

This bond can be implemented in the finite difference solver by imposing the terminal condition $g_N^j = 1 + c$ for $0 \le j \le M$ in addition to

$$g_{N/2+1}^j = h_{N/2+1}^j + c, \qquad 1 \le j \le M - 1, \qquad \$\, 1.22552. \tag{742}$$

Suppose that the coupon bond of the previous example is callable at par on the first coupon date. The price can be obtained using the Jamshidian trick. Since the first coupon payment coincides with the exercise date we do not include it in the calculation of $r_K$,

$$(1 + c)P(r_K, 1/2, 1) = 1 \implies K_2 = 0.769231. \tag{743}$$

The value of the callable coupon bond is thus given by

$$P(0, 1) + cP(0, 1/2) + cP(0, 1) - (1 + c)\mathbf{CZCB}(0, 1/2, 1, K_2) = 1.11914. \tag{744}$$

In terms of the finite difference solver the only difference from the straight bond is that we change the boundary condition at $n = N/2 + 1$ to allow the possibility of exercising the option,

$$g_{N/2+1}^j = \min(1, h_{N/2+1}^j) + c, \qquad 1 \le j \le M - 1, \qquad \$\, 1.11914. \tag{745}$$

We see that the callable feature reduces the value of the bond but it still trades above par. If the bond is continuously callable then it will always trade below par. In the continuously callable case no analytical formula is available and we must resort to the finite difference solver,

$$g_n^j = \begin{cases} \min(1, h_{N/2+1}^j) + c, & n = N/2 + 1, \\ \min(1, h_n^j), & n \ne N, N/2 + 1, \end{cases} \qquad 1 \le j \le M - 1, \qquad \$\, 0.861235. \tag{746}$$

### A.3.2 Effect of default probability

In the case when the bond is defaultable with constant default intensity $\lambda$, the pricing PDE is modified to

$$\frac{\partial g}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 g}{\partial r^2} - (r + \lambda)g(r, t) = 0. \tag{747}$$

The finite difference scheme is the same as before except that the continuation value is now given by

$$h_n^j = \frac{1}{1 + (r_j + \lambda)\Delta t} \left[ p g_{n+1}^{j+1} + (1 - 2p)g_{n+1}^j + p g_{n+1}^{j-1} \right]. \tag{748}$$

The PV of a defaultable bond can be regarded as a function of the free parameters $r(0)$ and $\lambda$. The present value of a zero-recovery, defaultable ZCB is given by the analytical formula,

$$\mathcal{P}_0(0, T) = e^{-(r(0) + \lambda)T + \frac{\sigma^2}{6}T^3} \tag{749}$$

and thus the price of a 1 year ZCB with default intensity $\lambda = 0.1$ is given by

$$\mathcal{P}_0(0,1) = \$\,0.673186 \tag{750}$$

Using the same boundary conditions for a non-defaultable ZCB and using the new formula for the continuation value we find agreement from the finite difference solver,

$$g_N^j = 1, \qquad 0 \le j \le M, \qquad \$\,0.673186. \tag{751}$$