Computational Finance – Financial Instruments

The Greek philosopher Thales of Miletus (*circa* 624-546 BC) is widely regarded as the first "thinker" to make use of the notion of an option. One version of the story has Thales as a frustrated philosopher who wanted to show his peer businessmen that he could use his intelligence to be rich if he wanted to, but it was just that he chose to remain poor.

One year, Thales predicted a bumper crop for that year's olive harvest. On the basis of this prediction, he paid a *small* deposit to all owners of olive presses to buy himself the *right* to rent all the olive presses at some standard rate. Imagine the owners of oil presses thinking this was a win-win. Not only were they getting a deposit ahead of time, but they were also going to be able to rent the presses at their standard rate. Indeed, the bumper olive crop did arrive, and Thales, the effective owner of all the olive presses, was able to make a small fortune renting out these presses at a significant premium to the price he had already agreed to rent the presses for.

This very first use of the notion of an *option* has already embedded in it many of the important reasons for the existence of financial derivatives. Most importantly, one wishes to somehow monetize a prediction for some quantity (the weather in Greece) which is not something that can be bought and sold as some good in some marketplace. As a result one writes a contract on some quantity (olive presses) which one can buy and sell. The key consideration here is that the future value of the olive press (the quantity which *can* be bought or sold) depends directly or indirectly on the weather in Greece (the quantity predicted, which *cannot* be bought or sold). Constructing such "derivative contracts" is itself a creative task, and in some sense represents the real ingenuity of Thales – he figured out a way to profit from a weather prediction based on commodities already existing. However, given the existence of such a derivative contract (which embodies the right to rent the olive presses at the standard rate in the future) an important computational task presents itself. How much should this "small deposit" (the *price* of the option) be that Thales should be willing to pay to own this right to rent? The price of this option naturaly would depend on: the probability of Thales' prediction on the weather coming true; the size of the bumper crop in that event (and hence the level of demand for the olive presses) which would indirectly determine the profit that Thales stood to make had his prediction come true; and, how valuable the profit that Thales stands to make really is, given that it only arrives at the end of the summer and he has to pay the deposit at the begining, denying himself all the consumption that that deposit could have bought him over the summer.

Rather than deal with goods in a barter economy, we will deal with money and within this framework we can explicitly quantify, through the notion of an interest rate, the last consideration above regarding how valuable future profit really is today.

1 Money

Roughly speaking, money can be equated to consumption, and it is more or less accepted that present consumption is worth more than an equal consumption at some future date.

> *Fundamental Principle:* Money today is worth more than the same amount of money tomorrow.

Intuitively, this statement says that if I consume today, I am better off.

Example: Suppose that \$1 today buys me a corn seed. If I plant it, in a year, this crop produces 1 corn cob (which I can eat) and 2 corn seeds. Therefore, the \$1 today buys me 1 seed, but in a year it is worth a cob and 2 seeds. If we assume that my labor, water, land, etc. would be idle unless I plant the seed, we see that 1 seed today is worth 2 seeds and a cob tomorrow.

To formalize this intuition, we say that

$$\$1 \text{ (today)} = \$(1+R) \text{ (in a year)}.$$

R is called the interest for the period of time from today to a year later. To make this more general, we may talk about any two times in the future, $t_1 < t_2$, and we can define the $[t_1, t_2]$ interest rate $R(t_1, t_2)$ by

$$\$1 \text{ (at } t_1) = \$(1 + R(t_1, t_2)) \text{ (at } t_2).$$

Thus, $R(t_1, t_2)$ is the additional monetary value needed to equate the \$1 at time t_1 with its equivalent amount at t_2.

For simplicity, we will assume that the interest rate over a given period of time (for example $[t_1, t_2]$) does not change with time. For example, suppose $t_1 = 1$ year and $t_2 = 2$ years. Then the value of $R(1, 2)$ today is the same as what it would be tomorrow.

The $[t_1, t_2]$-Period Interest Rate $R(t_1, t_2)$. We summarize the interest rate computation here for clarity, but generalize it to any amount of money X at time t_1.

Having X at time t_1 is equivalent to having $X \cdot (1 + R(t_1, t_2))$ at time t_2.

Consequently, if you give a bank X at time t_1, you can expect the bank to give you back $X \cdot (1 + R(t_1, t_2))$ at time t_2. Thus, your X has accrued an amount $XR(t_1, t_2)$ in interest. We will assume that you can *borrow and lend* at the same interest rate, which is a slight approximation because typically you can lend to a bank at a lower interest rate than you can borrow from the bank.

The Interest Rate Law. It suffices to know the function $R(0, t)$ because as we will now derive, there is a relationship between $R(t_1, t_2)$ and $R(0, t_1), R(0, t_2)$. Suppose we have an amount X which we place in the bank at time 0. We know that at time t_1, it grows to an amount

$$X \cdot (1 + R(0, t_1)).$$

By definition of $R(t_1, t_2)$, an amount X' at time t_1 grows to an amount $X' \cdot (1 + R(t_1, t_2))$ by time t_2; applying this equation to $X' = X \cdot (1 + R(0, t_1))$, we see that $X \cdot (1 + R(0, t_1))$ at time t_1 grows to an amount $X \cdot (1 + R(0, t_1)) \cdot (1 + R(t_1, t_2))$ by time t_2. Thus X at time 0 grows to

$$X \cdot (1 + R(0, t_1)) \cdot (1 + R(t_1, t_2))$$

at time t_2. But by definition of $R(0, t_2)$, this must equal $X \cdot (1 + R(0, t_2))$. We therefore have that

$$(1 + R(0, t_1)) \cdot (1 + R(t_1, t_2)) = 1 + R(0, t_2). \tag{1}$$

Multiplying out the LHS, we obtain the fundamental equation

$$R(0, t_2) = R(0, t_1) + R(t_1, t_2) + R(0, t_1) \cdot R(t_1, t_2) \tag{2}$$

Note that the above law must hold for any choice of t_1. Alternatively, we can obtain an expression for $R(t_1, t_2)$ only in terms of $R(0, t_1)$ and $R(0, t_2)$,

$$R(t_1, t_2) = \frac{R(0, t_2) - R(0, t_1)}{1 + R(0, t_1)}. \tag{3}$$

Since we can work only in terms of $R(0, t)$, we will usually simplify notation and use $R(t)$ to denote $R(0, t)$. Equation (2) places quite a strong constraint on what possibilities can be chosen for the interest rate function $R(t)$.

Exercise 1.1

The above discussion considered two periods $[0, t_1]$ and $[t_1, t_2]$. Generalize the interest rate law to N periods defined by the times $0, t_1, t_2, \ldots, t_N$. Specifically, show that for all $2 \le k \le N$,

$$\prod_{i=1}^{k}(1 + R(t_{i-1}, t_i)) = 1 + R(0, t_k).$$

(Consider X dollars put in the bank at time 0). This is called the multi-period compounding law.

In the above exercise, let $r_i = R(t_{i-1}, t_i)$, and let X_0 be put in the bank at time 0. Then at time t_k, the money has grown to

$$X_k = X_0 \cdot \prod_{i=1}^{k}(1 + r_i).$$

This is known as multi-period compounding. When all the r_i are constant equal to r, this expression becomes $X_k = X_0 \cdot (1 + r)^k$. The compounded value X_k is called the time-t_k *future value* of X_0. We can reverse this process. Suppose that at time t_k, we have an amount of money X_k. Then we know that this is equivalent to having an amount $X_0 = X_k / \prod_{i=1}^{k}(1 + r_i)$ at time 0, as this amount would grow to X_k by time t_k. This is known as the present value (PV) of X_k,

$$PV(X_k) = \frac{X_k}{\prod_{i=1}^{k}(1 + r_i)}.$$

Exercise 1.2

Specialize the above equation to the case where all the r_i are equal. In a practical sense (roughly speaking) what relationship among the t_i's do you expect for all the r_i to be the same.

We will now try to understand what this function $R(t)$ must look like.

A Stationarity Assumption. Suppose that the function $R(t_1, t_2)$ does not depend on the choice of origin. This means that it depends on the actual values of t_1, t_2 only through their difference $\tau = t_2 - t_1$. Let $R(\tau)$ be this function, where the stationarity is explicitly shown by R only depending on τ. Consider two times $t_1 = t$ and $t_2 = t + \Delta t$. The intuition is that we will eventually let Δt tend to zero. Rewriting Equation (1) using the stationarity assumption for the selected values of t_1, t_2, we have

$$1 + R(t + \Delta t) = (1 + R(t)) \cdot (1 + R(\Delta t)). \tag{4}$$

Let $D(\tau) = 1 + R(\tau)$. Clearly after zero time, no amount of interest could have accrued; it follows that $R(0) = 0$, or that $D(0) = 1$. After a little massaging, we can rewrite the above equation in a more useful form:

$$\begin{aligned} D(t + \Delta t) &= D(t) \cdot D(\Delta t), \\ &= D(t) \cdot (D(\Delta t) - D(0) + D(0)), \\ \Longrightarrow \qquad D(t + \Delta t) - D(t) \cdot D(0) &= D(t) \cdot (D(\Delta t) - D(0)). \end{aligned}$$

Now, using $D(0) = 1$ and dividing both sides by $D(t)\Delta t$,

$$\left(\frac{D(t + \Delta t) - D(t)}{\Delta t} \right) \frac{1}{D(t)} = \frac{D(\Delta t) - D(0)}{\Delta t}. \tag{5}$$

Notice that the LHS looks like $D'(t)/D(t)$ and the RHS looks like $D'(0)$. This becomes true in a more formal sense only in the limit $\Delta t \to 0$ and if $D(\tau)$ (i.e., $R(\tau)$) is a continuously differentiable function. When $D(\tau)$ is continuously differentiable, we can take the $\Delta t \to 0$ limit in Equation (5) to obtain a differential equation that $D(\tau)$ must satisfy. This differential equation can be uniquely solved using the boundary condition $D(0) = 1$ to obtain the only possible functional form for the interest rate function $D(\tau)$. To cut a long story short, the assumptions of stationarity and differentiability essentially uniquely determines the interest rate function.

Theorem 1.1 (Exponential Growth) *The unique stationary, continuously differentiable function $D(\tau)$ which satisfies (1) is $D(\tau) = e^{r\tau}$, where $r = D'(0)$ is a constant.*

Exercise 1.3

[For the analysis inclined] Prove Theorem 1.1 by showing that $D(\tau)$ must satisfy the differential equation $D'(t) = rD(t)$ and solve.

The constant r in the theorem is usually denoted the instantaneous *rate of interest*. The function $D(\tau)$ is sometimes called the *discount factor* to time τ for reasons that will become clear shortly. In terms of $R(t_1, t_2)$, the theorem states that

$$R(t_1, t_2) = e^{r(t_2 - t_1)} - 1.$$

We can now rewrite all the previous formulas using this form for $R(t_1, t_2)$.

Exercise 1.4

Consider an investment X_0 in the bank at time 0. Consider the value of the investment at the times $t_1, \ldots, t_N$, and let X_k be the value at time t_k. Using the formula $X_k = X_0 \prod_{i=1}^{k}(1+r_i)$, where $r_i = R(t_{i-1}, t_i)$ and the fact that $R(t_1, t_2) = e^{r(t_2-t_1)} - 1$, show that $X_k = X_0 e^{rt_k}$.

(Show that $\prod_{i=1}^{k}(1+r_i) = e^{rt_k}$.)

Analogous to exponential growth, we can now compute present values,

$$PV(X_k) = X_k e^{-rt_k} = \frac{X_k}{D(t_k)}.$$

The present value is the amount of money needed today to generate X_k at time t_k. The above formula is the reason for $D(\tau)$ being refered to as the discount function.

Streams of Payments (Cash Flows). How much money would we need today to give a stream of cashflows $X_1, X_2, \ldots, X_N$ at the future times $t_1, t_2, \ldots, t_N$. One might want to do this, for example, to make periodic payments in the future. We know how much money it takes today for any one of these payments, for example for payment X_k at time t_k, we need $PV(X_k)$ today. If we need to make all these payments, it seems reasonable that we would need the sum of these present values,

$$\begin{aligned} PV(X_1, \ldots, X_N; t_1, \ldots, t_N) &= \sum_{i=1}^{N} PV(X_i), \\ &= \sum_{i=1}^{N} \frac{X_i}{D(t_i)}, \\ &= \sum_{i=1}^{N} X_i e^{-rt_i}. \end{aligned}$$

Exercise 1.5

Suppose that you start with $X_0 = PV(X_1, \ldots, X_N; t_1, \ldots, t_N)$ as given in the formula above and you do the normal thing: put this amount in the bank; when payment X_1 is due at time t_1, you withdraw X_1 from the bank and pay it; similarily you payoff all the payments X_k at the required time t_k, as they become due.

Show that you will be able to make all the payments, i.e., the balance in your bank account will never be less than zero. Further, at the end of your last payment, you have no money in the bank account.

The last exercise really justifies our definition of the present value of a stream of cash flows.

Non-Stationary Setting. While we have defined the rate of interest in the statiionary setting, it is still convenient to define the rate of interest in a non-stationary setting via the discount function. The reason is (as we will soon see), the discount function is something that is set in the market-place by a financial instrument known as a *Bond*, specifically a *Zero Coupon Bond*. Thus, suppose that the discount function $D(t)$ is defined and is continuously differentiable. Then the instantaneous interest rate at time t, $r(t)$, is defined by

$$r(t) = \frac{d}{dt} \log D(t) = \frac{D'(t)}{D(t)}.$$

The function $r(t)$ is some times referred to as the *forward interest rate*, and its dependence on time is often termed the *term structure of the interest rate*. An alternative way to get something that looks like an interest rate is by the definition

$$\bar{r}(t) = \frac{\log D(t)}{t}.$$

This is also referred to as the *yield* and corresponds to the implied constant interest rate which matches $D(t)$ at time t. Its dependence on time is also the term structure of the interest rate. For the stationary setting, with an exponentially growing discount factor, the forward rate and the yield are equivalent. As will be seen later, the yield is typically an easier quantity to compute from real data.

Exercise 1.6

When will the forward rate be larger than the yield. When will it be smaller?

Exercise 1.7

[The Mortgage Calculation]

(a) A mortgage with principle P and N payments of size X at the times $t_1, \ldots, t_N$ is fair if you are indifferent about being the lender or borrower in the mortgage. For a given $r, P, \{t_1, \ldots, t_N\}$, determine the fair payment value X.
(Answer: $X = P / \sum_{i=1}^{N} e^{-rt_i}$.)
Specialize your answer to the case where $t_i = 1, \ldots, N$ (i.e., $t_i = i$), and obtain an answer that does not have a summation – i.e., get a closed form.

(b) Suppose that you can only afford to pay $\$500$ at the times $1, 2, 3, \ldots, 360$ (i.e. $t_i = i$). Compute the maximum mortgage that you can afford to take out. Assume that $e^r = 1.005$. (This roughly corresponds to a 30 year mortgage with a monthly payment of $\$500$ and a yearly interest rate of about 6%.)
Suppose that you need a mortgage amount of $\$100,000$. For how many months will you be paying off the mortgage.

Exercise 1.8

Show that the special case of $D(t) = e^{rt}$ which we derived in the stationary setting gives a constant (instantaneous) forward interest rate r.

Show that analogous to the formula $D(t) = e^{rt}$ (in the stationary case), in the non-stationary case, one has that

$$D(t) = e^{\int_0^t ds\ r(s)}.$$

Exercise 1.9

We give an alternate derivation of the formulas $D(t) = e^{rt}$ ($D(t) = e^{\int_0^t ds\ r(s)}$ in the non-stationary case.) by taking the continuous limit of compunding.

Divide the time period $[0, t]$ into n time periods each of length $\Delta t = t/n$. Let $r_i(\Delta t)$ be the interest rate function over the ith period. Assume that you invest X_0 in a bank at time 0 and let X_t be the amount accrued at time t.

Show that

$$X_t = X_0 \cdot \exp\left(\sum_{i=1}^{n} \log(1 + r_i(\Delta t))\right).$$

We can assume that $r_i(\Delta t)$ must go to zero as Δt goes to zero, as one should not be able to gain any interest over a period of time tending to zero. Hence, using the Taylor expansion for the logarithm, show that

$$X_t = X_0 \cdot \exp\left(\lim_{n\to\infty} \sum_{i=1}^{n} r_i(\Delta t)\right).$$

The exponent has exactly the form of an integral, providing $r_i(\Delta t)$ satisfies some regularity conditions. In particular if $r_i(\Delta t) = r(\frac{i}{n}t)\Delta t$ then we recover the Riemann limit for the integral form we expect.

1.1 The No Arbitrage Axiom.

We heuristically state an important concept here that we will make much more formal later on, namely the concept of no arbitrage.

In a market that is in a stable equilibrium, there does not exist a stream of cash flows accessible to an individual consisting only of non-negative cash flows, with at least one cash flow positive.

Essentially the intuition for why this is so is that if such a stream of cashflows did exist, then someone would immediately grab it. In fact, someone would want an infinite amount of it. If we assume the market is in equilibrium, then no one is grabbing opportunities here or there, hence the no arbitrage axiom. Note that it is an axiom, i.e. we simply believe it to be true. One particular implication of this is that in any accessible stream of cash flows must have present value zero.

Exercise 1.10

Show that the following axiom and the previous axiom of no arbitrage are equivalent.

In a market that is in a stable equilibrium, there does not exist a stream of cash flows with positive present value accessible to any individual.

(The "hard" part is to show that if there exists an accessible stream of cash flows with positive present value then there exists an accessible stream of non-negative cash flows, at least one of which is positive.)

Assume that a stream is accessible means it can either be bought or sold. Conclude that every accessible stream of cash flows must have present value zero.

(Hint: By selling a stream of cashflows with negative present value, show that one obtains a stream with positive present value.)

The no arbitrage axiom is a main tool in the pricing of complex deals in the financial marketplace. We will illustrate with a "complex deal", involving a stream of cash flows. Consider the following deal, which a friend asks you to commit to.

For \$100 at time t_1, you will get back X at time t_2.

How will you determine whether to enter into this deal. If we convert this into a stream of cashflows, it looks like $(-100, X; t_1, t_2)$. A natural approach would be to first compute the present value of this stream of cash flows,

$$PV(-100, X; t_1, t_2) = -100 \cdot e^{-rt_1} + X \cdot e^{-rt_2}.$$

If $PV(-100, X; t_1, t_2) \geq 0$, then ones natural instinct would be to engage in this deal. Thus a criterion for engaging in this deal is

$$X \geq 100 \cdot e^{-r(t_1 - t_2)}.$$

In essence, we are pricing this deal for the friend, saying that if he is willing to *pay* at least this amount at time t_2, then we are willing to engage in this deal. Are we justified in this intuition?

Lets approach it from the arbitrage opportunity perspective. Assume $X \geq 100 \cdot e^{-r(t_1 - t_2)}$, and we accept this deal. Now at time t_1, borrow \$100 from the bank, and give it to the friend. In this case, your cash flow at time t_1 netts out to zero. Now at time t_2, you will obtain X from your friend, and you will owe the bank $100 \cdot e^{r(t_2-t_1)}$. Your nett cash flows at time t_2 are therefore $X - 100 \cdot e^{r(t_2-t_1)} \geq 0$ by the assumption on X. Hence, we have constructed a stream of cash flows $(0, Y; t_1, t_2)$ where $Y \geq 0$. Certainly you cannot be averse to these cash flows. In fact, if $X > 100 \cdot e^{-r(t_1-t_2)}$, then $Y > 0$, and we have constructed an arbitrage opportunity.

Why would your friend every be willing to pay more than $100 \cdot e^{-r(t_1-t_2)}$? The only reasonable answer is that he does not have access to the bank, eg. he is a bad credit risk. In effect you are acting as a *broker* between your friend and the bank. In doing so you are "charging" a premium, $Y > 0$, but you are also assuming some of the risk that your friend does not pay you back. We will assume that all players in the market have equal access to all deals, so in this case, your friend will not pay more than $100 \cdot e^{-r(t_1-t_2)}$. More formally, $X > 100 \cdot e^{-r(t_1-t_2)}$ implies that you have access to an arbitrage opportunity, so we conclude that $X \leq 100 \cdot e^{-r(t_1-t_2)}$.

Exercise 1.11

Show that $X < 100 \cdot e^{-r(t_1-t_2)}$ allows your friend to construct an arbitrage opportunity.

As a result of the exercise, we conclude that $X = 100 \cdot e^{-r(t_1-t_2)}$.

Recap. What have we done? Essentially, we have *priced* a "complex" deal using arbitrage arguments. We examined various scenarios for the "price" X. We then argued that under some scenarios we could construct arbitrage opportunities, hence ruling out these prices. We were then left with only one possible price X that was consistent with no arbitrage. We were lucky to end up with only one *arbitrage-free* price. Sometimes there may be more than one, and sometimes there may be none. If there are none, it means that the mere existence of this deal, together with the other available deals creates arbitrage possibilities.

This activity is a large part of this course, and a large part of what goes on in big banks. Various complex deals are presented to a big bank. Some of them allow that bank to construct arbitrage opportunities[1], which are the favorable deals to the bank. These are the deals it would certainly desire to engage in.

2 The Bond

A contract that **guarantees** *you \$1 at time T is a* **zero coupon bond** *with face value \$1 and maturity T.*

One can buy and sell zero coupon bonds on the open marketplace. The question is how much should one pay today for such a contract. Let's assume that we are in a stationary market, with instanteneous interest rate r. Intuitively, we know that the present value of this contract is e^{-rT}, and so we should not be willing to pay more. Lets see how to prove this using the no arbitrage axiom.

[1] We have discussed pure arbitrage. There is also a notion of statistical arbitrage which becomes more relevent in the context of uncertainty, for example when the interest rate is not certain.

Suppose that the price of this contract is B, and suppose that $B < e^{-rT}$. In this case, borrow B dollars from a bank and buy the bond. At time T, you owe the bank $B \cdot e^{rT}$, but you are also getting a sure \$1 from the bond contract. Thus, your cash flow is $1 - B \cdot e^{rT} = e^{rT}(e^{-rT} - B) > 0$, where the inequality follows from the assumption on B. Lets look at the stream of your cash flows. Cash flow only occurs at the times 0 and T. At time 0, the cash flow is 0, and at time T, the cash flow is > 0. Thus, you have access to a stream of cash flows, all of which are non-negative, and at least one of which is positive. This contradicts the no arbitrage axiom, so $B \geq e^{-rT}$.

Exercise 2.1

Show that the assumption $B > e^{-rT}$ also leads to a contradiction of the no arbitrage axiom. In this case, sell the bond and place e^{-rT} of the proceeds in the bank. Evaluate your cash flows at time 0 and T.

Hence, conclude that $B = e^{-rT}$.

Coupon Paying Bonds. We can generalize our discussion to *coupon paying bonds*. Typically, a bond with face value \$1 has a last payment, and coupons at regular time intervals (typically every 6 months). The coupon payment tends to be analogous to the interest rate and will usually have a value of about $r/2$, so when $r = 0.05$ (corresponding to 5%), a typical coupon payment may be \$0.025. In practice, the face value may be \$100 and the coupon payment will then be \$2.50.

In general, a coupon paying bond can be formalized as a stream of payments $X_1, \ldots, X_N$ at the times $t_1, \ldots, t_N$.

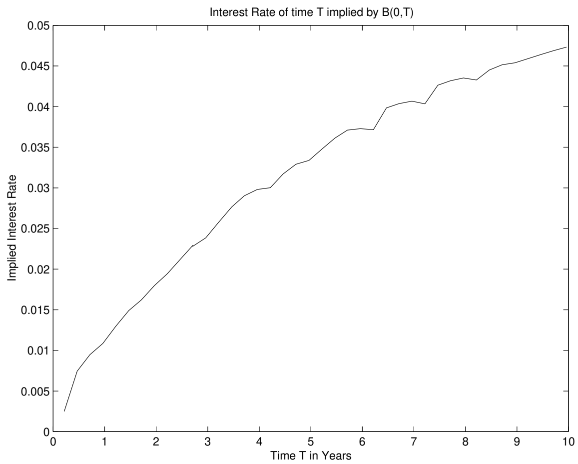
Exercise 2.2

Use an arbitrage argument to show that the price of the coupon paying bond with payments $X_1, \ldots, X_N$ at the times $t_1, \ldots, t_N$ should be

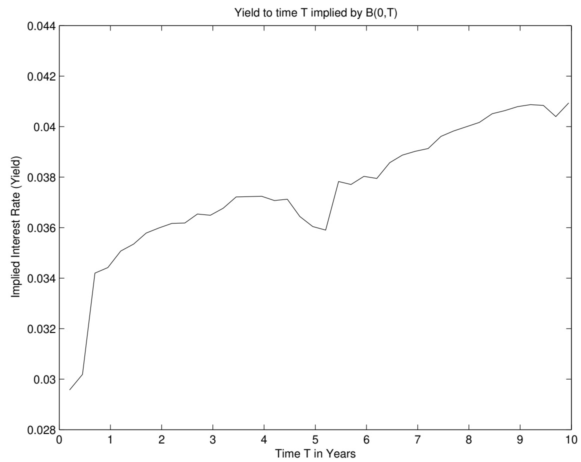
$$\sum_{i=1}^{N} X_i e^{-rt_i}.$$

2.1 The Market Price of Bonds.

While the Federal reserve may set the interest rate r, and banks may have their own lending/borrowing interest rate, the true measure of the value of present consumption versus the value of future consumption is set in the market place, because bonds (both zero coupon and coupon paying) are bought and sold in large quantities every day. Thus, the zero coupon bond with face value \$1 and maturity T has some price in the market. We will generally denote this price by $B(0, T)$, or $B(T)$ for short. By definition, $B(T) = 1/D(T)$, where $D(T) = 1 + R(T)$ is the discount factor defined earlier. Since the yield $\bar{r}(T)$ is defined as $\frac{1}{T} \log D(T)$, the market prices indirectly imply a term structure of interest rates: from the observed market prices $B(T)$, one can compute the implied term structure $\bar{r}(T) = -\frac{1}{T}\log B(T)$. We show two snapshots of this in the figures below, one from the market price data for zero coupon bonds available on August 29 2003, and one from data taken on September 2 2005.



August 29, 2003. (from bondsonline.com)



September 02, 2005. (from bondtrac.com)

These curves can be viewed in some sense as the markets perception on the two specified dates of how the interest rate should behave in the future. Given the zero coupon bond price, we can price any coupon paying bond. More generally, we can price any stream of cash flows.

Exercise 2.3

Suppose that the zero coupon bonds of any maturity are available in the market, priced at $B(0, T)$. Let C be a contract offering a stream of cash flows $X_1, \ldots, X_N; t_1, \ldots, t_N$. Use an arbitrage argument to show that the price of this contract today should be

$$\text{price}(C) = \sum_{i=1}^{N} B(t_i)X_i.$$

If the market were stationary, what would the mathematical form for $B(t)$ be?

(For a contract such as C, we will often not explicitly say price(C), but rather just use C to refer to the contract itself, as well as its price when the context is clear.)

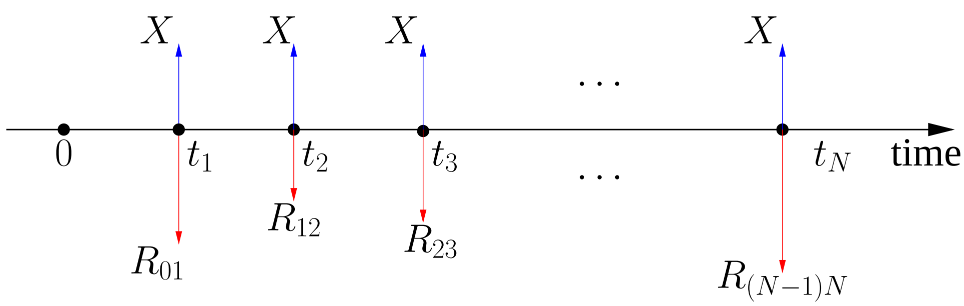
Exercise 2.4

The Interest Rate Swap. Agreeing that the markets are not stationary, the $[t_1, t_2]$ period interest rate $R(t_1, t_2)$ for a fixed period in the future may change with time.

An interest rate swap is the following somewhat bizzare contract.

> A set of times $t_1, \ldots, t_N$ are specified. The fixed payment or *par value* is X. The contract is between two parties. The *receiver* receives the fixed payment X at the times $t_1, \ldots, t_N$ from the *payer*. In return, the receiver pays the payer an amount $R(t_{i-1}, t_i; t_{i-1})$ at time t_i, for $i = 1, \ldots, N$. $R(t_{i-1}, t_i; t_{i-1})$ is the $[t_{i-1}, t_i]$ interest rate determined at time t_{i-1}. For convenience set $t_0 = 0$. The third argument of $R(\cdot)$ indicates that the value of the interest rate for the period $[t_{i-1}, t_i]$ is determined at the time t_{i-1}. We will use the notation $R_{(i-1)i}$ to denote the payment $R(t_{i-1}, t_i; t_{i-1})$.

The reason for calling it a swap should be clear – one is swapping a fixed payment X for a variable payment whose exact value will depend on a particular periods interest rate when that period arrives. So, right now, this interest rate is not known. Hence, from the present point of view, the variable payments are non-deterministic. This is a fairly complicated contract and it may be worth spending some time to understand it, before proceeding. In particular, it can be represented by the following picture.



The blue arrows indicate received payments, and the red arrows indicate payments out. The complication in this contract is that the payments out are not known at time 0. For example the payment R_{12} is only known at time t_1. Nevertherless, in this problem, we will investigate how to price this contract (i.e., determine the *equilibrium par value for* X) using an arbitrage argument. The $R_{(i-1)i}$ are typically called the *floating payments*, since they are not determined *a priori*, and X is typically called the fixed payment. Notice that the payments occur at time t_i, and that the exact value of the payments at time t_i are only known at t_{i-1}, which is when the floating payment gets determined.

1. Consider one of the floating payment $R_{(i-1)i}$. Construct the following "portfolio" which buys a zero coupon bond with maturity t_{i-1} and sells one with maturity t_i.
   (a) Show that the cash flow at time 0 is $B(t_i) - B(t_{i-1})$.
   (b) At time t_{i-1}, show that the value of this portfolio is $1 - B(t_i - t_{i-1}; t_{i-1})$, where the zero coupon value is the value determined at time t_{i-1}.
   (c) Use the relationship between $B(t)$ and $R(t)$ to conclude that the value of this portfolio at time t_{i-1} is

$$\frac{R_{(i-1)i}}{1 + R_{(i-1)i}}.$$

Note that this value is the value at time t_{i-1}, which is not known at time 0.

(d) Now consider the floating payment $R_{(i-1)i}$ which is made at time t_i. This payment is determined precisely at time t_{i-1}. Argue (for example using an arbitrage argument) that the value of this cash flow at time t_{i-1} is

$$\frac{R_{(i-1)i}}{1 + R_{(i-1)i}}.$$

(Remember that this cash flow only occurs at time t_i.)

(e) Use an arbitrage argument to conclude that the value of the floating payment $R_{(i-1)i}$ today must be equal to $B(t_i) - B(t_{i-1})$.

(f) Use an arbitrage argument to conclude that the total value today of all the floating payments is

$$1 - B(t_N).$$

(g) Using the result in the previous exercise, show that the total value of all the fixed payments is

$$X \cdot \sum_{i=1}^{N} B(t_i).$$

(h) Use an arbitrage argument to show that the value of X must be

$$\bar{X} = \frac{1 - B(t_N)}{\sum_{i=1}^{N} B(t_i)}$$

if there is to be no arbitrage opportunity. $\bar{X}$ is usually referred to as the *equilibrium par swap rate*. It is the rate at which one is indifferent between taking on the fixed or floating end of the contract.

2. Typically in the market place, there are published par swap rates and zero coupon rates, and there may be a difference (usually denoted the spread) between the par rate and the formula above that gives the par rate in terms of the zero coupon rates. What are possible explanations for this?

We obtained the "price" of a fairly complicated bond portfolio involving cash flows that were not determined at time 0 using a fairly sophisticated arbitrage argument. One might wonder how one comes up with such arbitrage arguments, it almost looks like magic. Later we will set up a systematic framework using risk neutral (martingale) measures for addressing such topics.

2.2 Bond Portfolio Immunization

Bond portfolios alone present several computational tasks that are generally non-trivial. We will now consider one such task, which is that of immunizing a bond portfolio. To understand the need for this, we need to cast the problem in a practical setting. Specifically, suppose that you have a payment X which is due at time t, and would like to have as little uncertainty as possible at time t. Hence, you would like to set aside some money now, in order to "guarantee" that you will be able to meet your commitment at time t.

The simplest thing to do is to purchase X zero coupon bonds with maturity t, which will cost you $X \cdot B(t)$ now. You now have no need to worry about meeting your commitment because at time t you will owe X which is exactly the amount that your bonds will furnish you with at that time. The problem that you will face when you attempt to implement this strategy is that you will generally not find zero coupon bonds available on the market with maturity exactly at time t; typically, the bonds are issued at discrete times, with discrete time to maturuty (usually bonds are issued quarterly and at time of issue, the time to maturity is usually either 3 months, 1, 5, 10, 30 years). As time passes, the time to maturity for a given bond will decrease, however there will still only be a discrete set of maturities available to you. All of a sudden, your task has taken a non trivial twist.

Let's try to formalize the problem we would like to solve, and at the same time simplify the setting to illustrate the essential issues. Your commitment is X at time t; in the market place are two bonds, with maturities $t_1 < t_2$. Lets for simplicity assume that the market percieves a stationary interest rate, and so there is a well defined instantaneous interest rate r. Thus, the prices of the two bonds are e^{-rt_1} and e^{-rt_2}. For simplicity, your strategy will be to buy α_1 and α_2 units of each bond (you will not hold cash to settle your debt – this more general setting will be treated in an exercise). What are α_1 and α_2 units of each bond worth at time t? We obtain this by multiplying their value today by e^{rt}. Since we need to settle our debt of X, we see that α_1, α_2 must satisfy the *settlement constraint*

$$\alpha_1 e^{-r(t_1-t)} + \alpha_2 e^{-r(t_2-t)} = X.$$

Exercise 2.5

Show that for any choice of α_1, α_2 satisfying the settlement constraint, the present value of the bond portfolio is the same. Why must this be the case?

How should we select α_1, α_2. The market being stationary means that there is a constant rate of interest r. This can be viewed as the perceived rate of interest. After some time Δt, this instantaneous interest rate may change to $r + \Delta r$. The settlement condition for the new interest rate $r + \Delta r$ at time Δt may not be satisfied any more. In fact the degree to which it is not satisfied can be viewed as an instability in our bond portfolio, or as a sensitivity to the interest rate. We can measure the degree to which we are *immunized* to changes in the interest rate by how much the settlement condition is violated. Thus, we consider

$$\epsilon(\Delta r; \alpha_1, \alpha_2) = \alpha_1 e^{-(r+\Delta r)(t_1-t)} + \alpha_2 e^{-(r+\Delta r)(t_2-t)} - X$$

as a measure of how sensitive we are to the interest rate. $\epsilon(\Delta r)$ measures how secure I am in the knowledge that I will be able to meet my debt with minimal additional cash flows at time t. We emphasize that if $\epsilon(\Delta r) > 0$ then it is certainly a good thing for you, as this means you expect to not only cover your commitment but also get cash flow back. However we view the goal as not to maximize profit, but rather to reduce uncertainty. We wish to make $\epsilon(\Delta r)$ a small, if not zero function of fluctuations in the interest rate. Lets formalize this notion. Expanding $\epsilon(\Delta r)$ as a power series, we obtain

$$\begin{aligned}\epsilon(\Delta r; \alpha_1, \alpha_2) &= a_0(\alpha_1, \alpha_2) + a_1(\alpha_1, \alpha_2)\Delta r + a_2(\alpha_1, \alpha_2)\Delta r^2 + \cdots, \\ &= \sum_{k=0}^{\infty} a_k(\alpha_1, \alpha_2)\Delta r^k.\end{aligned}$$

Exercise 2.6

Show that

$$\begin{aligned}a_0(\alpha_1, \alpha_2) &= 0; \\ a_1(\alpha_1, \alpha_2) &= \alpha_1(t - t_1)e^{-r(t_1 - t)} + \alpha_2(t - t_2)e^{-r(t_2 - t)}; \\ a_2(\alpha_1, \alpha_2) &= \tfrac{1}{2}\alpha_1(t - t_1)^2 e^{-r(t_1 - t)} + \tfrac{1}{2}\alpha_2(t - t_2)^2 e^{-r(t_2 - t)}.\end{aligned}$$

Obtain a general formula for $a_k(\alpha_1, \alpha_2)$.

A natural approach to minimizing the sensitivity to interest rate fluctuations is for ϵ to depend only on high order powers of Δr. Since the settlement constraint provides only one constraint, and we have two parameters, we can choose the parameters α_1, α_2 so that the first order dependence disappears and ϵ will fluctuate only in proportion to Δr^2. Thus we minimize the sensitivity to the interest rate by setting

$$a_1(\alpha_1, \alpha_2) = 0.$$

This plus the settlement constraint gives two equations in two unknowns, which can be solved to get *optimally immunized bond portfolio*. The details are an exercise.

Exercise 2.7

Show that the optimally immunized bond portfolio is obtained by setting

$$\begin{aligned}\alpha_1 &= -\frac{\rho}{1-\rho} X \cdot e^{-r(t - t_1)}, \\ \alpha_2 &= \frac{1}{1-\rho} X \cdot e^{-r(t - t_2)},\end{aligned}$$

where $\rho = (t - t_2)/(t - t_1)$.

Thus, while we do not know the interest rate fluctuations, we can minimize our exposure to it. An alternative way of writing the settlement constraint is to use present values,

$$\alpha_1 e^{-rt_1} + \alpha_2 e^{-rt_2} = X e^{-rt}.$$

Correspondingly, one could define the sensitivity ϵ by

$$\epsilon(\Delta r; \alpha_1, \alpha_2) = \alpha_1 e^{-(r+\Delta r)(t_1-\Delta t)} + \alpha_2 e^{-(r+\Delta r)(t_2-\Delta t)} - X e^{-(r+\Delta r)(t-\Delta t)}$$

Exercise 2.8

Show that using the sensitivities defined according to the present value constraint leads to exactly the same optimal portfolio.

The next exercise generalizes exactly this task. Specifically, assume that your commitments are $X_1, \ldots, X_m$ at the times $\tau_1, \ldots, \tau_m$, and in the market place you have access to zero coupon bonds with maturities $t_1, \ldots, t_n$. Assume that the current instantaneous interest rate is r.

Exercise 2.9

In this exercise we will discuss one approach to imunizing the bond portfolio that meets the commitments $(X_1, \ldots, X_m; \tau_1, \ldots, \tau_m)$. Suppose that the bond portfolio which you choose contains α_i units of the bond with maturity t_i.

(a) Show that the settlement constraint is given by

$$\sum_{i=1}^{n} \alpha_i e^{-rt_i} - \sum_{j=1}^{m} X_j e^{-r\tau_j} = 0.$$

(b) At time Δt, assume that the interest rate changes to $r + \Delta r$. We define the sensitivity with respect to fluctuation in the interest rate (Δr) over the time Δt, denoted $\epsilon(\Delta r; \alpha_1, \ldots, \alpha_n)$, to be the amount by which the settlement condition is violated. Show that

$$\epsilon(\Delta r; \alpha_1, \ldots, \alpha_n) = \sum_{i=1}^{n} \alpha_i e^{-(r+\Delta r)(t_i-\Delta t)} - \sum_{j=1}^{m} X_j e^{-(r+\Delta r)(\tau_j-\Delta t)}.$$

(c) Writing $\epsilon(\Delta r; \alpha_1, \ldots, \alpha_n)$ as a Taylor series in Δr, we have

$$\epsilon(\Delta r; \alpha_1, \ldots, \alpha_n) = \sum_{k=1}^{\infty} a_k(\alpha_1, \ldots, \alpha_n) \Delta r^k.$$

Show that

$$a_k(\alpha_1, \ldots, \alpha_n) = \frac{e^{r\Delta t}}{k!} \left[\sum_{i=1}^{n} \alpha_i (\Delta t - t_i)^k e^{-rt_i} - \sum_{j=1}^{m} X_j (\Delta t - \tau_j)^k e^{-r\tau_j} \right].$$

What is a_0?

(d) Argue that since there are n parameters, $\alpha_1, \ldots, \alpha_n$ (one for each bond), and since the settlement constraint provides one constraint, we should be able to set $n-1$ of the coefficients to 0. Which coefficients do you choose to set to 0 and why?

(e) The natural choice which gives optimal immunization is to set

$$a_1 = a_2 = \cdots = a_{n-1} = 0.$$

Define $a'_k(\alpha_1, \ldots, \alpha_n)$ by

$$a'_k(\alpha_1, \ldots, \alpha_n) = \sum_{i=1}^{n} \alpha_i t_i^k e^{-rt_i} - \sum_{j=1}^{m} X_j \tau_j^k e^{-r\tau_j}.$$

Show that $a_1 = a_2 = \cdots = a_{n-1} = 0$ if and only if $a'_1 = a'_2 = \cdots = a'_{n-1} = 0$. Note that this set of equations is *independent* of Δt.

(f) Show that this set of equations, together with the settlement constraint has a unique solution.

Here is an outline of the argument. Let $\beta_k = \sum_{j=1}^{m} X_j \tau_j^k e^{-r\tau_j}$ for $k = 0, \ldots, n-1$, and define $\boldsymbol{\alpha}, \boldsymbol{\beta}$ as follows.

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}, \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{n-1} \end{bmatrix}.$$

Show that the set of constraints can be written in the matrix form

$$\begin{bmatrix} \lambda_1 & \lambda_2 & \cdots & \lambda_n \\ \lambda_1 x_1 & \lambda_2 x_2 & \cdots & \lambda_n x_n \\ \lambda_1 x_1^2 & \lambda_2 x_2^2 & \cdots & \lambda_n x_n^2 \\ \lambda_1 x_1^3 & \lambda_2 x_2^3 & \cdots & \lambda_n x_n^3 \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1 x_1^{n-1} & \lambda_2 x_2^{n-1} & \cdots & \lambda_n x_n^{n-1} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_2 \\ \vdots \\ \beta_{n-1} \end{bmatrix},$$

where $\lambda_i = e^{-rt_i} > 0$ and $x_i = t_i$. More compactly, this can be written as

$$\Lambda \boldsymbol{\alpha} = \boldsymbol{\beta},$$

for an appropriately defined Λ. The task is to now show that Λ is invertible. This can be done by showing that

$$\Lambda = VL,$$

where V is a Vandermonde matrix and L is a positive diagonal matrix,

$$V = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ x_1^3 & x_2^3 & \cdots & x_n^3 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{bmatrix}, \qquad L = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

Clearly L is invertible, and it is well known that the Vandermonde matrix is invertible if and only if all the x_i are distinct. (The energetic may wish to prove this.) Show that in our case, all the x_i are distinct to conclude the proof.

This exercise shows that the immunization problem reduces to solving a linear system involving a Vandermonde matrix. Clearly this system can be solved by inverting the Vandermonde matrix. We refer the interested reader to texts on matrix analysis (eg. [?]) for efficient algorithms to invert Vandermonde matrices. When the bond maturities have the form $t_i = i \cdot t$, then the Vandermonde matrix V has a very special form, and can be inverted in closed form, [?].

(g) Suppose one also held cash. How does our general solution apply even in this setting?

Exercise 2.10

Consider a stream of commitments of $\$100$ at each of the times $1, 2, 3$. Let the interest rate be 0.05, and assume that there are 10 bonds in the market place with maturities $\frac{1}{2}, 1\frac{1}{2}, 2\frac{1}{2}, \ldots, 9\frac{1}{2}$. You would like to build an immunized bond portfolio to guarantee this cashflow stream.

(a) What is the maximum value of k for which you can be immunized to fluctuations up to order Δr^k.

(b) Construct the bond portfolio which is immunized up to order Δr^k for the value of k in the previous part.

(c) Plot the amount by which the settlement constraint is violated as a function of the fluctuation in the interest rate for interest rate fluctuations $\Delta r \in [-0.02, 0.02]$. The settlement constraint violation is given by

$$\epsilon(\Delta r) = e^{(r+\Delta r)\Delta t} \left[\sum_{i=1}^{n} \alpha_i e^{-(r+\Delta r)t_i} - \sum_{j=1}^{m} X_j e^{-(r+\Delta r)\tau_j} \right].$$

Select a value of Δt for your plot, for example $\Delta t = \frac{1}{12}$ corresponds to 1 month.

Exercise 2.11

For the general case of one commitment and n bonds, one could define the settlement constraint and sensitivity ϵ either using the present value or the value at the time of the commitment.

Show that both ways lead to the same optimally immunized portfolio.

Exercise 2.12

A much harder bond portfolio immunization problem arises if there are a large number (hundreds) of bonds with different maturities available, and there are constraints on the size of the bond portfolio. For example, imagine there are 1000 bond maturities, but your bond portfolio is limited to size 3.

Naturally, any 3 bond maturities will serve to immunize the settlement constraint violation against fluctuations of order 2, but what is the best set of 3 maturities to select. One criterion is that the coefficient of the 3rd order fluctuation should be minimized. More generally, suppose that there are n maturities available, but the portfolio is limited to size ℓ.

(a) Formulate the problem as one of finding the optimal set of bonds which immunize the portfolio to fluctuations of order $\ell - 1$ and minimizes the magnitude of the order ℓ coefficient of the fluctuation. Specifically show that the problem reduces to the following mixed integer-linear program (MILP),

$$
\begin{aligned}
&\min z, \quad \text{s.t.}\\
&\alpha_i \in \{0,1\},\\
&\sum_{i=1}^{n} \alpha_i \le k,\\
&\sum_{i=1}^{n} \alpha_i t_i^k e^{-rt_i} - \sum_{j=1}^{m} X_j \tau_j^k e^{-r\tau_j} = 0, \qquad k \in \{0,\ldots,\ell-1\},\\
&\sum_{i=1}^{n} \alpha_i t_i^\ell e^{-rt_i} - \sum_{j=1}^{m} X_j \tau_j^\ell e^{-r\tau_j} \le z,\\
&-\sum_{i=1}^{n} \alpha_i t_i^\ell e^{-rt_i} + \sum_{j=1}^{m} X_j \tau_j^\ell e^{-r\tau_j} \le z.
\end{aligned}
$$

(b) Formulate a brute force strategy for solving this problem, and determine its running time complexity.

(c) Formulate a gready heuristic strategy for solving this problem and determine its run time complexity. Run some experiments to com compare the performance of the greedy with the brute force.

(d) [**hard, open- ended**] Can you formulate a heuristic with a provable approximation guarantee on the quality of the solution with respect to the optimal solution. One approach would be to relax this MILP to an LP and then use a randomized rounding technique to obtain a valid solution. Then show that the expected value of the objective after randomized rounding is within some approximation factor of the objective in the relaxed LP solution. Since the solution to the relaxed LP gives an upper bound on the solution to the MILP, you have your result.

Exercise 2.13

Other important complex contracts derived from bonds are **interest rate caps** and **interest rate swaptions**. We define these complex contracts here, and the energetic reader may want to ponder how to price these contracts.

2.3 Yield Curve Interpolation from Coupon Bond Data

Just as a reminder, the zero coupon price for the bond with maturity t is $B(t)$, and the yield $Y(t)$ is defined by $Y(t) = -\frac{\log B(t)}{t}$. Typically, bonds are only available at discrete maturities. A sample of the daily yields as published by the US treasury department (http://www.ustreas.gov/) is given in the table below.

Date	1 mo	3 mo	6 mo	1	2	3	5	7	10	20	30
01/02/90	–	7.83	7.89	7.81	7.87	7.90	7.87	7.98	7.94	–	8.00
01/03/90	–	7.89	7.94	7.85	7.94	7.96	7.92	8.04	7.99	–	8.04
01/04/90	–	7.84	7.90	7.82	7.92	7.93	7.91	8.02	7.98	–	8.04
01/05/90	–	7.79	7.85	7.79	7.90	7.94	7.92	8.03	7.99	–	8.06
⋮											
08/26/08	1.67	1.71	1.95	2.19	2.35	2.64	3.06	3.37	3.79	4.43	4.40
08/27/08	1.58	1.67	1.93	2.16	2.31	2.58	3.02	3.34	3.77	4.41	4.38
08/28/08	1.63	1.74	1.98	2.19	2.37	2.62	3.09	3.42	3.79	4.41	4.38
08/29/08	1.63	1.72	1.97	2.17	2.36	2.60	3.10	3.45	3.83	4.47	4.43

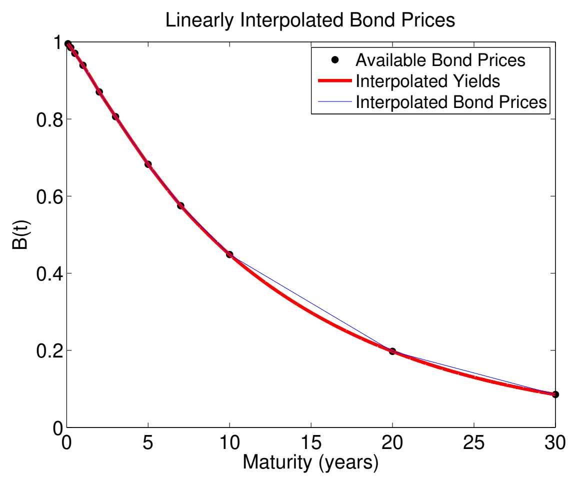
Notice that in the early days, some maturities were not available since the US government was not issuing those bonds. The immediate problem is one of obtaining the present value of a cashflow at a maturity which is not one of the discrete maturities available. One needs to interpolate bond prices in order to obtain prices at other maturities.

The typical properties one would like of any interpolation method are that it be continuous, and that it be local. Local means that the value of the interpolated price at some maturity τ depends only on the available maturities close to τ. An additional bonus is that the interpolated function has a continuous derivative. There are many approaches for interpolation of yield curves [**?**]. The simplest is the linear interpolation which is local, continuous, though not continuously differentiable.

Even with linear interpolation, the results differ depending on whether one interpolates the bond prices $B(t)$ or whether one interpolates the yields $Y(t)$, and obtain the bond prices from the resulting interpolated yield.

For a set of maturities $t_1 \le t_2 \le \cdots \le t_n$ and corresponding values $y_1, y_2, \ldots, y_n$, for a given t such that $t_i \le t < t_{i+1}$, the interpolated value is given by

$$I(t) = y_i \frac{t_{i+1} - t}{t_{i+1} - t_i} + y_{i+1} \frac{t - t_i}{t_{i+1} - t_i}.$$



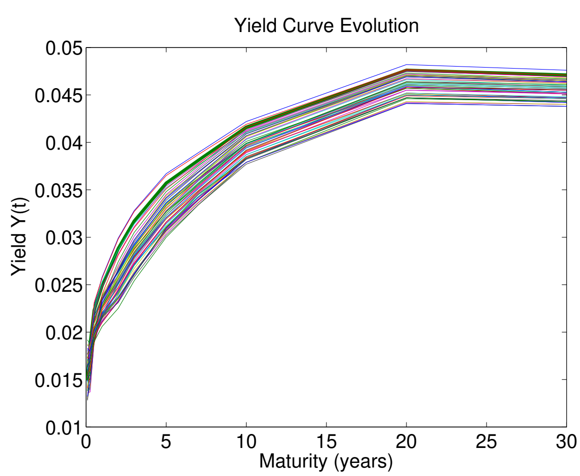
Linear interpolation of the yields gives smoother looking bond prices. One can now use the interpolated bond prices for computing present values.

2.4 Yield Curve Interpolation from Swap Data

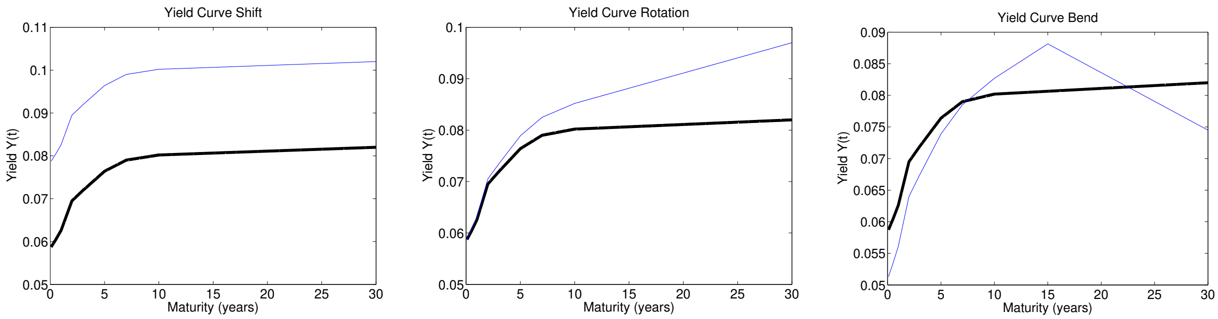
Exercise 2.1 gives a relationship between the par swap rates and the bond prices. Swap rates are also periodically published by the Federal reserve, again at discrete maturities. The swap rates could also be used to infer the yields, which can then be interpolated.

2.5 Data Driven Yield Curve Dynamics

If we look at the yield curve over time, it is definitely not constant. The daily evolution of the yield curve over a short period of time is shown in the next figure. Each line is the yield curve for a particular day.



While the yield curve is not constant, it is believed that the yield curve exhibits only a few main degrees of freedom which capture most of its dynamics. Typically three degrees of freedom are used, often denoted by the shift (the whole curve shifts up or down), the rotation (the whole yield curve rotates) and the bend (the center of the yield curve moves in one direction while the extremes move in the other).



The fact that a small number of degrees of freedom should suffice to describe yield curve dynamics suggests that it should be possible to effectively model yield curve dynamics using some kind of Markov Models.

2.5.1 Modeling Yield Curve Dynamics Using Hidden Markov Models

Hidden Markov Models are very useful for modeling processes in computational finance, and we will use the context of yield curve modeling to introduce them. A Hidden Markov Model (HMM) is a state based model. The market exists in various states, for example the high interest rate state, the low interest rate state, the upward sloping yield curve state, the inverted yield curve state (downward sloping), etc. The actual yield curve depends on the state the market is in. Finally, the market may transition from one state to the next, according to some transition probability distribution.

Suppose that there are K state $\mathbf{S} = \{s_1, \ldots, s_K\}$. If the market is in state s_i, it will transition to state $s_j \in \mathbf{S}$ with probability $P_{ij} \geq 0$, where $\sum_{j=1}^{K} P_{ij} = 1$. The matrix $\mathbf{P}$ is called the state transition matrix.

Given that the state of the market on a particular day is s_i, the observed yield curve $\mathbf{x}$ will be drawn from some (state dependent) distribution $P_i(\mathbf{x})$. The Hidden Markov Model is then fully specified by the initial state (or initial probability distribution for the initial state), the transition matrix $\mathbf{P}$ and the state dependent densities P_i. Let $\pi_i(t)$ be the probability that the state at time t is s_i. Then the probability distribution for the yield curve at time $t+1$ is given by

$$P_{t+1}(\mathbf{x}) = \sum_{i=1}^{K} \pi_i(t) \sum_{j=1}^{K} P_{ij} P_i(\mathbf{x}).$$

$P_{t+1}(\mathbf{x})$ is the probability distribution for the yield curve at time $t+1$. The first sum is over the possible states at time t, weighting each state by the probability of being in that state. The second sum considers all possible states that one can transition to from the state at time t, weighting by the probability of transitioning to that state and the probability of observing $\mathbf{x}$ from the state transitioned into. Note that we can obtain a relationship between $\pi_j(t)$ and $\pi_i(t+1)$:

$$\pi_i(t+1) = \sum_j \pi_j(t) P_{ji},$$

or in more compact vector notation,

$$\boldsymbol{\pi}(t+1) = \mathbf{P}^T \boldsymbol{\pi}(t).$$

Exercise 2.14

Show that if $\sum_i \pi_i(t) = 1$, then $\sum_i \pi_i(t+1) = 1$.

By induction one then has that

$$\boldsymbol{\pi}(t) = (\mathbf{P}^T)^t \boldsymbol{\pi}(0).$$

We will assume that the state dependent yield curve distribution $P_i(\mathbf{x})$ is a Normal distribution. The number of dimensions in the vector $\mathbf{x}$ is equal to the number of maturities for which the yield curve data exists. Thus,

$$\begin{aligned} P_i(\mathbf{x}) &= N(\mathbf{x}; \mu_i, \Sigma_i), \\ &= \frac{1}{(2\pi)^{d/2}|\Sigma_i|^{1/2}} exp\left(-\frac{1}{2}(\mathbf{x}-\mu_i)^T \Sigma_i^{-1}(\mathbf{x}-\mu_i)\right). \end{aligned}$$

The Hidden Markov Model is therefore fully specified by $\boldsymbol{\pi}(0), \mathbf{P}, \{\mu_i, \Sigma_i\}_{i=1}^K$. The first task is to estimate, or *callibrate* these parameters of the HMM to observed market yield curves.

2.5.2 Immunization of Bond Portfolios to Data Driven Yield Curve Dynamics

3 The Stock

A share of stock in a company entitles the owner to a part ownership in the company. The payoff for owning stock in a company typically comes in two forms: dividends (cash flows) that are distributed by the company to share holders and capital gain of the company – the company becomes worth more, hence the value of your part ownership in the company has increased. We will not spend much time discussing the ins and outs of stock valuation, and how one would analyze the financial accounts of a company in order to determine its value. Rather, here we will give a very simple idea of how much one should be willing to pay for a share of stock in a company.

Assume that cashflows in the form of dividends are expected to be distributed yearly in the amounts of $X_1, X_2, \ldots$, and assume that the 1 year interest rate $R(1)$ is known. In this case let P_0 be the be the price of the share now, then using the tools of the previous section, we know that the price P_0 should be the present value of the future stream of cash flows,

$$\begin{aligned} P_0 &= PV(X_1, X_2, X_3 \ldots; 1, 2, 3 \ldots), \\ &= \frac{X_1}{1+R(1)} + \frac{X_2}{(1+R(1))^2} + \frac{X_3}{(1+R(1))^3} + \cdots. \end{aligned}$$

Similarly, the price of the share at time 1, denoted P_1, would be the present value computed at time 1,

$$P_1 = \frac{X_2}{1+R(1)} + \frac{X_3}{(1+R(1))^2} + \cdots .$$

Multiplying P_0 by $1 + R(1)$, we see that

$$R(1)P_0 = X_1 + P_1 - P_0.$$

The RHS is the cash flow after 1 year plus the gain in the stock price (the capital gain). This sum is typically called the economic earnings per share (EPS), which includes the dividend plus any capital gain. Thus we see that

$$\frac{P_0}{EPS} = \frac{1}{R(1)}.$$

The LHS is typically called the price to earnings ratio (P to E ratio). Typically the yearly interest rate $R(1)$ is around 0.05, in which case the P to E ratio of a company should be about 20. Note the use of $R(1)$ assumes the dividends X_i are deterministic. Since companies have non-trivial default risks, one should replace $R(1)$ with the corresponding implied interest rate for the company's bonds. This could be significantly higher than $R(1)$, implying a lower "fair" price to earnings ratio. Companys with a significantly lower P to E ratio that the cair value of $1/R(1)$ could be considered a good buy, of course, assuming that the previous P to E ratio is a good indicator of the future P to E ratio.

The formula above explains a well known behavior in the markets. As the interest rate rises, so the P to E ratio will fall, and since typically the EPS will not rise, this means that the price will fall. Thus it is often the case the the Federal Reserve uses the yearly interest as a way to stimulate or hold back the stock markets, as a way to control inflation.

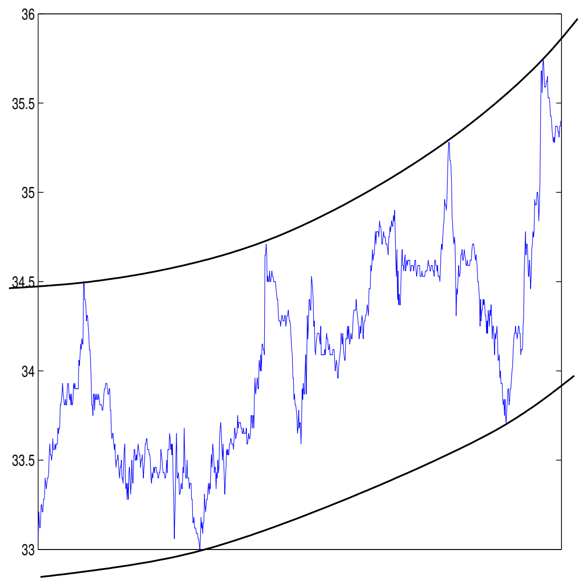
3.1 Predicting Stock Prices

Of course, if we (the author) could do this, then we would stop writing immediately, get rich and go home. Thus, it is likely that your author cannot predict stock prices, however many people believe that it is possible to predict stock prices. For those who are in this category, we will briefly discuss some approaches that one might take. Specifically, on the basis of what indicators should one try to do prediction. We briefly discuss some approaches that people have found useful. Are they profitable? well many of these techniques have been profitable when applied to historical data. The real question is whether they will remain profitable on future data. An alternative way to look at this problem is that the systems that have survived are precisely the ones that are profitable on historical data. If you try many systems, one of them will be profitable, by chance. There is no reason to believe that this system will be profitable in the future. These issues fall into the realm of inference and learning from data. When can one conclude that the best system you have picked with respect to performace on historical data is really a system which is above random.

To drive the point further, suppose that all the systems you try are really just random. Inevitably some of these systems will do better than others on historical data. Picking one of these good performers has still left you with a random system. What has happened? This type of issue is sometimes called *data snooping* or *overfitting*. We will not address these issues here. Rather, we will discuss some indicators that you may want to use in developing a predictor. How you will determine whether your predictor performs well in practice is an issue you undertake at your own risk. In fact, this task is an entire area of research itself, within the realm of statistical learning theory. For an introduction to these issues, see our book *Learning From Data*, by Abu-Mostafa, Magdon-Ismail and Lin.

3.1.1 Plotting or Charting Methods

The basic idea is to plot the price $P(t)$ and try to observe/detect trends. An example is shown below for a small sample of the price data from IBM, a fairly heavily traded stock.



The band seems to indicate both the trending behavior of the price of IBM stock, in addition to defining its region of activity. Extrapolating the upper and lower boundaries into the future in the natural way suggested by the picture, one might expect that a trading strategy which sells when the price approaches the upper boundary and buys when it approaches the lower boundary should be profitable. Along these lines, several tactics are used.

Support and Resistance Levels. Support and resistance levels are "psychological" barriers that the price appears not to break. A resistance is an upper barrier. Typically, the price will approach this upper barrier and then get reflected down. If one could determine a resistance, then it might be possible to make money by shorting the stock when the price approaches the resistance. A support is a lower barrier below which the price is not expected to fall.

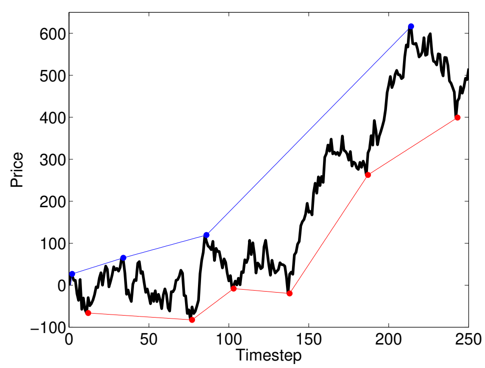
Typical resistances might be the previously attained high, or a round number that is above the current trading price. For example if the current trading price is \$34.50, then a support might be \$35.00. Similarily, a support tends to be the previous low, or a round number below the price. In the previous example, a support might be \$34.00.

The reason for the round numbers is a purely psyghological one, because most people will place limit buy orders at a round number below the trading price and similarly for limit sell orders above the trading price.

Trade Lines. Trade lines are similar to support and resistance in that they try to determine and extrapolate the trading region of a stock price by joining together successive maxima (to obtain the upper trade line) and successive minima (to obtain the lower trade line). This is illustrated with the small sampling from IBM stock price in the picture.

Bollinger Bands. Bollinger bands define the upper resistance curve and the lower support curve as some number of standard deviations above and below a specified moving average indicator (see below for the definition of moving average indicators). The standard deviation is typically computed on the same window length as the moving average.

Exercise 3.15



In this example we will explore an algorithm to construct the lower support curve and the upper resistance curve as illustrated in the above example. As we see, the algorithm reasonably captures what we would intuitively expect of these lower and upper envelopes.

Let $\mathbf{S}_0 = \{s_0, s_1, \ldots, s_T\}$ be the time series of prices at the times $0, 1, \ldots, T$. For simplicity, assume that all the s_i are distinct. A local maximum of $\mathbf{S}_0$ occurs at time t if $s_{t-1} < s_t$ and $s_{t+1} < s_t$. Suppose that the local maxima of $\mathbf{S}_0$ occur at the times $t_1, t_2, \ldots, t_k$. We define the level-1 upper envelope of $\mathbf{U}_1$ as the time series $u_0, u_1, \ldots, u_T$ which linearly interpolates between the points $\{(0, s_0), (t_1, s_{t_1}), (t_2, s_{t_2}), \ldots, (t_k, s_{t_k}), (T, s_T)\}$, which is the sequence of local maxima together with the first and last point.

We define the level-2 upper envelope $\mathbf{U}_2$ as the level-1 upper envelope of the time series U_1. Similarly we define $\mathbf{U}_k$ recursively as the level-1 upper envelope of U_{k-1}.

Analogously, the level-1 lower envelope $\mathbf{L}_1$ is defined as the linear interpolator of the local minima of $\mathbf{S}_0$, and the level-k lower envelope $\mathbf{L}_k$ is the level-1 lower envelope of L_{k-1}.

The higher level envelopes are successively smoothed versions of the lower level envelopes.

(a) Implement this algorithm and show its result on a randomly generated time series. The example shown in this exercise is the level-2 envelope of the time series.

(b) Show that the envelopes are invariant to scaling of the price or time axes. This is a useful property of this approach, i.e., it does not depend on any absolute thresholds or scales.

3.1.2 Quantitative Indicators

Moving Averages. A moving average is an average (over some probability distribution) of the previous prices of the stock. The *resolution* of the moving average is related to the time interval separating the points at which the stock price is sampled (the smaller the time interval, the higher the resolution). To be more specific, let τ be the time interval, and let $p_0, p_1, p_2, \ldots$ be a probability distribution over $0, 1, 2, \ldots$. Then a moving average has the form

$$\begin{aligned} MV(t) &= \sum_{i=0}^{\infty} p_i S(t - i\tau), \\ &= p_0 S(t) + p_1 S(t-\tau) + p_2 S(t-2\tau) + p_2 S(t-3\tau) + \cdots . \end{aligned}$$

Note that this sum typically gets truncated because the time series of stock prices does not go back to $-\infty$. Some typical choices for the probability distributions are

Uniform Fixed Window.

$$p_i = \frac{1}{T+1},$$

for $i = 1 \ldots T$. This choice for p_i corresponds to a arithmetic average over the previous $T+1$ time steps,

$$MV(t) = \frac{1}{T+1} \sum_{i=0}^{T} S(t - i\tau).$$

Typically, one may use such moving averages with different choices of the *window size* T. For example, if $\tau = 1$ trading day, then some useful choices of T are $T = 10$ trading days (2 weeks), $T = 60$ trading days (3 months), $T = 120$ trading days (6 months). A good indication of the market trend migh be obtained by looking at (for example) when the 10 day moving average crosses the 60 day moving average to indicate a change in the trend.

Exercise 3.16

You are given a time series for the stock price, $S_1, S_2, \ldots, S_N$. Assume that $\tau = 1$ and that the window size is T.

(a) For what values of t is $MV(t)$ defined?

(b) Give a *linear time* algorithm (linear in N and independent of T) to compute the time series $MV(t)$ for all values of t on which it is defined.

Exponentially Weighted Moving Average.

$$p_i = (1 - e^{-\lambda})e^{-\lambda i}.$$

This form of a moving average gives more weight to recent values of the stock price, and may be useful when the market dynamics is changing often. λ is often called the exponential decay factor in the weighting. One can roughly equate $\frac{1}{\lambda}$ to an effective fixed window width, since once i becomes greater than $\frac{1}{\lambda}$, the probabilities drop very sharply.

Exercise 3.17

You are given a time series for the stock price, $S_1, S_2, \ldots, S_N$. Assume that $\tau = 1$ and that the exponential decay factor is λ. You may assume that $S_i = 0$ for $i < 1$.

(a) Give an algebraic expression (a summation formula) for $MV(t)$.

(b) For what values of t is $MV(t)$ defined?

(c) Give a *linear time* algorithm (linear in N) to compute the time series $MV(t)$ for all values of t on which it is defined.
(Hint: First show that $MV(t+1) = e^{-\lambda}MV(t) + (1 - e^{-\lambda})S_{t+1}$.)

(d) How would your answers above change if you did not assume that $S_i = 0$ for $i < 1$, i.e., you computed the weighted sum only back till $i = 1$.
(Hint: you need to renormalize the p_i's at each time. Try to do this efficiently to still maintain a linear time algorithm. One approach is to define $A(t) = \sum_{i=1}^{t} e^{-\lambda(t-i)}S_i$, and $B(t) = \sum_{i=1}^{t} e^{-\lambda(t-i)}$. Now relate $MV(t)$ to $A(t)$ and $B(t)$, and show how to efficiently update $A(t)$ and $B(t)$.)

Moving averages are methods for attempting to "smooth" out the noise (random fluctuations), and extracting the true trend. Many people argue that there is predictive power in moving averages, however, our position here is a more neutral one of simply providing the reader with a possible indicator for predictive use.

The Relative Strength Index. The relative strength index R_N is the percentage of *up* moves over the last N moves. The relative strength index also gives some indication of the trend. When $R \geq 70\%$ it seems to indicate that the market has been in an up trend. How to use this predictively may of course depend on the market. Will the market continue to trend up or revert back down? Similarily, when $R \leq 30\%$ it seems to indicate that the market has been in a down trend.

Oscillators. An oscillator is a form of indicator that is commonly known as a *stochastic*. We describe a very simple form of oscillator, κ_N which is computed over the previous N time periods,

$$\kappa_N = \frac{\text{Current Price} - \text{Low}}{\text{High} - \text{Low}}.$$

The high and the low above are computed over the previous N periods. κ_N measures how significant the current trend is.

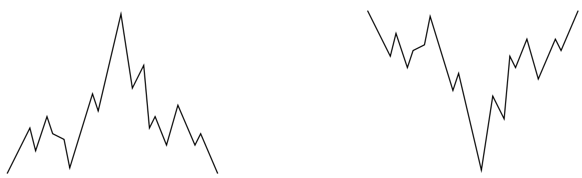
Exercise 3.18

Assume that N is given. Give linear time algorithms to compute the relative strength index and the oscillator as functions of t for a given stock price time series.

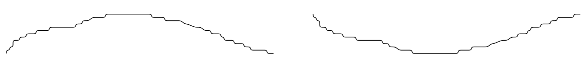
3.1.3 Specific Patterns.

Over time, qualitative traders have recognized the presence of certain patterns. Having extablished the existence of these patterns, one might argue that if the begining of such a pattern is observed, one should be able to predict the behavior by using the remaining part of the pattern as a guide.

Some common patterns are: the **head and shoulders pattern**, and its reverse;



the **saucer top and botom**;



etc.

Exercise 3.19

Given a data base of patterns and a a stock price time series, develop efficient algorithms to extract the set of paterns from the data base that are K_1, K_2-*active* at time t.

We define the active patterns as follows. View each pattern P as a string $P = p_1 p_2, \ldots, p_{|P|}$. Also view the stock price time series up to time t, $S(t) = S_0 S_1 \ldots S_t$ as a string. A pattern P is K_1, K_2-*active* at time t if some prefix of length ℓ, $K_1 \leq \ell \leq K_2$, of the corresponding pattern string P matches the corresponding suffix of length ℓ of the string $S(t)$.

Define the database size D as the total length of all the patterns in the database. Let M_t be the total length of the patterns that are K_1, K_2-active at time t, and let $M = \sum_t M_t$ be the total length of the active patterns for this stock time series (M is the size of the output).

Suppose that you are allowed to preprocess the database of patterns, and that the stock price time series and the value of K_1, K_2 are inputs to the algorithm. With $O(D)$ preprocessing, your algorithm should run in $O((K_2 - K_1 + 1)NK + M)$.

(Hint: You may want to preprocess your database into a suffix tree.)

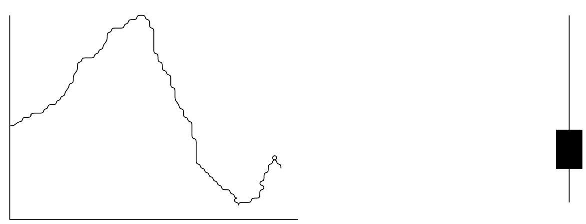
A typical application of such algorithms would be to define the stock price series as a string over a three letter alphabet: down (-1), no significant move (0), up (+1).

The pattern strings are also similarily defined. One then takes all the K-active patterns at a time t. Those patterns whose matching prefix is a proper prefix give a prediction of the future. These patterns can be "voted together" somehow to give a prediction of the future prices of the stock. If, for example, the prediction is sufficiently positive, one might then consider a buy trade.

[**Extension, Open Ended**] Suppose now that exact matches are not required, but certain transformations of a pattern, including noise in the data (approximate matching) should be allowed for. Formalize this problem and try to develop efficient algorithms or heuristics to solve the problem.

The Japanese Candlestick. An interesting pictorial representation of a short period of the stock price series is given by the Japanese candlestick. There are two kinds of candlestick, the solid and the clear.

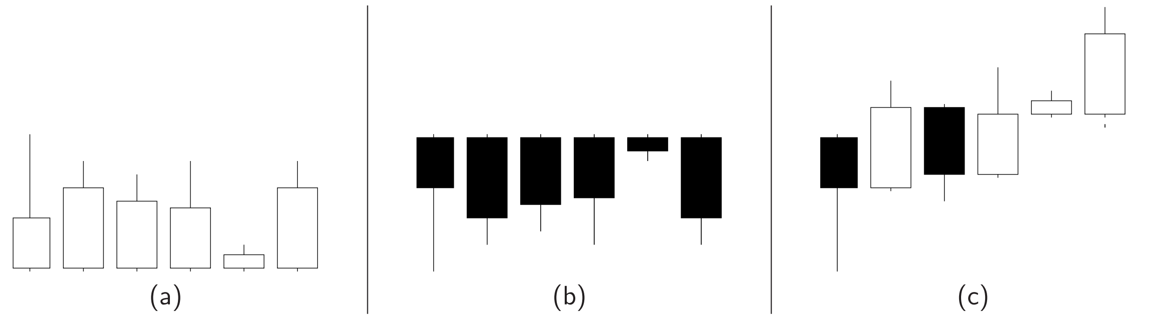
Typically the period represented by the candlestick may be one day. The clear candlestick indicates that the close (C) was above the open (O). The height of the bar indicates the magnitude of the difference $|C - O|$. The high (H) is represented by upper tip of the top line and the low (L) by the lower end of the bottom line. All height differences are proportional to their true values. An example of a market behavior and the corresponding candlestick are given below.



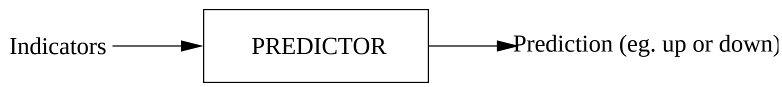
Note that a given market behavior can easily be converted to a candlestick, but the candlestick cannot uniquely be converted into a market behavior. However it gives a good idea, for example if the close is below the open, it probably means that the low is after the high, as this picture would result in smaller overall price movements.

Exercise 3.20

Give the market behaviors that would correspond to the following candlestick sequences,



Recap. The important point we wish to emphasize is the we have presented a set of indicators which are very good at giving us a picture of how the market has been behaving in the past (for example has it been recently trending up, down or neither). The real goal, however would be to predict how the market will behave *in the future*,



Prediction, however is not the content of our discussion. You may, of course at your own risk, take for granted that the market will behave in the future, as it did in the past. On a lucky day you may make some money. Some believe that it should be possible to predict based on such indicators. Some believe you are better off trying to predict the prices tomorrow based on the length of mini-skirts being sported by women today – these people essentially believe that the only way to go is to do due diligence research on a company, its management, its products, its competition, etc, and obtain a better estimate of its EPS than the market has. We wash ourselves of these tasks of prediction because there are plenty of other computational tasks that are eagerly awaiting our attention.

Data Snooping (Overfitting) Before we go on, we will address one important issue, namely that of *data snooping*. Specifically, how to evaluate a predictor.

Suppose that John TinyTrader comes to you with a predictor and shows you that over the last 10 months, it predicted the market movement (up or down) correctly. This person then offers you this system for a price. Do you want it? The problem you face is one of trying to *evaluate* a predictor without knowing the *process* that created the predictor. The warning we give to you is that this is an extremely dangerous task. Your evaluation of the predictor depends heavily on the process that created the predictor. In the absence of any knowledge regarding the process that created the predictor, a prudent way to proceed would be to assume the worst. To illustrate the dilemma, consider the following three distinct processes that could have resulted in John TinyTrader approaching you:

1. John TinyTrader is the only trader in the market. One day he dreams up a system, and immediately gives you the system. You test it on the next 10 months. It works! Mr. TinyTrader now approaches you, and asks how well did my system do, and are you willing to pay for it?

2. There are a large number of TinyTraders on the market, John, Jack, Ann, ... For concreteness suppose that there are about 1024 such TinyTraders, and they all have a different set of predictions for the next 10 months (if any two have the same sets of predictions, we can ignore one of them). Clearly no one will approach you if they made mistakes. John happens to be the TinyTrader who gets it all right. Mr. TinyTrader now approaches you.

3. John TinyTrader is the only trader in the market. One day he dreams up a system and tests it on the next 10 months. It failed to predict the 8th month correctly. John continues testing his system, and lo and behold, it correctly predicts months 9-18 correctly. Mr. TinyTrader now approaches you with his data on the most recent ten months.

Under which process of creating the trading system would you pay the most. If your intuition said system 1 appears worth the most, followed by 3, then 2, then your intuition is not bad. If we rephrase each of these creation processes, perhaps the choice will become clearer:

1. I have a system which I use in the market for the next 10 months. It makes money, this is real money!

2. I have 1024 different systems, I need to pick one. So I look at the performance over 10 months, and luckily for me, one of them did perfectly. Was I really lucky? No! One system is guaranteed to perform well, as there are only 1024 possible outcomes over the next 10 months, and so one of my systems must be perfect. What can I conclude about this system? not much.

3. I have a system. I will continue testing until it performs perfectly for 10 consecutive months. Then I will approach you with my system. You should convince your self that a *random* system will eventually perform perfectly for 10 consecutive months. Don't be fooled by the argument

   > "...but over 19 months, the system predicted 18 months correct."

   It could have just gotten lucky. The fact is this system is guaranteed to be presented to you eventually, and it could be no better than random.

The underlying difference between system 1 and the other two is the notion of choice. In a (not so vague way) systems 2 and 3 are the result of someone being able to *choose* a system from among a set to present. Naturally TinyTrader will select the best. Such issues of how the availability of choices affects your belief in the quality of the system is a topic beyond the scope of our discussion, and falls in the arena of Learning, specifically the sub field of *generalization* – how does the performance of the system you have seen so far generalize to the future. The generalization ability is intimately related to the degree of choice one has in choosing which system to show.

To drive the point home, consider a uniform random variable over the interval $[0, 1]$. If you randomly pick one and show it to me, it will have an expected value of $\frac{1}{2}$ which one might argue is a fair representation of the distribution. If, however, you pick 10 such random numbers independently from the *same* distribution, and then *choose* to show me the largest of these, this is no longer a fair representation of the distribution. In fact the number you show me will on the average be much larger than $\frac{1}{2}$. The next exercise will convince you of this fact.

Exercise 3.21

Show that the expected value of the maximum of n independent samples of a uniform random variable over $[0, 1]$ is $n/(n + 1)$.

Exercise 3.22

[**The Postal Scam.**] You are a gambling man and every week you place a small bet on the Monday night (American) football game. One Firday you got a postcard with only the following on it

Ravens <u>Cardinals</u>

You watch the game on Monday and the Cardinals win. For the next 4 consecutive Fridays, you get similar postcards with the following

Bills <u>Falcons</u>

<u>Panthers</u> Bengals

<u>Bears</u> Browns

Broncos <u>Cowboys</u>

On each corresponding Monday, you watch the game and to your surprise, the sequence of winners is Falcons, Panthers, Bears and Cowboys.

You eagerly await the next postcard on the next Friday, which to your surprise arrives, with the following on it

Colts Packers

You are confused; you turn over the postcard and find a message. Call 1-900-XXX-XXXX to find out which team is underlined. When you call, they ask you for $10 for the answer. You need to make a quick decision. Do you bite (assuming that you can place any sized bet on the Colts-Packers game in Las Vegas).

More specifically, what is the maximum you would be willing to pay for this information. You may make some reasonable assumptions on the price of a postcard stamp.

Does your answer to the above problem change if there is a bet-limit of $100 that you can make on the game in Las Vegas?

The moral is that while we do not engage in the task of prediction, we warn that to evaluate a predictor carefully, it is imperative that one knows the process that created the predictor. Trying only to evaluate it based on the resulting performance is flawed. In a conservative world, one should assume the worst for the process that created the predictor.