Lecture Notes in Finance 2 (MiQE/F, MSc course at UNISG)

Paul Söderlind[1]

26 December 2025

[1]University of St. Gallen. *Address:* s/bf-HSG, Unterer Graben 21, CH-9000 St. Gallen, Switzerland. *E-mail:* Paul.Soderlind@unisg.ch. Document name: Fin2All.TeX.

Contents

Chapter 14

Foreign Exchange

This chapter discusses some special properties of exchange rates and how they impact investments abroad. In particular, it studies deviations from uncovered interest rate parity (UIP), carry trades, and how currency risk can be hedged.

14.1 Investing in Foreign Currency

14.1.1 The Return from Holding Currency

Investing in a foreign currency typically means that you buy that currency, lend on the foreign money market and eventually buy back domestic currency. To define the return, let S_t be today's price, measured in domestic currency, of one unit of foreign currency. This means that we treat foreign currency as any other asset. Also, let R^*_{ft} be the foreign risk-free rate between $t-1$ and t. The *return*, measured in domestic currency, is then

$$R_t = (1 + R^*_{ft})\frac{S_t}{S_{t-1}} - 1. \tag{14.1}$$

Remark 14.1 *(*Details of the currency return R_t) In $t-1$, invest S_{t-1} (of domestic currency) to buy one unit of foreign currency and lend it on the foreign money market. After one period you have $1 + R^*_f$ units of foreign currency, which buys $(1 + R^*_f)S_t$ units of domestic currency (this is the payoff). The gross return is payoff/investment, which is $(1 + R^*_f)S_t/S_{t-1}$. See Figure 14.1.*

The return of the foreign investment *in excess of the domestic risk-free rate* is then

$$R^e_t = (1 + R^*_{ft})\frac{S_t}{S_{t-1}} - (1 + R_{ft}). \tag{14.2}$$

$t-1$ $\qquad\qquad\qquad\qquad$ t

(a) buy 1 foreign currency unit: pay S_{t-1}
(b) lend on foreign money market

(a) collect $1+R_f^*$
(b) sell foreign currency:
get $(1+R_f^*)S_t$

Return: $R_t = (1+R_f^*)S_t/S_{t-1}-1$

UIP: $\mathrm{E}_{t-1}R_t = R_f \Leftrightarrow$
$\mathrm{E}_{t-1}S_t/S_{t-1} = (1+R_f)/(1+R_f^*)$

Figure 14.1: Return on currency investment

Clearly, an appreciation of the foreign currency (or a depreciation of the domestic currency), along with a high foreign and low domestic risk-free rate, positively impacts the return.

Example 14.2 *With* $(S_{t-1}, S_t, R_f^*, R_f) = (1.20, 1.25, 0.06, 0.04)$

$$R_t^e = (1+0.06)\frac{1.25}{1.20} - (1+0.04) = 0.064.$$

With $S_t = 1.20$ *the excess return is* $0.06 - 0.04 = 0.02$*. Instead with* $S_t = 1.177$ *the excess return is close to zero*

$$R_t^e = (1+0.06)\frac{1.177}{1.20} - (1+0.04) \approx 0.$$

Remark 14.3 *(*Indirect exchange rate quotation) If you instead work with exchange rate quotes that use the number of foreign currency units needed to buy one domestic currency unit,* $\tilde{S}$*, then replace* S *by* $1/\tilde{S}$ *in the previous equations. Appendix 14.5 discusses details.*

In practice, *risk-free returns are from zero-coupon bonds* (bills). We can thus rewrite $1+R_{ft}$ in terms of an interest as

$$1+R_{ft} = (1+Y_{t-1})^m, \tag{14.3}$$

where Y_{t-1} is an effective *interest rate* determined in $t-1$ and m is the fraction of a year between date $t-1$ and t, for instance, $m = 1/12$ for a month. Notice that the interest rate is dated $t-1$, since we know already then how much we earn on the bond between $t-1$ and t. For the foreign market, we have $1+R_{ft}^* = (1+Y_{t-1}^*)^m$.

Using this in (14.2) gives the excess return on the foreign investment as

$$R_t^e = (1 + Y_{t-1}^*)^m \frac{S_t}{S_{t-1}} - (1 + Y_{t-1})^m. \tag{14.4}$$

14.1.2 Covered Interest Rate Parity

To avoid arbitrage opportunities, the forward price in $t-1$ for delivery of one unit of foreign asset in t must obey

$$F_{t-1} = \frac{1 + R_{ft}}{1 + R_{ft}^*} S_{t-1}. \tag{14.5}$$

This is an application of the spot-forward parity, which for the FX market is often called covered interest rate parity (CIP). The *forward premium*, $F_{t-1}/S_{t-1} - 1$, reflects the interest rate difference: a higher value means that the domestic interest rate is higher than the foreign. See Sercu (2009) for more details. In practice, there are some deviations from CIP, depending on which interest rates (secured or unsecured?) that are used.

Example 14.4 *Using the same numbers as in Example 14.2 we get*

$$F_{t-1} = \frac{1 + 0.04}{1 + 0.06} \times 1.20 \approx 1.177.$$

Proof (of (14.5)) Replace the risky strategy in (14.2) by "locking in" the FX rate with a forward contract (replace S_t by F_{t-1}) to get $(1 + R_{ft}^*)\frac{F_t}{S_{t-1}} - (1 + R_{ft})$. This risk-free excess return must be zero, or else arbitrageurs step in. Rearrange as (14.5). $\square$

Remark 14.5 *(Alternative expression of (14.2)) Use CIP to rewrite the excess return (14.2) as* $R_t^e = (1 + R_{ft})(S_t/F_{t-1} - 1)$. *This is sometimes approximated by* $S_t/F_{t-1} - 1$.

14.1.3 Uncovered Interest Rate Parity

The uncovered interest rate parity (UIP) says that the expected exchange rate ($\mathrm{E}_{t-1} S_t$) must be such that the *expected excess return* from the currency speculation in (14.2) is zero

$$\mathrm{E}_{t-1} R_t^e = 0. \tag{14.6}$$

This means that investing on the foreign money market (and then changing back to the domestic currency) has the *same expected returns* as investing on the domestic money market—in spite of having different risks. This would happen if investors are risk-neutral or if currency returns have no systematic risk. Notice that this is very different from CIP, which only rules out arbitrage opportunities and says nothing about expectations or risk. A somewhat more flexible form of UIP would add a constant risk premium to (14.6).

A zero expected excess return in (14.2) means that we must have

$$\frac{\mathrm{E}_{t-1} S_t}{S_{t-1}} = \frac{1 + R_{ft}}{1 + R_{ft}^*}. \tag{14.7}$$

Using CIP (14.5), we can also write UIP (14.7) as $\mathrm{E}_{t-1} S_t = F_{t-1}$, that is, the expectation should equal the forward price.

UIP thus says that the foreign currency is expected to appreciate (that is, $\mathrm{E}_{t-1} S_t / S_{t-1} > 1$) if the foreign interest rate is lower than the domestic. In this way, the foreign investment gains from the (expected) exchange rate movement, but loses from the interest rate—leaving the (expected) return the same as in the domestic market.

Example 14.6 *(UIP) Using the same number as in Example 14.2, UIP says that*

$$\mathrm{E}_{t-1} S_t = 1.20 \times \frac{1 + 0.04}{1 + 0.06} = 1.177,$$

so the foreign currency is expected to depreciate as it has a higher interest rate.

Empirical evidence is mixed but often reveals considerable deviations from UIP, potentially due to either *(a)* significant shifts in risk premia over time or *(b)* systematic disparities between expectations and historical exchange rate movements, including large surprises or even non-rational expectations.

Empirical Example 14.7 *Figure 14.2 shows results from regressions of exchange rate depreciations ($S_t / S_{t-1} - 1$) on the lagged forward premium ($F_{t-1} / S_{t-1} - 1$). The slope coefficients are mostly far from one (1), and negative in most cases.*

14.1.4 Carry Trade

A common FX strategy is to borrow a low interest rate currency (CHF and JPY?), buy a high interest rate currency (AUD?) and lend on its money market. This is called a *carry trade*. The excess return of a carry trade is given by R_t^e in (14.2). However, a carry trade need not borrow the domestic currency. For instance, a US investor could borrow JPY and lend AUD.

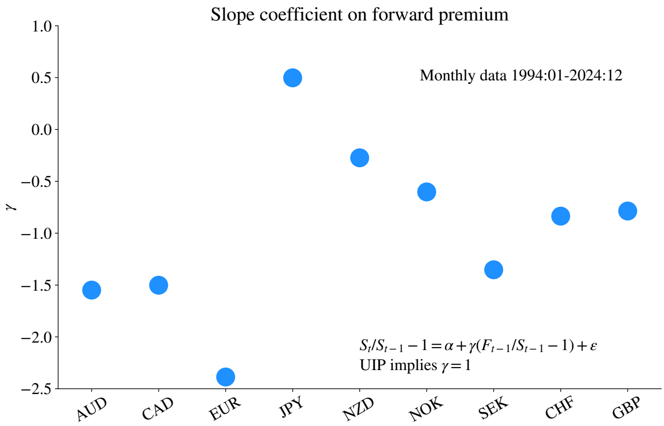


Figure 14.2: Regression of exchange rate changes on forward premia

This strategy has a positive return if the high (low) interest rate currency depreciates (appreciates) less than suggested by UIP, but clearly also carries the risk of the opposite happening. Empirical evidence suggests that carry trades have generated positive average returns, but are exposed to (intermittent) dramatic losses.

Empirical Example 14.8 *Figures 14.3–14.4 illustrate the performance of a monthly carry trade implemented on 10 key currencies. The strategy performs much better than an equally weighted investment in all currencies (financed by borrowing USD), but suffers in periods of high uncertainty (as measured by VIX).*

14.2 Currency Risk in Foreign Investments

We now consider an investment in a *risky foreign asset*. The definition of the return is similar to (14.1), except that we replace the safe foreign return (R^*_{ft}) with a risky foreign return (R^*_t).

The gross return (measured in domestic currency) of this investment is

$$1 + R_t = (1 + R^*_t)\frac{S_t}{S_{t-1}}. \tag{14.8}$$

Take logs to get the log return

$$r_t = r^*_t + \Delta s_t, \tag{14.9}$$

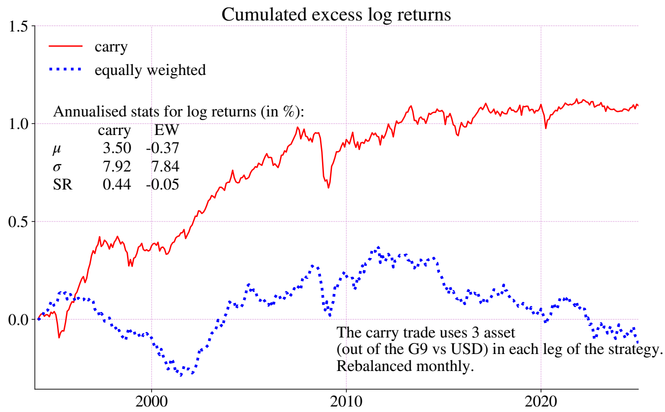


Figure 14.3: Cumulated return on currency investment

where r_t^* is the log foreign return, $\ln(1 + R_t^*)$, and Δs_t is the change of the log exchange rate, $\ln(S_t/S_{t-1})$. Notice that our investor gains if the *(a)* foreign asset (equity?) increases in value ($r_t^* > 0$) and *(b)* if the foreign currency increases in value (appreciates) relative to the domestic currency ($\Delta s_t > 0$). See Figure 14.5 for an empirical illustration.

Example 14.9 *(Investing abroad) Consider a US investor buying British equity in period* $t-1$*:* 5.5 *GBP per British share × 1.6 USD per GBP = 8.8 USD,and selling in* t*:* 5.1 *GBP per British share × 1.9 USD per GBP = 9.69 USD. The gross return for the US investor (in USD) is* $1 + R = (1 - 0.073) \times (1 + 0.188) = 1.10$*. Taking logs gives* $\ln(1 + R) = 0.096$.

From (14.9) the mean and variance of the log return are

$$\mathrm{E}\, r_t = \mathrm{E}\, r_t^* + \mathrm{E}\, \Delta s_t \tag{14.10}$$

$$\mathrm{Var}(r_t) = \mathrm{Var}(r_t^*) + \mathrm{Var}(\Delta s_t) + 2\,\mathrm{Cov}(r_t^*, \Delta s_t). \tag{14.11}$$

Notice that a negative covariance (the foreign local return is high at the same time as the foreign currency depreciates) may reduce the variance of the return measured in domestic currency.

Empirical Example 14.10 *See Tables 14.1 and Figure 14.5 for an empirical illustration.*

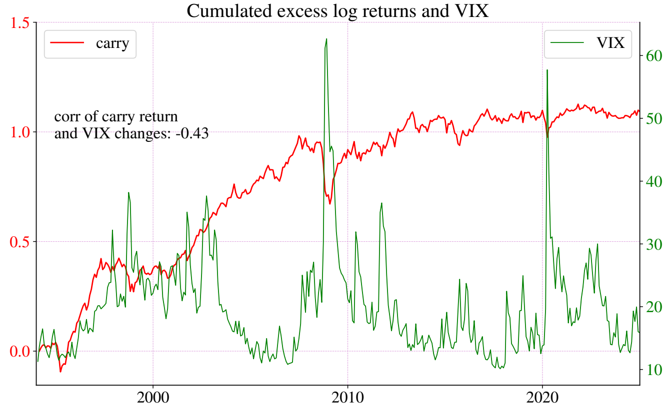


Figure 14.4: Return on currency investment, see Figure 14.3 for details

14.3 Hedging Exchange Rate Movements

International equity or bond investments often involve considerable exchange rate risk. It may be useful to hedge that risk. For instance, the investment strategy may be based on industry analysis ("pick promising pharma companies across the globe"), while the currency exposure is just unwanted risk which requires a different type of analysis. Unless the covariance is very negative (as discussed above) this may motivate hedging the currency exposure.

The most common ways of hedging the exchange rate risk involve forward and option contracts (mostly for short horizons) or swap contracts (longer horizons). Alternatively, a partial hedge is achieved by financing the investment by borrowing on the foreign market. In that way only the profit, not the entire investment, is exposed to exchange rate risk.

To illustrate how a forward contract might help, suppose we could lock in the period t exchange rate by entering a forward contract in $t-1$. If so, the return of the foreign investment (but measured in domestic currency) changes from (14.8) to

$$1 + R_t^{hedged} = (1 + R_t^*)\frac{F_{t-1}}{S_{t-1}}, \tag{14.12}$$

where the currency risk is eliminated.

The practical problem with (14.12) is that the foreign return, R_t^*, typically is not known in $t-1$, so we do not know how many units of currency to hedge via forward

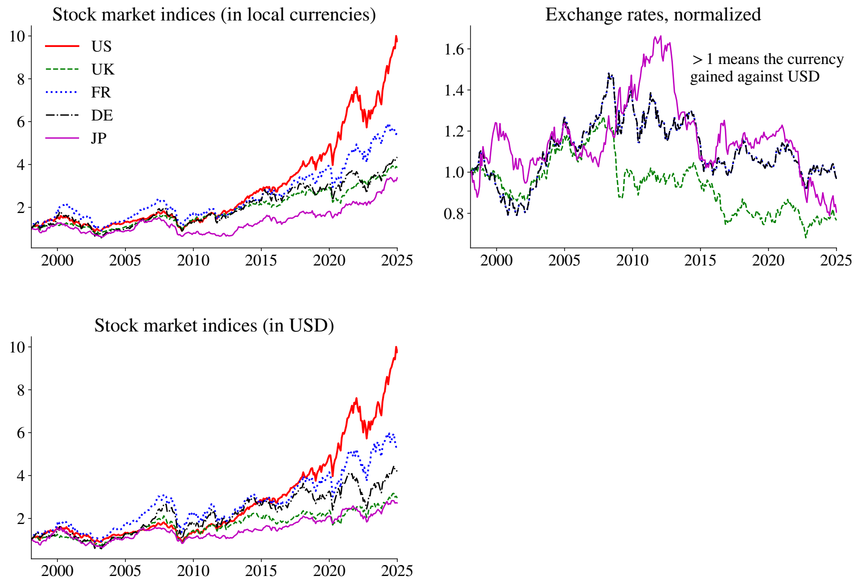


Figure 14.5: International stock market indices and exchange rates

contracts. One possibility is to only hedge the investment, in which case the right hand side changes to $F_{t-1}/S_{t-1} + R_t^* S_t/S_{t-1}$ so only the foreign return is exposed to currency risk.

Remark 14.11 *(*(14.12) when the foreign return is risk-free) Use the forward-spot parity (14.5) to substitute for the forward price in (14.12) to see that the currency-hedged return on the foreign bond market equals the domestic risk-free rate.*

14.4 Explaining Exchange Rates

Financial models of exchange rate returns focus on their correlations (betas) with commonly used market indices. Typically, it is found that exchange rates are fairly disconnected from other markets. This has prompted other finance approaches, including Ranaldo and Söderlind (2010) who demonstrate that the link to other markets depends on market volatility, and Lustig, Roussanov, and Verdelhan (2011) who show how currencies are related to different FX market factors.

Another set of models, mostly from macro and trade economics, focus on the "fundamentals" of currencies, like trade balance and transaction needs. These are

	Local currency	Exchange rate	2*Cov	in USD
		Contribution to average		
US	8.46	0.00		8.46
UK	5.02	−0.99		4.02
FR	6.35	−0.12		6.23
DE	5.45	−0.12		5.33
JP	4.52	−0.80		3.72
		Contribution to variance		
US	2.49	0.00	0.00	2.49
UK	1.82	0.72	0.18	2.73
FR	3.13	0.86	0.31	4.29
DE	4.24	0.86	0.32	5.42
JP	2.97	1.06	−1.30	2.72

Table 14.1: Contribution to the average and variance (annualized, in %) of the log returns for a US investor investing in different foreign equity markets, 1998:01-2024:12

briefly discussed in later sections.

Empirical Example 14.12 *(CAPM regressions for FX returns.) Figure 14.6 shows results from regressing FX excess returns on the U.S. equity market (as a proxy for the global equity market). The evidence suggests that* (a) *many currencies are positively related to the equity market; (b) but the link is fairly weak (R^2 values are low); and* (c) *there are some important cross-sectional differences where (traditionally) low interest currencies (JPY, CHF) are less correlated with the equity market than other currencies. The second subfigure suggests that CAPM still give moderate pricing error, on average over time.*

14.4.1 Purchasing Power Parity and the Real Exchange Rate?

The basic idea of the purchasing power parity (PPP) is that a product should *cost the same at home and abroad* (when measured in a common currency). If this is not the case, then goods arbitrage (buying goods in the cheap country and selling them in the expensive one) will take place, driving down demand for the currency of the more expensive country which leads to an depreciation of its currency.

The strong assumption about goods arbitrage can be relaxed by instead assuming that goods may differ across countries, but that the import/export demand is

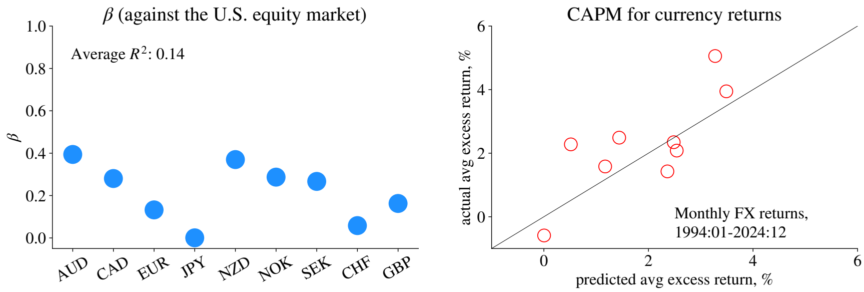


Figure 14.6: βs of currency excess returns, measured in USD

somewhat price elastic. The *real exchange rate* (the relative price of foreign and domestic goods, measured in the domestic currency) is often used as an indicator of the competitiveness of a country. If the domestic price is too high, then export will decrease and import will increase, leading to a trade deficit. This puts pressure on the exchange rate in the same way as discussed above. The mechanism is thus that the real exchange rate puts pressure on the (nominal) exchange rate.

Empirical tests strongly refutes this set of theories for price and exchange rate *levels*, but may work reasonably well for changes over the long run (10+ years). In particular, it points at the important link between inflation (which drives up prices) and depreciations, which is a well established fact over longer runs. In the short run, the causality seems to be the reverse: (nominal) exchange rate movements cause movements in the real exchange rate (competitiveness). Also, it is observed that price levels, when measured in a common currency, are higher in wealthier countries. Once we adjust for that, we get a better measure of over/under valuation of the currency.

14.4.2 Interest Rates?

The exchange rate often appreciates when the central bank raises the interest rate. This typically happens very quickly. One possible explanation is financial flows: if international investors want to benefit from the higher interest rates, then they first need to buy the currency. However, according to UIP, the interest rate hike causes an immediate appreciation, followed by a slow depreciation.

Empirical tests suggests that high interest rate currencies can continue to appreciate for several years (this forms the basis for carry trades), but that they typically eventually suffer a sudden depreciation.

14.4.3 Transactions? (Business Cycles and Financial Flows)

The business cycle theory for exchange rates goes back to first principles to ask the question: why do we hold a currency (cash or cash-like assets)? After all, cash is typically not a good savings instrument (cash is eroded by inflation and there are typically better investment vehicles).

Instead, the key use of a currency is that it facilitates transactions, which suggests that both business cycle conditions (which drive the transaction volumes for goods and services) and financial flows are the most important factors behind exchange rates.

Empirical tests of these models find that also they have some explanatory power over longer horizons.

14.5 Appendix: Exchange Rate Quotation*

14.5.1 Direct and Indirect Quotation

An exchange rate is the price of one currency in terms of another currency. There are clearly *two ways of quoting* this price: the price (measured in domestic currency) of one unit of foreign currency ("direct quotation"), or the price (measured in foreign currency) of one unit of domestic currency ("indirect quotation"). A reasonably established set of quotations and symbols exists in the interbank market, but in other settings, either type of quotation is possible; one should always verify.

Example 14.13 *As an example, Datastream/Refinitiv defaults to reporting "how many USD you pay for one AUD", but "how many CAD you pay for one USD".*

Example 14.14 *For a Swiss investor in 2014, a direct quotation mean that EUR 1 cost CHF 1.2 ("EUR 1 = CHF 1.2"), and an indirect quotation that CHF 1 cost EUR 0.8333 (sometimes written as "CHF 1 = EUR 0.8333").*

Remark 14.15 *(The meaning of CHF/EUR) These lecture notes follow the convention that CHF/EUR (or $S^{CHF/EUR}$) denotes how many CHF you have to pay for each EUR, for instance, $S^{CHF/EUR} = 1.2$. Clearly, $S^{x/y} = 1/S^{y/x}$, for instance, $S^{EUR/CHF} = 0.8333$. (In contrast, the interbank FX market often use EURCHF to denote the same thing, that is, how many CHF you pay for one EUR.)*

Remark 14.16 *(*Currency codes, according to ISO 4217) USD, EUR, JPY, GBP, AUD, CAD, CHF, CNY (Chinese yuan), SEK (Swedish krona), MXN (Mexican peso).*

14.5.2 Cross Rates

Exchange rate across "smaller" currencies are often established indirectly and are therefore called "cross rates": as a combination of two trades. For instance, suppose you own CHF and want to buy CAD (Canadian dollars). It may well be that this involves two trades: use the CHF to buy USD and then use the USD to buy CAD. (Even 15 years after the collapse of the Bretton-Woods system in the early 1970s almost all currency trades went via the USD. Since then there are more direct trades, but trade via the USD still dominates.)

Example 14.17 *(The implicit trade in a cross rate) (a) Buy one USD, costs* 0.95 *CHF; (b) use the one USD to buy* 1.25 *CAD; (c) in total you have paid* 0.95 *CHF and got* 1.25 *CAD. Therefore, the implied price (in CHF) per AUD is* $0.95/1.25 \approx 0.76$. *(You can memorize this as "CHF/USD×USD/CAD=CHF/CAD") See Figure 14.7 for an illustration.*

Remark 14.18 *(*The implicit trade in a cross rate, using* $S^{x/y}$ *notation) In the previous example,* $S^{CHF/USD} = 0.95$ *and* $S^{USD/CAD} = 1/1.25$, *so* $S^{CHF/USD} = S^{CHF/USD} S^{USD/CAD} = 0.95/1.25 \approx 0.76$. *In general, cross rates mean that* $S^{x/y} = S^{x/z} S^{z/y}$.

If there is a way to trade without going through another currency (and there typically is), then the price on this market should be very close to the cross rate. If not, there would be an arbitrage opportunity.

14.5.3 Log Rates

A fair amount of exchange rate analysis is done in terms of log rates. For that reason, this section summarizes how the previous expressions look like in logarithmic terms.

Remark 14.19 *(Log FX returns, (14.1)–(14.2)) Let* r_t *be the log return,* $\ln(1 + R_t)$. *From (14.1), it can be written* $r_t = r^*_{ft} + \Delta s_t$, *where* r^*_{ft} *is the log foreign gross risk-free rate,* $\ln(1+R^*_{ft})$, *and* Δs_t *is the relative change of the exchange rate,* $\ln S_t/S_{t-1}$. *Subtract* $r_{ft} = \ln(1 + R_{ft})$ *to get the excess log return* $r^e_t = \Delta s_t + r^*_{ft} - r_{ft}$, *which is the log version of (14.2).*

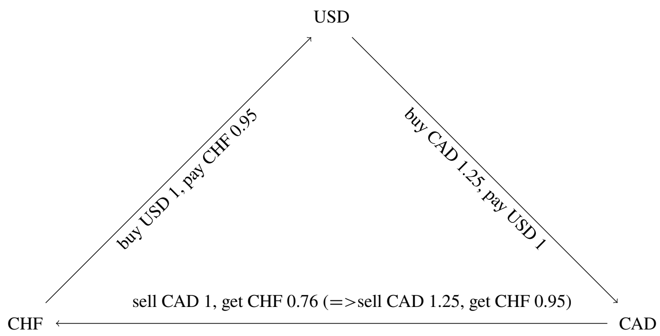


Figure 14.7: Cross-rates

Remark 14.20 *(Log FX returns, (14.4)) Equation (14.3) can be used to rewrite the excess log return in Remark 14.19 as* $r_t^e = \Delta s_t + m(y_{t-1}^* - y_{t-1})$*, where* $y = \ln(1 + Y)$*. This is the log version of (14.4).*

Remark 14.21 *(Log FX returns) Take logs of (14.5), rearrange and use in the excess log return in Remark 14.19 to get* $r_t^e = \Delta s_t - (f_{t-1} - s_{t-1}) = s_t - f_{t-1}$*, which is the log version of the result in Remark 14.5. Also, the interest rate differential can be written* $r_{ft}^* - r_{ft} = s_{t-1} - f_{t-1}$*.*

Chapter 15

Forwards and Futures

This chapter provides details on forwards, futures, and swaps. These instruments are important in themselves, but will also play important roles in (later) discussions of bonds and options. One of the key results of the chapter is the forward-spot parity, which links the pricing of forward contracts to the current spot price.

15.1 Derivatives

Remark 15.1 *(On the notation) The notation is kept short. The current period is assumed to be* $t = 0$ *and the derivative expires in* $t = m$*, which means* m *years later. Time subscripts and indicators of time to maturity are typically suppressed, unless strictly needed in the context. For instance, instead of* $F_0(m)$ *we often use* F *denote the forward price (contracted in* $t = 0$*, expiring in* $t = m$*) and similarly for interest rates (*y *instead of* $y_0(m)$*). Also, instead of* S_0 *we use* S*, but we keep the subscript on* S_m *which is the value of the underlying asset at expiration of the derivative.*

Derivatives are assets whose payoff depend on some underlying asset (for instance, the stock of a company). The most common derivatives are futures contracts (including forward contracts) and options. Derivatives are in zero net supply, so a contract must be issued (a short position) by someone for an investor to be able to buy it (long position). For that reason, gains and losses on derivatives markets sum to zero.

15.2 Present Value

The present value of Z units paid m periods (years) into the future is

$$\mathrm{PV}(Z) = (1+Y)^{-m} Z, \text{ or} \tag{15.1}$$

$$= e^{-my} Z, \tag{15.2}$$

where Y is effective spot interest rate on an m-period loan, and y is the continuously compounded m-period interest rate ($y = \ln(1+Y)$). As usual, the *interest rates are expressed on an annual basis*; hence, m should measure time in years. For instance, $m = 1/4$ means a quarter of a year (3 months).

Example 15.2 *(Present value) With $y = 0.05$ and $m = 3/4$ we have the present value $e^{-3/4 \times 0.05} Z \approx 0.963Z$.*

15.3 Forward Contracts

15.3.1 Definition of a Forward Contract

A forward contract specifies, among other details, the expiration date, which asset should be delivered (the "underlying asset"), and the agreed payment for it, referred to as the forward price F. See Figure 15.1 for an illustration of the timeline.

The profit (payoff) of a forward contract *at expiration* is straightforward to calculate. Let S_m be the spot market price of the underlying asset at expiration (in m). Then, for the *buyer* of a forward contract the

$$\text{payoff of a forward contract} = S_m - F. \tag{15.3}$$

The reason is that, at expiration, the owner of the forward contract pays F to get the asset which is worth S_m. See Figure 15.2 for an illustration of the payoff (at expiration) as a function of the underlying price, S_m. (The payoff function will look more interesting for options.) Similarly, the payoff for the *seller* (or issuer) of a forward contract is $F - S_m$ (she buys the asset on spot market for S_m, gets F for asset according to the contract). This sums to *zero*, irrespective of the value of the underlying asset.

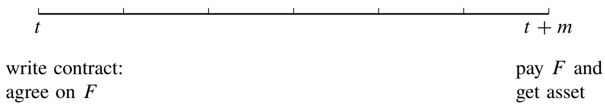


Figure 15.1: Timing convention of forward contract

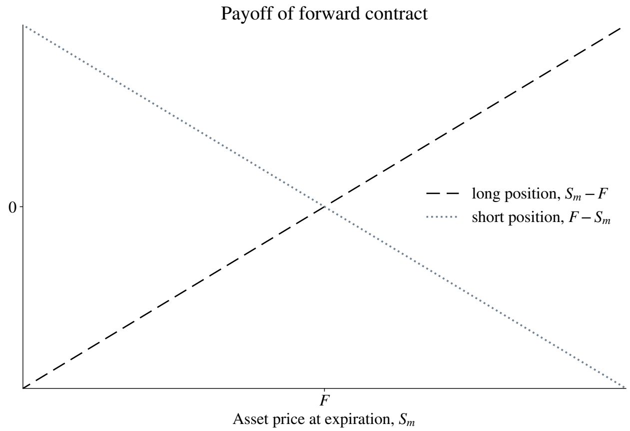


Figure 15.2: Payoff of forward contract at expiration

15.3.2 Forward-Spot Parity

A forward contract entails both a right (to get the underlying asset at expiration) and an obligation (to pay the forward price at expiration), so it is perhaps not obvious what the value of it is. However, in the absence of trading costs, a no-arbitrage argument shows that the following proposition must hold.

Proposition 15.3 *(Forward-spot parity, no dividends) The present value of the forward price, F, on an asset without dividends (until the expiration of the forward contract) equals the spot price:*

$$e^{-my}F = S \text{ or } F = e^{my}S, \tag{15.4}$$

where S is the current spot price of the underlying asset and y is m-period spot interest rate.

(If you prefer effective interest rates, then (15.4) reads $F = (1+Y)^m S$.)

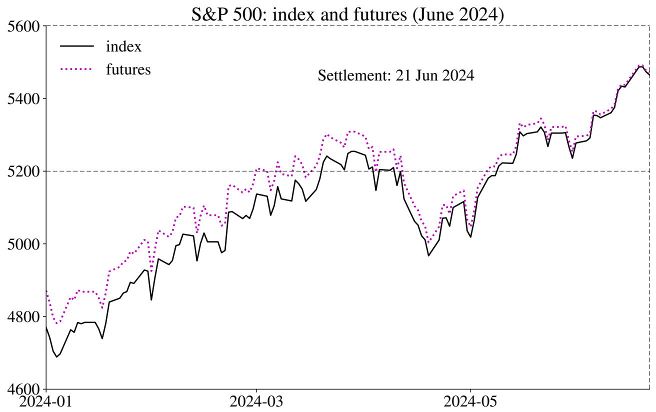


Figure 15.3: S&P 500 index level and futures

With a positive interest rate, the forward price is higher that today's underlying price. The forward-spot parity means that a *covered strategy* should have the same return as the risk-free rate: buy the underlying asset now (S) and issue a forward contract, and get the forward price (F) at expiration. This is a risk-free strategy with a gross return of $F/S = e^{my}$, if the parity holds. See Hull (2022) 5 and 8–9 and McDonald (2014) 6–8 for more details.

Example 15.4 *(Forward-spot parity) With* $y = 0.05$*,* $m = 3/4$ *and* $S = 100$ *we have the forward price* $e^{3/4 \times 0.05} 100 \approx 103.82$*.*

Proof (of Proposition 15.3) Issue a forward contract, borrow $e^{-my}F$, and buy one unit of the asset. This will for sure give zero at expiration (hand over the asset, collect F and repay the loan). The cost today must therefore be zero, or there is an arbitrage opportunity. $\square$

Example 15.5 *(Arbitrage when Proposition 15.3 does not hold) Assume the same parameters as in Example 15.4, except that* $F = 105$*. Today: issue a forward contract, borrow* $e^{-3/4 \times 0.05} 105 \approx 101.14$ *and buy the underlying asset for* 100*. You have made a risk-free profit of* 1.14*. (At expiration, hand over the underlying and collect the forward price—which is just enough to repay the loan.)*

Proposition 15.6 *(Forward-spot parity, continuous dividends) When the dividend* *is paid continuously as the rate δ (of the price of the underlying asset), then*

$$e^{-my}F = Se^{-m\delta}, \text{ so} \tag{15.5}$$

$$F = Se^{m(y-\delta)} \tag{15.6}$$

Notice that the dividends decrease the forward price. The intuition is that the forward contract does not give the right to these dividends so its present value is the underlying asset value stripped of the present value of the dividends.

Proof (*of Proposition 15.6) Portfolio A: enter a forward contract, with a present value of $e^{-my}F$. Portfolio B: buy $e^{-m\delta}$ units of the asset at the price $e^{-m\delta}S$, and then collect dividends and reinvest them in the asset. Both portfolios give one asset at expiration, so they must have the same costs today. □

Example 15.7 *(Forward-spot parity) Based on Example 15.4, but with a continuous dividend rate of* $\delta = 0.01$*, we get* $F = e^{0.75\times(0.05-0.01)}100 \approx 103.04$*, which is lower than before.*

Notice that the forward prices converges to the underlying price at expiration of the futures. Before that it can deviate because of delayed payment (+) and no participation in dividend payments (−).

Empirical Example 15.8 *Figure 15.3 show the underlying price and the futures price on S&P 500 developed over six months. (A futures price is typically very close to a forward price, as discussed below.)*

Proposition 15.9 *(*Forward-spot parity, discrete dividends) Suppose the underlying asset pays the dividend* d_i *at* m_i $(i = 1, ..., n)$ *periods into the future (but before the expiration date of the forward contract). To do the proper discounting, let* $y(m_i)$ *be today's* m_i*-period interest rate. If the dividends are known already today, then the forward price satisfies*

$$e^{-my(m)}F = S - \textstyle\sum_{i=1}^{n} e^{-m_i y(m_i)} d_i. \tag{15.7}$$

Proof (of Proposition 15.9*) Portfolio A: enter a forward contract, with a present value of $e^{-my}F$. Portfolio B: buy one unit of the asset at the price S and sell the rights to the known dividends at the present value of the dividends. Both portfolios give one asset at expiration, so they must have the same costs today. □

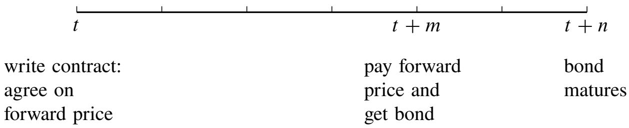


Figure 15.4: Timing convention of forward contract on a bond

15.3.3 Application: The Forward Price of a Bond

Consider a forward contract (expiring in m) on a zero coupon bond that matures in n (assuming $n > m$). See Figure 15.4 for an illustration of the timeline.

By the forward spot parity (15.4), today's forward price is

$$F = e^{my(m)}B(n) = B(n)/B(m), \tag{15.8}$$

where $B(n)$ is the price of an n-period bond today and $B(m) = e^{-my(m)}$ is the price of an m-period bond (with a face value of 1). Thus, the forward price is just the price of a long-maturity bond relative to that of a short-maturity bond.

Example 15.10 *(Forward price of a bond) Let* $(m, n, B(m), B(n))$ *be* $(5, 7, 0.779, 0.657)$. *Then,* $F = 0.657/0.779 \approx 0.843$.

15.3.4 Application: The Forward Price of Foreign Currency

Let S be the price (measured in domestic currency) of one unit of foreign currency. Investing in foreign currency effectively means investing in a foreign interest bearing instrument which earns the continuous interest rate ("dividend") $\delta = y^*$, so (15.6) gives

$$F = Se^{m(y-y^*)}. \tag{15.9}$$

This is called the *covered interest rate parity* (CIP).

Example 15.11 *(CIP) With* $S = 1.20, m = 1, y = 0.0665$ *and* $y^* = 0.05$ *we have*

$$F = 1.20e^{0.0165} = 1.22.$$

15.3.5 The Return on Holding a Forward Contract until Expiration

Suppose you enter a forward contract today and hold it until it expires in period m. You do not pay anything up front in, but you have pledged to pay F in period m, which has a present value of $e^{-my}F$. You could put this amount on a bank (money market) account and consider it your investment (often referred to as a *prepaid forward contract*). The payoff is clearly the value of the underlying asset at expiration: S_m. The gross return is therefore

$$1 + R = \frac{S_m}{F} e^{my}. \tag{15.10}$$

For an asset with continuous (or no) dividends, the forward-spot parity (15.6) then shows that the gross return is just $S_m e^{m\delta}/S$, which is the same as holding the underlying asset (and collecting the dividends, if any).

15.3.6 The Value of an Old Forward Contract

Consider a forward contract that expires in $t+m$, although the contract was written at some earlier point in time ($\tau < t$) and specified a forward price of F_τ (time subscripts are needed for the analysis here). The value of this contract in t is

$$W_t = e^{-my}(F_t - F_\tau), \tag{15.11}$$

where F_t is today's forward price on the same underlying asset (and same expiration date). This value, W_t, is what someone would pay in order to buy the old forward contract. The intuition is that an owner of an old (τ) forward contract can short sell a new forward contract (t) and thereby cancel all risk—and stand to win $F_t - F_\tau$ at expiration. The present value of this is (15.11). Clearly, for a new contract ($t = \tau$), the value is zero.

Proof (15.11) An investor sells (issues) a forward contract in t. At expiration, this will give $F_t - S_{t+m}$. If she buys an old forward contract (paying W_t today), the payoff of that is $S_{t+m} - F_\tau$ at expiration. Hence, the total portfolio has the payoff $F_t - F_\tau$, which is risk-free so it must earn the risk-free rate: $(F_t - F_\tau)/W_t = e^{my}$. Rearrange to get (15.11). $\square$

15.4 Forwards versus Futures

A forward contract is typically a private agreement between two investors—and can therefore be tailor-made. A futures contract is similar to a forward contract (write contract, get something later at a pre-determined price), but is typically traded on an exchange—and is therefore standardized (amount, maturity, settlement process). As for the settlement, it is either in cash (paying the value of the underlying asset) or physical (delivering the underlying asset). The latter is not used for synthetic/complex assets like equity indices since it would involve considerable transaction costs.

Another important difference is that a forward contract is settled at expiration, whereas a futures contract is settled daily (*marking-to-market*). This essentially means that gains and losses (due to price changes, see (15.11)) are transferred between issuer and owner daily—but kept at an interest bearing account at the exchange. The counterparties have to post *initial margins*—and the marking-to-market then adds to/subtracts from the margin accounts. If the amount decreases below a certain level (maintenance margin), then a *margin call* is issued to the investor—informing him/her to add cash to the margin account. See Example 15.12.

The margin requirements for an investor is governed by his/her overall portfolio (for instance, it is smaller if the portfolio includes negatively correlated positions) and is set by statistical measurements of the portfolio risk (see the *CME SPAN* system).

Example 15.12 *(Margin account) Consider a margin account of a buyer of a futures contract, where the maintenance margin = 0.75×initial margin with the initial margin some 3-12% of the notional value of the contract. For simplicity, assuming a zero interest rate it could be*

Day	*Futures price*	*Daily gain*	*Posting of margin*	*Margin account*
0	100		4	4
1	99	-1		3
2	97	-2	2	3
3	99	2		5

On day 2, the investor received a margin call to add cash to the account—to make sure that the maintenance margin (here 3) is kept. Notice that the overall profit on day 3 is the difference of what has been put into the margin account $(4+2)$ *and the final balance (5), that is,* -1*. This is also the* cumulative *daily gain* $(-1-2+2=-1)$*.*

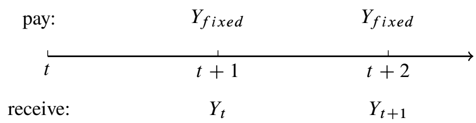


Figure 15.5: 2-year fixed-for-floating interest rate swap

With marking to market this is all that happens: no payment of the futures price and no delivery of the underlying asset. However, it is equivalent to what happen without marking to market, since at expiration, the gain is $99 - 100 = -1$ *(futures = underlying at expiration).*

If interest rates change randomly over time (and they do), the rate at which the money on the margin account is invested at will be different from the rate when the futures was issued. This risk of this happening is reflected in the futures price.

Instead and more theoretically, if the interest rate path were non-stochastic (and there was no counterparty risk), then the forward and futures prices would be the same. See the proposition below. (The proof is in the Appendix) In practice, the difference between forward and futures prices is typically small.

Proposition 15.13 *(Forward vs. futures prices, non-stochastic interest rates) The forward and futures prices would be the same (a) if there were no counterparty risk; (b) and if the interest rate only changed in a non-stochastic way.*

15.5 Swap Contracts

Swap contracts involve the exchange of two payment streams over a predefined period.

For instance, in a *fixed-for-floating interest rate swap* as illustrated in Figure 15.5, counterparty A pays a fixed interest rate at the end of each period (in the figure simplified to be each year) to counterparty B, while B the pays a floating rate (defined by referencing to an existing asset or index), In this case, this is very similar to a portfolio of forward contracts: the fixed rate is the forward price and the underlying assets are the values of the floating rates (for each respective quarter). Therefore, the pricing of the fixed leg of the swap contract could be derived from forward contracts (and vice versa).

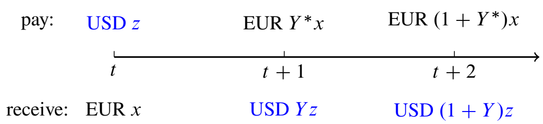


Figure 15.6: 2-year fixed-for-fixed currency swap

An *FX swap* is typically just a spot buy of currency and a contracted agreement to sell it back (for a fixed price F) in a predetermined future period. This is basically a spot transaction combined with a forward contract. It can also be though of as an exchange of loans: one counterparty lends one currency to another counterparty, who in turn lends another currency.

As an another example, Figure 15.6 illustrates a *fixed-for-fixed currency swap*. It is essentially two loans, but in different currencies: counterparty A borrows EUR (and pays interest on that), and lends USD (and receives interest on that). Counterparty B does the opposite. It is called fixed-for-fixed since both interest rates are fixed.

15.6 Appendix: Some Extra Proofs*

Proof (of Proposition 15.13*) To simplify the notation, let $t = 0$ and $m = 2$. Also, let r_s be the continuously compounded rate at which you accumulate interest on the margin account between days s and $s+1$ (that is, $r = y/365$) and f_s be the futures price on day s. *Strategy A*: have e^{-r_1} long futures contracts on (the end of) day 0, pre-commit to increase it to 1 on day 1 and keep all settlements on the margin account. This gives

Day (s)	Settlement	Futures Position (EOD)	Margin Account (EOD)
0		e^{-r_1}	0
1	$e^{-r_1}(f_1 - f_0) = A$	1	A
2	$f_2 - f_1 = B$	0	$e^{r_1}A + B$,

where EOD means end of day. The end-value of strategy A is therefore $f_2 - f_0$, which equals $S_2 - f_0$ since the value at expiration is the value of the underlying asset. *Strategy B*: be long one forward contracts, which gives a payoff on day 2 of $S_2 - F_0$. Both strategies take on exactly the same risk, so the prices must be the same: $f_0 = F_0$. (The proof relies on knowing r_1 already on day 0.) □

Chapter 16

Interest Rate Calculations

This chapter first presents details on different types of bonds (zero coupon bonds, coupon bonds), and later discusses the yield curve, forward interest rates and the yield to maturity. It can be seen as a general description of the instruments on the money and bond markets.

16.1 Zero Coupon Bonds

16.1.1 Zero Coupon Bond Basics

Remark 16.1 *(On the notation) The notation is kept short. Time subscripts and indicators of time to maturity are typically suppressed. That is, we use B and Y instead of $B_t(m)$ and $Y_t(m)$, unless the indicator for the trading date (t) and/or time to maturity (m) are important in the specific context. This is mostly unproblematic, but sometimes we need expressions like $my_t(n-m)$, which says m times the $(n-m)$-period interest rate in t, which we write as $m \times y_t(n-m)$ to make it less confusing.*

Consider a zero coupon bond (also called a “discount” or “bullet” bond) that

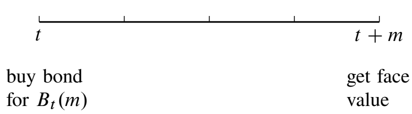


Figure 16.1: Timing convention of zero coupon bond

costs $B_t(m)$ in t and pays the face value in $t+m$ (we will often use the short hand notation B). The time to maturity (also called "tenor" or "term"), m, is measured in years (for instance, $m = 1/2$ means half a year). See Figure 16.1 for an illustration of the timeline.

The gross return (payoff divided by price) from investing in this bond is $1/B$, as the face value is here normalized to unity. The relation between the *bond price* B and the *effective (spot) interest rate* $Y(m)$ is

$$1/B = (1+Y)^m, \tag{16.1}$$

$$B = (1+Y)^{-m}, \tag{16.2}$$

$$Y = B^{-1/m} - 1. \tag{16.3}$$

Equation (16.1) states that the interest rate Y is an *annualized rate of return* derived from investing B and receiving the face value (here normalized to 1) m years later. Similarly, (16.2) says that the bond price is the present value of the face value (again, 1). Equation (16.3) solves for the interest rate in terms of the bond price.

Example 16.2 *(Effective rates) Consider a six-month bill so* $m = 0.5$*. Suppose* $B = 0.95$*. From (16.1) we then have that*

$$1/0.95 = (1+Y)^{0.5}, \text{ so } Y \approx 0.108.$$

Remark 16.3 *(A face value of 100) In case the face value is* X *(say, 100) instead of* 1*, then the bond price will be* X *times higher than with a face value of 1. The left hand side of (16.1) will be* X/B *and give the same interest rate. In practice, bond quotes are typically expressed in percentages (like 97, often leaving out the % sign) of the face value, whereas the discussion here effectively uses the fraction of the face value (like 0.97).*

The relation between the interest rate and the price depends on the time to maturity (m): prices on long-maturity bonds are more sensitive to interest rate changes than prices on short-maturity bonds. The relationship is also slightly convex. These features will be important when we discuss hedging bond portfolios. See Figure 16.2 for an illustration.

We also have the following relation between the bond price and the *continuously compounded interest rate* (y)

$$1/B = e^{my}, B = e^{-my}, \text{ and } y = -(\ln B)/m. \tag{16.4}$$

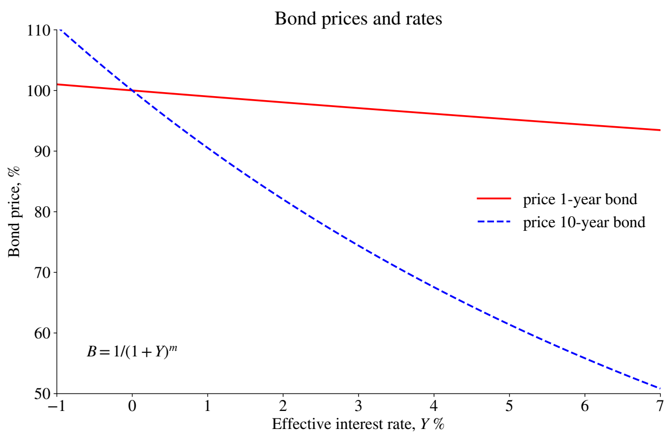


Figure 16.2: Interest rate vs. bond price

Example 16.4 *(Continuously compounded rate) Using the numbers as in Example 16.2, (16.4) gives*

$$1/0.95 = e^{0.5y}, \text{ so } y \approx 0.103.$$

Some fixed income instruments (in particular, interbank loans) are quoted in terms of a *simple interest rate* ($\tilde{Y}$), which is related to the bond price according to

$$1/B = 1 + m\tilde{Y},\ B = \frac{1}{1 + m\tilde{Y}}, \text{ and } \tilde{Y} = \frac{1/B - 1}{m}. \tag{16.5}$$

Example 16.5 *(Simple rates) Consider a six-month bill so $m = 0.5$. Suppose $B = 0.95$. From (16.5) we then have that*

$$0.95 = \frac{1}{1 + 0.5\tilde{Y}}, \text{ so } \tilde{Y} \approx 0.105.$$

Remark 16.6 *(The transformation from one type of interest rate to another*) We have*

$$\begin{aligned}
Y &= e^{y} - 1 \text{ and } Y = (1 + m\tilde{Y})^{1/m} - 1 \\
y &= \ln(1 + Y) \text{ and } y = \ln\left(1 + m\tilde{Y}\right)/m, \\
\tilde{Y} &= [(1 + Y)^{m} - 1]/m \text{ and } \tilde{Y} = (e^{my} - 1)/m.
\end{aligned}$$

The different interest rates (effective, continuously compounded, and simple) are typically quite similar, except at very high rates. See Figure 16.3 for an illustration.

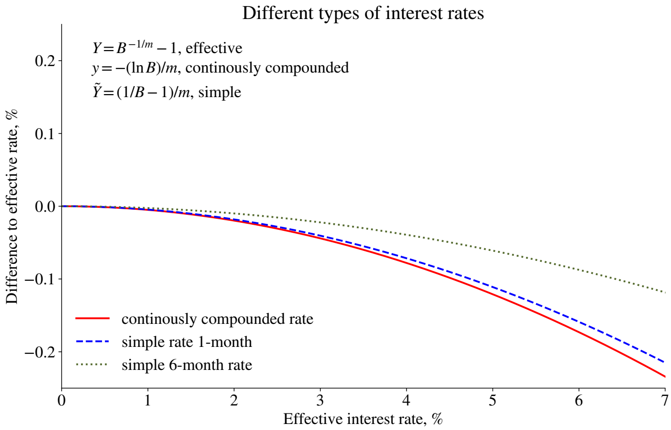


Figure 16.3: Different types of interest rates

Example 16.7 *(Different interest rates) For* $m = 1/2$, $Y = 0.108$, $y = 0.103$ *and* $\tilde{Y} = 0.106$

$$1.053 \approx (1 + 0.108)^{0.5} \approx e^{0.5 \times 0.103} \approx 1 + 0.5 \times 0.105.$$

The *yield curve* is the interest rate plotted (or tabulated) against the time to maturity, see Figure 16.4 for a 3-D plot of how the U.S. yield curves have shifted over time.

Empirical Example 16.8 *(U.S. yield curve over time) Figure 16.2 shows how the U.S. yield curve has changed over time. The level is typically lower after the Volcker deflation in the early 1980s, but there are still some variations (especially after 2022). Also, the slope changes in non-trivial ways: the yield curve are mostly up-ward sloping, but there are many instances of almost flat curves, and sometimes even down-ward sloping ones.*

16.1.2 The Return from Holding a Zero Coupon Bond

The log *return from holding a zero coupon bond* from t to $t + s$ is clearly the relative change in the bond price

$$r_{t+s} = \ln \frac{B_{t+s}(m-s)}{B_t(m)}, \tag{16.6}$$

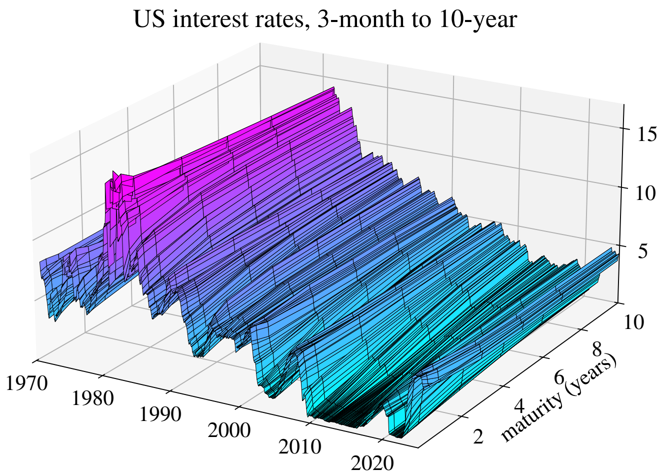


Figure 16.4: US yield curves

where the subscripts indicate the trading date and the terms in parentheses indicate time to maturity. Notice that the bond's time to maturity decreases with time, here from m to $m-s$. Also, notice that r_{t+s} defines a traditional log return over s periods, in contrast to interest rates which are expressed on a "per year" basis.

Example 16.9 *(Bond return) If the bond price decreases from* 0.95 *to* 0.86*, then* (16.6) *gives the log return*

$$\ln \frac{0.86}{0.95} = -0.1.$$

Substituting for the bond prices in (16.6) gives

$$\begin{aligned} r_{t+s} &= -m \times [y_{t+s}(m-s) - y_t(m)] + s \times y_{t+s}(m-s), \\ &= -m[y_{t+s} - y_t] + s y_{t+s}, \end{aligned} \tag{16.7}$$

where the second line uses simplified notation (suppressing indication of the maturity). We use this expression to study some special cases to highlight key properties of bond returns.

The *first special case* is when we hold the bond until maturity ($m = s$), in which case $r_{t+m} = m y_t$. This mean earning today's interest rate for m years.

The *second special case* is a very short holding period (s is very small). The second term in (16.7) is then virtually zero, so we can write

$$r_{t+s} \approx -m[y_{t+s} - y_t] \text{ when } s \approx 0. \tag{16.8}$$

This value is clearly negative if the interest rate change is positive—and even more so if the time to maturity (m) is long. (Clearly, nothing forces the investor to sell the bond at $t + s$, but this represents the capital gain/loss when the position is valued according to the market price.) See Figure 16.5 for an illustration (disregard the coupon bond for now). Also, see Elton, Gruber, Brown, and Goetzmann (2014) 21–22 and Hull (2022) 4 for more detailed discussions.

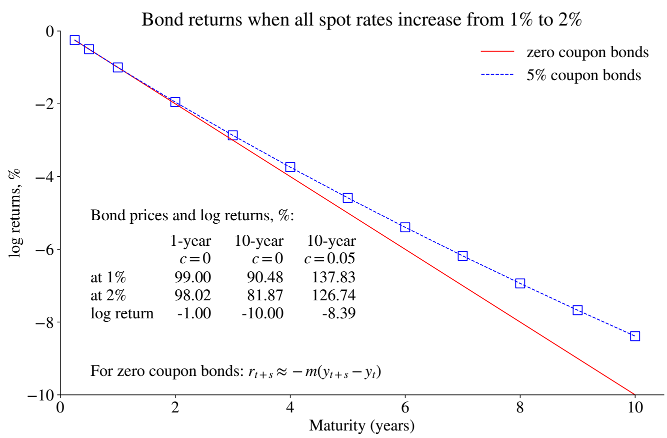


Figure 16.5: Returns after interest rate changes

Example 16.10 *(Bond returns vs interest rate changes) Figure 16.5 includes a table with some calculations. For instance, the 10-year zero coupon bond is worth* $e^{-10\times0.01} \approx 90.48\%$ *at the 1% interest rate, but* $e^{-10\times0.02} \approx 81.87\%$ *at the 2% rate. This gives a log return of* -10%*, as in (16.8).*

The *third special case* is an unchanged flat yield curve. In this case, all interest rates in (16.7) are the same (and here denoted y), so we get

$$r_{t+s} = sy, \tag{16.9}$$

which mean earning (today's) interest rate for s years. The reason is that the bond starts out as a m-maturity bond, but becomes an $(m - s)$-maturity bond—and the latter has a higher price (if $y > 0$), as illustrated in Figure 16.6.

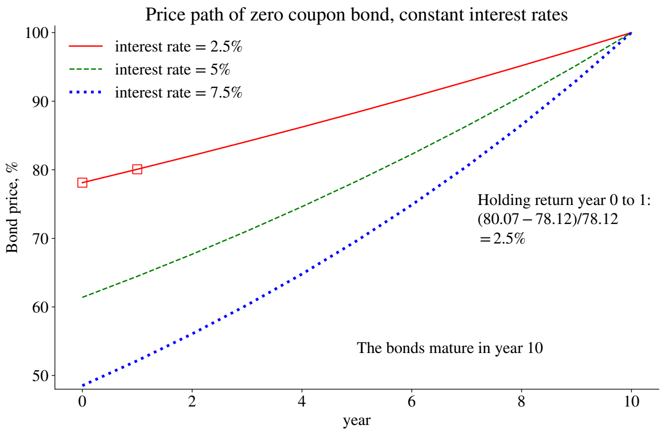


Figure 16.6: The price of a zero coupon bond maturing in year 10, constant interest rates

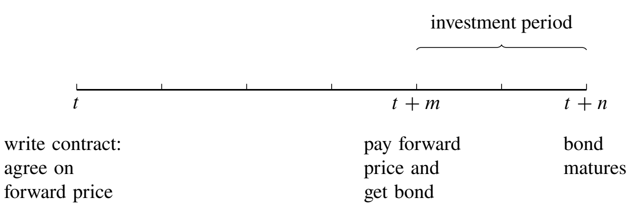


Figure 16.7: Timing convention of forward contract

16.2 Forward Rates

A forward contract on a bond allows an investor to lock in an interest rate for a future investment period. Consider entering a forward contract in $t = 0$: it specifies *(a)* the amount the investor has to pay in period m (the forward price, F), and *(b)* which bond that will be delivered, in particular, one that matures in n, where $n > m$. See Figure 16.7 for an illustration of the timeline.

The forward-spot parity says that the present value of the forward price on a zero-coupon bond must equal today's spot price

$$[1 + Y(m)]^{-m} F = B(n). \tag{16.10}$$

Entering a forward contract represents a commitment to an investment in period

m. If we keep the bond until maturity in n, then this investment spans $n - m$ years. The gross return, which is known already today, is $1/F$. A per-year effective rate of return, referred to as the *forward rate* (Γ), is defined analogously to an interest rate

$$1/F = (1 + \Gamma)^{n-m}\,. \tag{16.11}$$

By using the relation between bond prices and yields (16.1), the forward rate can be written

$$\Gamma = \frac{[1 + Y(n)]^{n/(n-m)}}{[1 + Y(m)]^{m/(n-m)}} - 1. \tag{16.12}$$

(All values in this expression are determined today, in $t = 0$.) This expression demonstrates that the forward rate depends on both interest rates and, consequently, the general shape of the yield curve. Actually, the forward rate can be interpreted as the "marginal cost" of extending the loan's maturity. See Figure 16.8 for an illustration.

Example 16.11 *(Forward rate) Let $m = 0.5$ (six months) and $n = 0.75$ (nine months), and suppose that $Y(0.5) = 0.04$ and $Y(0.75) = 0.05$. Then (16.12) gives*

$$\Gamma = \frac{(1 + 0.05)^{0.75/0.25}}{(1 + 0.04)^{0.5/0.25}} - 1 \approx 0.07.$$

See Figure 16.8 for an illustration.

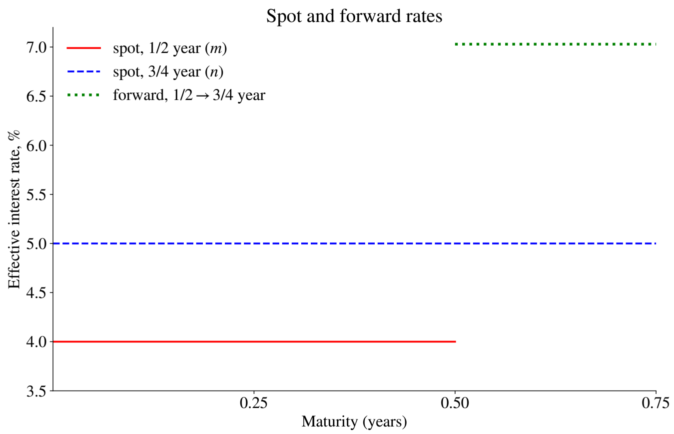


Figure 16.8: Spot and forward rates

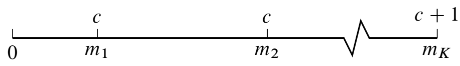


Figure 16.9: Timing convention of coupon bond

Remark 16.12 *(Forward Rate Agreement (FRA)) An* FRA *is an over-the-counter contract that secures an interest rate during a future period in exchange for a floating rate. The FRA does not involve any lending or borrowing; rather, it provides compensation for deviations between the future floating interest rate and the agreed forward rate. An FRA is similar to a one-period, and typically short-term, interest rate swap.*

Remark 16.13 *(Alternative way of deriving the forward rate*) Rearrange (16.12) as*

$$[1 + Y(m)]^m \,(1 + \Gamma)^{n-m} = [1 + Y(n)]^n \,.$$

This says that compounding $1 + Y(m)$ *over* m *years and then* $1 + \Gamma$ *for* $n - m$ *years should give the same amount as compounding* $1 + Y(n)$ *over* n *years.*

16.3 Coupon Bonds

Remark 16.14 *(On the notation) These notes often use* P *instead of* $P_t(c, m_1, ..., m_K)$ *to denote the price of a coupon bond. The indicator of the trading date (*t*), coupon rate (*c*) and time until coupon payments* $m_1, ..., m_K$ *are only needed when important in the specific context.*

16.3.1 Coupon Bond Basics

Consider a bond which pays coupons, c, on K occasions ($m_1, m_2, .., m_K$ years from now), and the face (or par) value, normalized to 1, at maturity (m_K). As before, m_k is measured in years. See Figure 16.9 for an illustration.

A coupon bond can be considered a portfolio of zero coupon bonds: c of them maturing in m_1, another c in m_2,..., and $c + 1$ in m_K. The price of the coupon bond (P) must therefore equal the price of the portfolio

$$P = \textstyle\sum_{k=1}^{K} B(m_k)c + B(m_K) \tag{16.13}$$

where $B(m_k)$ is the price of a zero coupon bond maturing m_k years later. This is

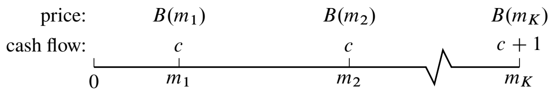


Figure 16.10: Using zero-coupon bonds to value a coupon bond

illustrated in Figure 16.10. Using the relation between (zero coupon) bond prices and spot interest rates in (16.1), equation (16.13) can also be written

$$P = \sum_{k=1}^{K} \frac{c}{[1 + Y(m_k)]^{m_k}} + \frac{1}{[1 + Y(m_K)]^{m_K}}. \tag{16.14}$$

This shows that coupon bond price is just the present value of the cash flow from coupons and the face value, but where the discounting is done by the different spot interest rates. In these calculations, P represents the full (invoice) price of the bond, which can differ from the quoted price (also called “clean price”) by an accrued interest rate term. See the appendix on market conventions for a discussion. Also, see McDonald (2014) 9 and Fabozzi (2004) for more detailed discussions.

The same valuation principle can be applied to more complicated cash flow processes, such as a portfolio of bonds. Suppose the bond portfolio pays the cash flow cf_k in m_k years from now, as illustrated in Figure 16.11. This cash flow could include both coupon payments and face values. The pricing expressions (16.13)–(16.14) can then be generalised to

$$P = \sum_{k=1}^{K} B(m_k) cf_k \tag{16.15}$$

$$= \sum_{k=1}^{K} \frac{cf_k}{[1 + Y(m_k)]^{m_k}}. \tag{16.16}$$

Clearly, setting $cf_k = c$ for $k \leq K - 1$ and $cf_K = c + 1$ gives (16.13) and (16.14).

Remark 16.15 *((16.16) with continuously compounded rates*) The bond price can also be written* $P = \sum_{k=1}^{K} cf_k / \exp[m_k y(m_k)]$.

Remark 16.16 *(Floating Rate Notes) FRNs are bonds with floating coupon payments, typically indexed to some reference interest rate (for instance, T-bills). They are particularly common on the corporate bond market. Since the coupons are not known in advance, the approach in this section is not applicable. In a way, they are more similar to a combination of a coupon bond plus an interest rate swap (discussed below).*

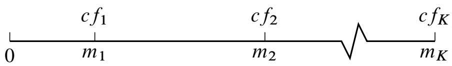


Figure 16.11: Timing convention of bond portfolio

Example 16.17 *(Coupon bond prices) For bonds with 1 and 2 years until maturity, (16.13) can be written in matrix form as*

$$\begin{bmatrix} P(1) \\ P(2) \end{bmatrix} = \begin{bmatrix} c(1)+1 & 0 \\ c(2) & c(2)+1 \end{bmatrix} \begin{bmatrix} B(1) \\ B(2) \end{bmatrix},$$

where we use $P(m)$ and $c(m)$ to indicate the price and coupon rate for the m-year coupon bond. For instance, with $(B(1), c(1)) = (0.95, 0)$ and $(B(2), c(2)) = (0.90, 0.06)$ we have that

$$\begin{bmatrix} P(1) \\ P(2) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0.06 & 1.06 \end{bmatrix} \begin{bmatrix} 0.95 \\ 0.90 \end{bmatrix}, \text{ which gives } \begin{bmatrix} P(1) \\ P(2) \end{bmatrix} \approx \begin{bmatrix} 0.95 \\ 1.01 \end{bmatrix}.$$

Example 16.18 *(Coupon bond price at par) Suppose $B(1) = 1/1.06$ and $B(2) = 1/1.091^2$. The price of a bond with a 9% annual coupon with two years to maturity is then*

$$\frac{0.09}{1.06} + \frac{0.09}{1.091^2} + \frac{1}{1.091^2} \approx 1.$$

This bond is (approximately) sold "at par", that is, the bond price equals the face (or par) value (which is 1 in this case).

Example 16.19 *("Bootstrapping") Reconsider Example 16.17, but suppose we instead have information about prices of the coupon bonds—and that we want to know the implied prices of the zero coupon bonds. This can be done by solving the equations for $B(1)$ and $B(2)$. That means we solve*

$$\begin{bmatrix} 0.95 \\ 1.01 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0.06 & 1.06 \end{bmatrix} \begin{bmatrix} B(1) \\ B(2) \end{bmatrix} \text{ to get } \begin{bmatrix} B(1) \\ B(2) \end{bmatrix} \approx \begin{bmatrix} 0.95 \\ 0.90 \end{bmatrix}.$$

(More details on bootstrapping are given in an appendix.)

Remark 16.20 *(STRIPS, Separate Trading of Registered Interest and Principal of Securities*) A coupon bond can be split up into its embedded zero coupon bonds—and traded separately (as zero coupon bonds).*

16.3.2 Yield to Maturity

The effective *yield to maturity* (abbreviated ytm, and also referred to as the redemption yield), θ, of a bond portfolio is the internal rate of return that satisfies the following relationship

$$P = \sum_{k=1}^{K} \frac{cf_k}{(1+\theta)^{m_k}}. \tag{16.17}$$

where the portfolio has the cash flow cf_k in $m_1, m_2, ..., m_K$ years. This equation can be solved (numerically) for θ. Bonds are commonly quoted based on their yield to maturity, rather than their price. For a *par bond* where $P = 1$, the yield to maturity is equal to the coupon rate. For a zero coupon bond, the yield to maturity equals the spot interest rate.

Example 16.21 *(Yield to maturity) A 4% (annual coupon) bond with 2 years to maturity. Suppose the price is 1.019. The ytm is 3% since it solves*

$$1.019 \approx \frac{0.04}{1+0.03} + \frac{1.04}{(1+0.03)^2}.$$

Example 16.22 *(Yield to maturity of a par bond) A 9% (annual coupon) par bond with 2 years to maturity has a ytm of 9% since*

$$\frac{0.09}{1+0.09} + \frac{1.09}{(1+0.09)^2} = 1.$$

Example 16.23 *(Yield to maturity of a portfolio) A 1-year discount bond with a ytm (effective interest rate) of 7% has the price* $1/1.07$ *and a 3-year discount bond with a ytm of 10% has the price* $1/1.1^3$*. A portfolio with one of each bond has the ytm that solves*

$$\frac{1}{1.07} + \frac{1}{1.1^3} = \frac{1}{1+\theta} + \frac{1}{(1+\theta)^3}, \text{ so } \theta \approx 0.091.$$

This is clearly not the average ytm of the two bonds.

Remark 16.24 *(Approximate ytm*)* $\theta \approx 2[(c + (1-P)/K]/(1+P)$ *is sometimes used as an approximation. For the bond in Example 16.21 we would get* $\theta \approx 3.02\%$*. However, this approximation becomes less precise when the bond price is far from par.*

16.3.3 The Return from Holding a Coupon Bond

To calculate the return from holding a coupon bond *until maturity*, it is necessary to specify *how the coupons are reinvested.* If the coupons are reinvested through forward contracts (agreed upon today), the return is the same as that of a zero-coupon bond. This result is intuitive because the investor purchases the bond now and receives no payments until maturity, similar to a zero-coupon bond. This is summarised in the following proposition (the proof is at the end of the section).

Proposition 16.25 *(Return from holding a coupon bond until maturity) If the coupons are reinvested by forward contracts, then the (annualized) return on holding the bond until maturity is the current spot rate on a zero coupon bond with the same maturity.*

Note that this result holds regardless of the coupon rate, so it could be said that coupons do not matter for returns. (With other assumptions about how the coupons are reinvested, it might be different.)

Example 16.26 *(Holding a coupon bond until maturity) Suppose that the spot (zero coupon) interest rates are 4% for one year to maturity and 5% for 2 years to maturity (the zero coupon bond prices are* $B(1) = 0.962$ *and* $B(2) = 0.907$*). A 3% coupon bond with 2 years to maturity must have the current price*

$$\frac{0.03}{1.04} + \frac{0.03 + 1}{1.05^2} \approx 0.963.$$

However, the value of the bond portfolio at maturity, if the coupon is reinvested by a forward contract, is

$$0.03 \times \frac{0.962}{0.907} + 0.03 + 1 \approx 1.062,$$

so the gross return over two years is approximately $1.062/0.963 \approx 1.102$*. Compare that to* $(1 + 0.05)$^2*, which is approximately the same (some differences due to rounding).*

A more hypothetical (text book) case is when we (in the future) can reinvest at a rate equal to today's yield to maturity. The next proposition summarizes this (the proof is in an appendix).

Proposition 16.27 *(Return from holding a coupon bond until maturity, a special case) If all coupons are reinvested in assets that generate returns equal to today's yield to maturity of the bond (*θ*), then the (annualized) rate of return will be* θ*.*

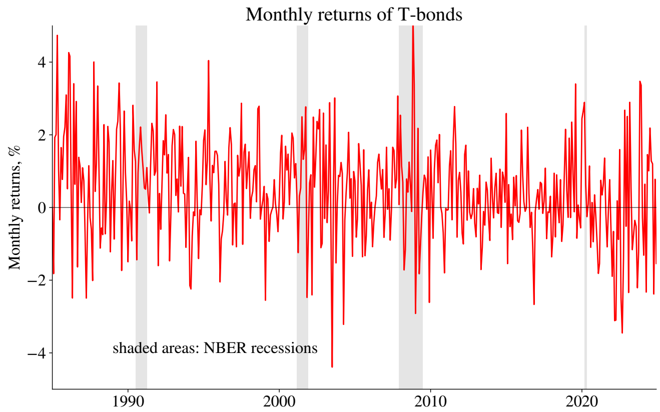


Figure 16.12: Returns on an index of U.S. Treasury bonds

In contrast, the gross return from holding a coupon bond until a period *before maturity* depends on both the bond price development and the value of the (reinvested) coupon payments received until the bond is sold. When there are changes in the interest rate level and we sell the bond before maturity, then the bond price changes often dominate: lower interest rates mean capital gains and vice versa (just like for zero coupon bonds). For long-maturity bonds, the effects can be considerable. See Figure 16.5 for an illustration.

In the special case where the coupons are locked in by forwards, then the bond is effectively transformed into a zero-coupon bond, so the return is same as on an m_K-year zero coupon bond bought in $t = 0$ and sold in s (with $s \leq m_K$), similar to Proposition 16.25. (Details are in an appendix.)

Empirical Example 16.28 *Figure 16.12 shows monthly returns on a basked of U.S. T-bonds. These returns are less volatile than equity returns, but still show non-trivial movements.*

Proof (of Proposition 16.25) Consider a 2-year coupon bond. From (16.13), the price of the bond is $P_t = B_t(1)c + B_t(2)(c + 1)$. From (16.11), we know that the forward contract for the first coupon has a gross return (from $t = 1$ to $t = 2$) of $B_t(1)/B_t(2)$. The value of the reinvested coupon and the face value at maturity is then $cB_t(1)/B_t(2) + c + 1$. Dividing by P_t (the investment) gives $1/B_t(2)$, so the return on buying and holding (and reinvesting the coupons) this

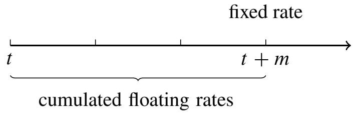


Figure 16.13: Timing convention of an OIS swap with one payment

coupon bond is the same as the 2-year spot interest rate. (The extension to more years is straightforward.) □

16.4 Other Credit Instruments

16.4.1 Overnight Indexed Swap (OIS)

Overnight indexed swaps (OIS) have have supplanted the earlier LIBOR market for lending and borrowing between financial institutions, as well as for valuing derivatives. In its simplest form, such a contract (agreed upon at $t = 0$) specifies a fixed payment in m (the OIS rate) against receiving an accumulated value, which is approximately an average of the realised overnight ("floating") interest rates between $t = 0$ and m. See Figure 16.13 for an example. (Also, see the appendix on bond market conventions for details on the accumulation of the floating interest rates.) These contracts typically have a notional face value that scales the payment.

Remark 16.29 *With a notional value of 1000, an OIS rate of 4% and an accumulated floating rate of 3%, the payment at the end of a 3-month contract is* $1000 \times (0.04 - 0.03) \times 0.25$.

For longer-maturity contracts (m exceeding a year), the contract often involves periodical payments (typically every three months), where the fixed OIS rate is compared with the cumulative overnight interest rates since the last payment. See Figure 16.14 for an illustration.

Empirical Example 16.30 *Figure 16.15 shows the Euro OIS rates (1m to 12m) since late 2019.*

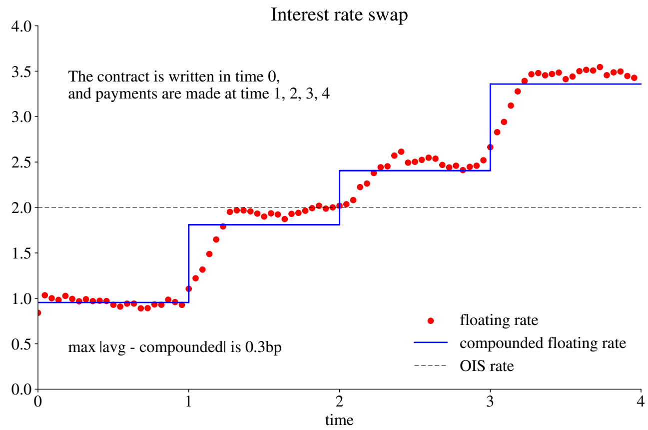


Figure 16.14: OIS with several payments

16.4.2 Repo

In a repurchase agreement (*repo*), investor A sells a security to investor B, with an agreement to repurchase it at a predetermined price at some specific future time (the next day, after a week, etc.). The *repo rate* is calculated as the relative difference between the initial and the repurchase price.

The contract essentially implies that investor A borrows cash, while investor B borrows the asset. Investor B is said to have made a reverse repo, and can sell the asset to someone else. This is a way of shortening the security, so the repo rate is low if there is a demand for shortening the security.

A *haircut* (for instance, 3%) indicates that the collateral (security) has market value that is 3% higher than the agreed price in the repo. This provides a safety margin to the lender—since the market price of the security could decrease over the life span of the repo.

Example 16.31 *(Long-short bond portfolio). First, buy bond X and use it as collateral in a repo (the repo borrowing finances the purchase of the bond). Second, enter a reverse repo where bond Y is used as collateral and sell the bond (selling provides cash for the repo lending).*

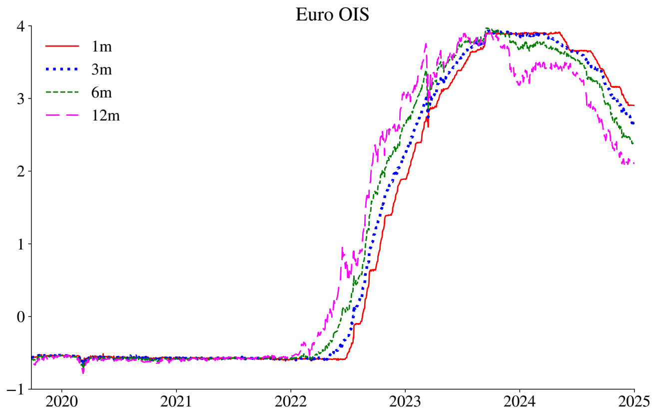


Figure 16.15: Euro OIS rates

16.4.3 Collateralized Debt Obligations

A collateralized debt obligations (CDO) is a repackaging of a portfolio of bonds (referred to as "collaterals"), in which the claims are divided into tranches with varying degrees of seniority. For instance, with junior, mezzanine and senior tranches, the higher tranches are often protected against any losses (unless they are dramatic/total). In contrast, the junior tranche is similar to equity.

CDOs are created for two main reasons. First, it is a way for the issuer (typically a bank), to "package and sell off." This is a way to shrink the balance sheet for the bank (securitisation) but still earn a fee. Second, a CDO transforms a portfolio of risky bonds to *(a)* some safe bonds and *(b)* some very risky ones. This opens up new possibilities for investors. For instance, it may allow risk averse investors (including pension funds) to invest into the safe tranches, while they would otherwise not dare (or be allowed to) invest into the original bonds.

The correlation between the defaults of the bonds within the CDO is a critical factor. The idea of tranching (in particular, to regard the senior tranche as safe) depends on the assumption that not all bonds default at the same time. Underestimating the correlation can result in significant overpricing of the senior tranches, as was frequently observed during the financial crisis 2008–9.

Another important aspect of the CDO is whether the originator (bank) holds the junior trance or not. If it does, then it has the incentives to screen the borrowers/monitor the loans, otherwise not.

16.4.4 Credit Default Swaps

A credit default swap (CDS) is a financial instrument that provides insurance against the default on a bond. Often, the CDS is a contract where one investor pays a premium (say, every quarter) in return for an insurance in case a bond defaults. Many CDS contracts are priced under the convention that a default implies only 40% of the face value can be recovered (referred to as the recovery rate), with the remainder considered lost. In such cases, it is the probability of default which is the main driver of the pricing.

If you hold a portfolio of one risky bond and a CDS on it, then you effectively own a risk-free bond. The other way around is to buy one risk-free bond and issue a CDS, which gives effectively the same as owning the risky bond. This straightforward observation is essential for understanding how the CDS premium is calculated.

Example 16.32 *(CDS payoffs) Table 16.1 gives an example of the payoffs from a CDS.*

year	Prob of survival to year t end	Prob of default in year t	Expected spread payment	Expected payment from insurance	Expected PV of net payment
1	0.98	0.02	$0.98s$	0.02×0.6	$0.98s - 0.012$
2	0.95	0.03	$0.95s$	0.03×0.6	$0.95s - 0.018$
Sum					$1.93s - 0.03$

Table 16.1: Example of the payment flows of a 2-year CDS with an assumed recovery rate of 0.4 and a risk-free interest rate of zero. The CDS spread is denoted s.

16.5 Appendix – Estimating the Yield Curve*

The (zero coupon) spot rate curve is of particular interest: it helps us price other bonds or portfolios of bonds—and it has a clear economic meaning ("the price of time").

In some cases, the spot rate curve is actually observable—for instance from swaps and STRIPS. In other cases, the instruments traded on the market include some zero coupon instruments (bills) for short maturities (up to a year), but perhaps only coupon bonds for longer maturities. This means that the spot rate curve needs to be calculated (or estimated). This section describes different methods for doing that.

16.5.1 Direct Calculation of the Yield Curve ("Bootstrapping")

We can sometimes calculate large portions of the yield curve directly from bond prices by a method called "bootstrapping."

For instance, with coupon bonds maturing in the next three periods, (16.13) can be used to write

$$\begin{bmatrix} P(1) \\ P(2) \\ P(3) \end{bmatrix} = \begin{bmatrix} c(1)+1 & 0 & 0 \\ c(2) & c(2)+1 & 0 \\ c(3) & c(3) & c(3)+1 \end{bmatrix} \begin{bmatrix} B(1) \\ B(2) \\ B(3) \end{bmatrix},$$

which is a recursive (triangular) system of equations. We can solve for the zero-coupon bond prices $B(1)$, $B(2)$ and $B(3)$ and then use (16.1) to transform to spot interest rates. See Example 16.19.

Unfortunately, the bootstrap approach is tricky to use. First, there are typically gaps between the available maturities, at least outside the U.S. treasury market. One way around that is to interpolate. Second (and quite the opposite), there may be several bonds with the same maturity but with different coupons/prices, so it is hard to calculate a unique yield curve. This could be solved by forming an average across the different bonds or by simply excluding some data. Alternatively, we could use another method than the bootstrap.

16.5.2 Estimating the Yield Curve with Regression Analysis

Recall equation (16.13) which expresses the coupon bond price in terms of a series of discount bond prices. It is reproduced here

$$P = \sum_{k=1}^{K} B(m_k)c + B(m_K). \tag{16.18}$$

If we attach a random error to the bond prices, then this looks very similar to regression equation: the coupon bond price is the dependent variable, the coupons are the regressors, and the discount function (discount bond prices) are the coefficients to estimate.

However, if there are more coupon dates than bonds, then we cannot estimate all the necessary zero coupon bond prices from data (fewer data points than coefficients).

The way around this is to decrease the number of coefficients by assuming that the discount function, $B(m)$, is a linear combination of some J predefined functions of maturity, $g_1(m)$,..., $g_J(m)$,

$$B(m) = 1 + \textstyle\sum_{j=1}^{J} a_j g_j(m), \tag{16.19}$$

where $g_j(0) = 0$ since $B(0) = 1$ (the price of a bond maturing today is one).

This can be estimated as follows *(1)* specify the $g_j(m)$ functions; *(2)* use (16.19) in (16.13) to calculate the implied bond prices; *(3)* estimate a_1,..., a_j by minimizing the squared pricing error by a least-squares approach (see, for instance, Campbell, Lo, and MacKinlay (1997) 10). One possible choice of $g_j(m)$ functions is a polynomial, $g_j(m) = m^j$. Another common choice is to make the discount bond price a spline (see McCulloch (1975)).

Example 16.33 *(Quadratic discount function) With a quadratic discount function*

$$B(m) = a_0 + a_1 m + a_2 m^2,$$

we get from (16.13)

$$\begin{aligned} P(m_K) &= \textstyle\sum_{k=1}^{K} B(m_k)c + B(m_K) \\ &= \textstyle\sum_{k=1}^{K} \left(a_0 + a_1 m_k + a_2 m_k^2\right) c + \left(a_0 + a_1 m_K + a_2 m_K^2\right). \end{aligned}$$

Collect all constants into a first regressor, then all terms that are linear in m into a second regressor and finally all terms that are quadratic in m into a third regressor

$$P(m_K) = a_0 \underbrace{(Kc+1)}_{\text{term 0}} + a_1 \underbrace{(c\textstyle\sum_{k=1}^{K} m_k + m_K)}_{\text{term 1}} + a_2 \underbrace{(c\textstyle\sum_{k=1}^{K} m_k^2 + m_K^2)}_{\text{term 2}}.$$

For a 1-year bonds that pays no coupons and a 2-year bond that pays a 6% coupons at $m_1 = 1$ and $m_2 = 2$, we have the following matrix of regressors (the bonds are on different rows)

Bond ↓	*term 0*	*term 1*	*term 2*
1-year, 0%	1	1	1
2-year, 6%	$2 \times 0.06 + 1$	$0.06 \times (1+2) + 2$	$0.06 \times (1^2 + 2^2) + 2^2$
(=)	1.12	2.18	4.30.

The a_0, a_1, and a_2 can be estimated by OLS if we have data on at least two bonds. This method can, however, lead to large errors in the fitted yields (if not the prices).

Example 16.34 *(Cubic discount function**) With a cubic discount function*

$$B(m) = a_0 + a_1 m + a_2 m^2 + a_3 m^3,$$

we get

$$P\ (m_K) = a_0(Kc+1) + a_1(c\textstyle\sum_{k=1}^{K} m_k + m_K) + a_2(c\sum_{k=1}^{K} m_k^2 + m_K^2) \\ + a_3(c\textstyle\sum_{k=1}^{K} m_k^3 + m_K^3).$$

Empirical Example 16.35 *Figure 16.16 shows the estimation of the German yield curve for one trading day using a cubic discount function, based on a cross-section of government bonds.*

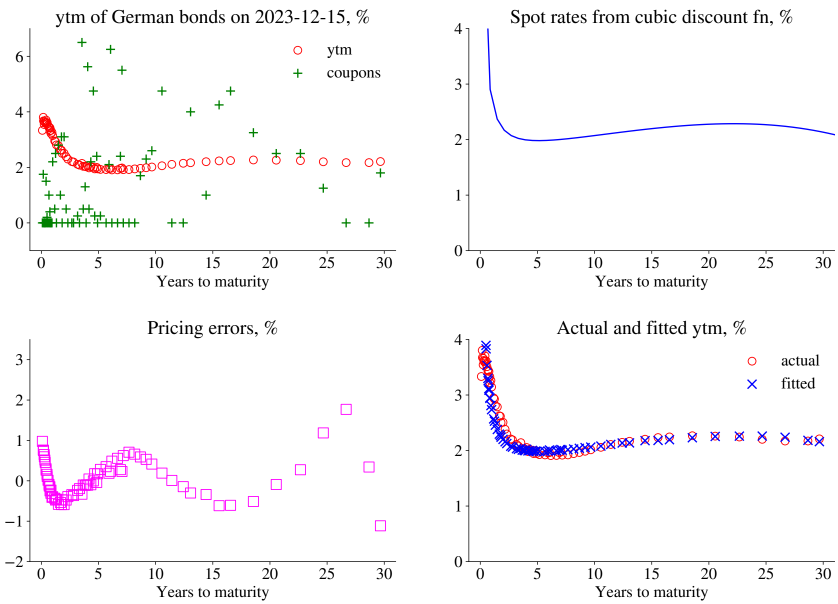


Figure 16.16: Estimated yield curves

16.5.3 Estimating a Parametric Forward Rate Curve*

Yet another approach to estimating the yield curve is to start by specifying a function for the instantaneous forward rate curve, and then calculate what this implies for the discount bond prices (discount function)

Let $f(m)$ denote the instantaneous forward rate with time to settlement m. The *extended Nelson and Siegel forward rate function* (Svensson (1995)) is

$$f(m) = \beta_0 + \beta_1 \exp(-m/\tau_1) + \beta_2 \frac{m}{\tau_1} \exp(-m/\tau_1) + \beta_3 \frac{m}{\tau_2} \exp(-m/\tau_2), \quad (16.20)$$

where $\beta_0, \beta_1, \beta_2, \tau_1, \beta_3, \tau_2$ are parameters (β_0, τ_1 and τ_2 must be positive, and $\beta_0 + \beta_1$ must also be positive—see below). The original Nelson and Siegel function sets $\beta_3 = 0$. Note that in either case

$$\lim_{m \to 0} f(m) = \beta_0 + \beta_1, \text{ and}$$

$$\lim_{m \to \infty} f(m) = \beta_0,$$

so $\beta_0 + \beta_1$ corresponds to the current very short spot interest rate (an overnight rate, say) and β_0 to the forward rate with settlement very far in the future (the asymptote).

The spot rate implied by (16.20) is (integrate to see that)

$$\begin{aligned} y(m) = \beta_0 + \beta_1 \frac{1 - \exp(-m/\tau_1)}{m/\tau_1} + \beta_2 \left[\frac{1 - \exp(-m/\tau_1)}{m/\tau_1} - \exp\left(-\frac{m}{\tau_1}\right) \right] \\ + \beta_3 \left[\frac{1 - \exp(-m/\tau_2)}{m/\tau_2} - \exp\left(-\frac{m}{\tau_2}\right) \right]. \quad (16.21) \end{aligned}$$

One way of estimating the parameters in (16.20) is to substitute (16.21) for the spot rate in (16.4), and then minimize the weighted sum of the squared price errors (differences between actual and fitted prices), with 1/maturity (or 1/duration) as the weights (a practice used by many central banks).

Empirical Example 16.36 *Figure 16.17 shows the estimation of the German yield curve for one trading day, using a Nelson-Siegel approach. Compare with Figure 16.16, especially the fitted rates at short maturities.*

16.5.4 Par Yield Curve

A par yield is the coupon rate at which a bond would trade at par (that is, have a price equal to the face value). Setting $P = 1$ in (16.13) and solving for the implied

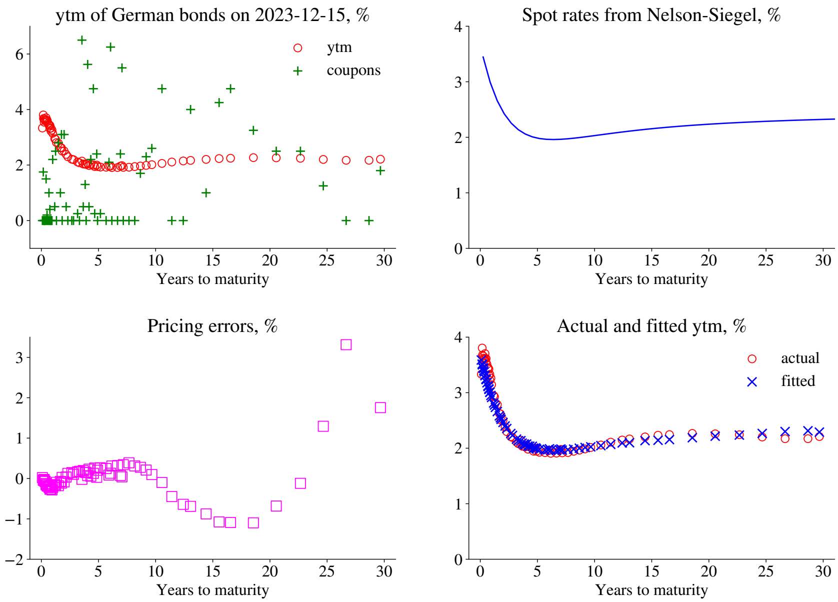


Figure 16.17: Estimated yield curves

coupon rate gives

$$c = \frac{1}{\sum_{k=1}^{K} B(m_k)} \left[1 - B(m_K)\right], \text{ or} \tag{16.22}$$

$$= \frac{1}{\sum_{k=1}^{K} \frac{1}{[1+Y(m_k)]^{m_k}}} \left[1 - \frac{1}{[1+Y(m_K)]^{m_K}}\right]. \tag{16.23}$$

Typically, this is very similar to the effective spot interest rates (on zero coupon bonds).

Example 16.37 *Suppose $B(1) = 0.95$ and $B(2) = 0.90$. We then have*

$$1 = (0.95 + 0.9)c + 0.9, \textit{ so } c = \frac{1}{0.95 + 0.9}(1 - 0.9) \approx 0.054.$$

When many bonds are traded at (approximately) par, the par yield curve (16.22) can be obtained by just plotting the coupon rates. In practice, the yield to maturity is used instead (to partly compensate for the fact that the bonds are only approximately at par)—and the gaps (across maturities) are filled by interpolation. (Recall that for a par bond, the yield to maturity equals the coupon rate.) This is basically the way the Constant Maturity Treasury yield curve, published by the US Treasury, is constructed.

16.5.5 Swap Rate Curve

The swap rates for different maturities can also be used to construct a yield curve.

16.6 Appendix – Conventions on Important Markets*

16.6.1 Compounding Frequency

Suppose the interest rate r is compounded n times per year. By comparing with the definition of the effective interest rate (with annual compounding) in (16.1) we have

$$1/B = (1 + r/n)^n = 1 + Y. \tag{16.24}$$

Clearly, as $n \to \infty$, the expression in (16.24) goes to e^y, where y is the continuously compounded rate.

This shows how we can transform from semi-annual ($n = 2$) or quarterly ($n = 4$) compounding to annual compounding (and vice versa).

16.6.2 US Treasury Notes and Bonds

The convention for *US Treasury notes and bonds* (issued with maturities longer than one year) is that coupons are paid semi-annually (as half the quoted coupon rate), and that yields are semi-annual effective yields. (This applies also to most US corporate bonds and UK Treasury bonds.)

However, both are quoted on an annual basis by multiplying by two. The quoted *yield to maturity*, ϕ, solves

$$P = \sum_{k=1}^{K} \frac{c/2}{(1+\phi/2)^{n_k}} + \frac{1}{(1+\phi/2)^{n_K}}, \tag{16.25}$$

where the bond pays coupons $c/2$, in $n_1, n_2, ..., n_K$ half-years. By using (16.24), the yield quoted, ϕ, can be expressed in terms of an annual effective rate.

Example 16.38 *A 9% US Treasury bond (the coupon rate is 9%, paid out as 4.5% semi-annually) with a yield to maturity of 7%, and one year to maturity has the price*

$$\frac{0.09/2}{1+0.07/2} + \frac{0.09/2}{(1+0.07/2)^2} + \frac{1}{(1+0.07/2)^2} = 1.019.$$

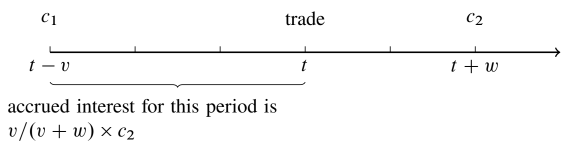


Figure 16.18: Accrued interest

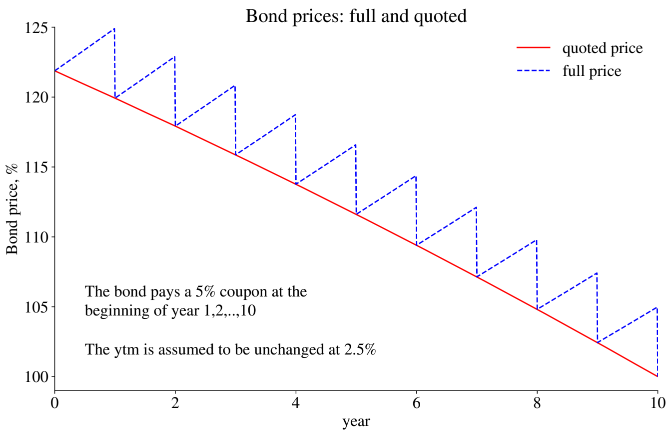


Figure 16.19: Full and quoted bond prices

From (16.24), we get that the yield to maturity rate expressed as an annual effective interest is $(1+0.035)^2-1\approx 0.071$.

16.6.3 Accrued Interest on Bonds

The quotes of bond prices (as opposed to yields) are not the full price (also called the dirty price, invoice price, or cash price) the investor pays. Instead, the full price is

full price = quoted price + accrued interest.

The buyer of the bond (buying in t) will typically get the next coupon (trading is "cum-dividend"). The accrued interest is the faction of that next coupon that has been accrued during the period the seller owned the bond. It is calculated as

accrued interest = next coupon $\times$ days since last coupon/days between coupons.

For instance, for US Treasury notes bonds, the next coupon is half the coupon rate and the days count uses actual days. See Figures 16.18 –16.19.

16.6.4 US Treasury Bills

Discount Yield

US Treasury bills have no coupons and are issued in 3, 6, 9, and 12 months maturities—but the time to maturity does, of course, change over time. They are quoted in terms of the (banker's) *discount yield*, Y_{db}, which satisfies

$$B = 1 - mY_{db}, \text{ where } m = \text{days}/360, \text{ so} \tag{16.26}$$

$$Y_{db} = (1 - B)/m. \tag{16.27}$$

Notice the convention of m =days/360. (If the face value is different from one, then we have $Y_{db} = [\text{face}-B]/(\text{face}\times m)$.)

From (16.1) and (16.26) it is the clear that the effective interest rate and the continuously compounded interest rates are

$$Y = [1 - mY_{db}]^{-1/m} - 1 \tag{16.28}$$

$$y = -\ln(1 - mY_{db})/m. \tag{16.29}$$

Sometimes, the bills are quoted in terms of a *bond equivalent yield*, which is the simple interest rate (16.5) but using the convention of 365 days per year.

Example 16.39 *A T-bill with 44 days to maturity and a quoted discount yield of 6.21% has the price* $1-(44/360)\times 0.0621 \approx 0.9924$. *The bond equivalent (simple) interest rate is* $(1/0.9924 - 1)365/44 \approx 6.35\%$.

16.6.5 European Bond Markets

The major continental European bond markets (in particular, France and Germany) typically have annual coupons and the accrued interest is calculated according to the "actual/actual" convention, that is, as

accrued interest = next coupon $\times$ days since last coupon/365 (or 366).

(The computation is slightly more complicated for the UK and the Scandinavian countries, since they have ex-dividend periods.)

16.6.6 Short Term Reference Rates

The short term reference rates in the U.S. (SOFR) and EU (ESTR), used in overnight indexed swaps (OIS) and other contracts are based on *backward looking* compounding of overnight (one day) rates. These overnight rates are considered almost risk-free (repos in the US, unsecured in the euro area) because they apply to large financial institutions and are short term.

The compounding is done by the formula

$$\text{compounded rate(over } d_b \text{ business days)} = [\Pi_{i=1}^{d_b}(1 + m_i \tilde{Y}_i) - 1]/m,$$

where $\tilde{Y}_i$ is a simple interest rate applicable over m_i of calendar days, measured as a fraction of the year ($m_i = 1/360$ or 1/365 if it's just one day, but 3/360 or 3/365 for Fridays and similarly for other business days followed by holidays), d_b is the number of business days and m the total number of calendar days as a fraction of the year ($m = 10/360$ or 10/365 if the contract spans 10 calendar days). Notice that the $1/m$ term makes this an annualised rate.

This formula is a mix of effective compounding over business days, since $1 + m_i \tilde{Y}_i = (1 + Y_i)^{m_i}$ where Y_i is an effective rate, and simple averaging on non-business days and with the scaling by $1/m$.

Remark 16.40 *(The traditional formula) The formula for the compounded rate is often written*

$$[\Pi_{i=1}^{d_b}(1 + \frac{n_i}{N}\tilde{Y}_i) - 1]\frac{N}{d_c},$$

where d_b is the number of business days, n_i the number of calendar days for which the rate $\tilde{Y}_i$ applies, N number of days per year (according to the market convention) and d_c the total number of calendar days. (Both FED and ECB use this expression.)

Example 16.41 *(Compounded rate) For two business days, the compounding in euro area or US (where $N = 360$) could be*

$$[(1 + 0.02/360)(1 + 0.03/360) - 1]360/2 \approx 0.025.$$

The difference to a simple average increases as the variability of the one-day rates does and the number of days increases.

16.7 Appendix – More Proofs and Details*

16.7.1 Proof and Details of Proposition 16.27

Proof (of Proposition 16.27) Consider a 2-year coupon bond with ytm θ. From (16.17), the price of the bond is

$$P = \frac{c}{1+\theta} + \frac{c+1}{(1+\theta)^2}.$$

If we can reinvest the first coupon payment to give the return θ, it is worth $c(1+\theta)$ are maturity—and we also receive $c+1$ at maturity. Divide the end value with the initial investment (the bond price P)

$$\frac{c(1+\theta)+c+1}{c/(1+\theta)+(c+1)/(1+\theta)^2} = (1+\theta)^2.$$

□

16.7.2 Proof and Details of the Return on Holding a Coupn Bond for s Periods

Proposition 16.42 *(Bond holding return, a special case) Suppose we reinvest the coupons with forward contracts—as if we were going to hold the bond until maturity m_K. Holding the bond until $t+s$ ($s \le m_K$) gives the total gross return $B_{t+s}(m_K - s)/B_t(m_K)$. This implies that the portfolio has the same return as an m_K-year zero coupon bond bought in t, which becomes an $m_K - s$ zero coupon bond in $t+s$.*

For instance, with a 3-year bond, the gross return on holding the bond for one year is $B_{t+1}(2)/B_t(3)$, while the gross return from holding it for two years is $B_{t+2}(1)/B_t(3)$.

Clearly, the strategy to reinvest the coupons with forward contracts essentially turns this into an m_K-year zero coupon bond (where you invest in t but do not receive any payoffs until $t+m_K$). The return of the strategy is thus the same as on holding this zero coupon bond for s years. Once again, with other assumptions about how the coupons are reinvested, the result is different.

Proof (of Proposition 16.42*) Consider a 3-year coupon bond which we hold for 1 year. Enter forward contracts like in the proof of Proposition 16.25. The value of this portfolio in $t+1$ must be the present value of the value at maturity, that is,

$$B_{t+1}(2)\left[\frac{B_t(1)}{B_t(3)}c + \frac{B_t(2)}{B_t(3)}c + c + 1\right],$$

where $B_{t+1}(2)$ denotes the price in $t+1$ of a two-year zero coupon bond. Dividing by the bond price in t

$$P_t = B_t(1)c + B_t(2)c + B_t(3)(c+1)$$

gives the gross return

$$1 + R_{t+1} = B_{t+1}(2)/B_t(3).$$

□

Example 16.43 *(Holding a coupon bond for one year) Use the same numbers as in Example 16.26 and assume that the interest rates are unchanged. The present value in $t+1$ of the value at maturity is*

$$0.962 \times 1.062 = 1.022.$$

Dividing by the bond price P_t, the gross return is

$$\frac{1.022}{0.963} \approx 1.06.$$

Using Proposition 16.42 directly gives $B_{t+1}(1)/B_t(2)$, which is approximately the same. Instead, if the interest rates change so $B_{t+1}(1) = 0.957$, then the return is $0.957 \times 1.062/0.963 \approx 1.055$, *which is the same as $B_{t+1}(1)/B_t(2)$.*

Notice that in the *special case* of holding the bond until maturity ($s = m_K$), then Proposition 16.42) shows that $1 + R_{t+s} = 1/B(m_K)$ (since $B_{t+s}(0) = 1$), which is the same result as in Proposition 16.25). In this case, the bond earns the spot interest rate $Y(m_K)$ per year.

Also, notice that in the very *special case* of a flat and unchanged yield curve (with the interest rate Y for all maturities), then then Proposition 16.42) shows that the return is

$$1 + R_{t+s} = (1 + Y)^s, \tag{16.30}$$

so the return is just accumulated interest rates. See Figure 16.20 for an illustration.

Remark 16.44 *(Realized forwards*) Sometimes another set of assumptions (labelled "realized forwards") is used to analyse the return on holding a coupon bond. In this case, the coupons are reinvested at the spot rates prevailing at the time of the coupon payment. However, it is assumed that those future spot rates will actually be equal to today's forward rates (hence "realized"). This is clearly unrealistic, but can be used to gauge the expected return on holding the bond, at least if today's forwards are close approximations of the expected future spot rates. The result is similar to Proposition 16.42.*

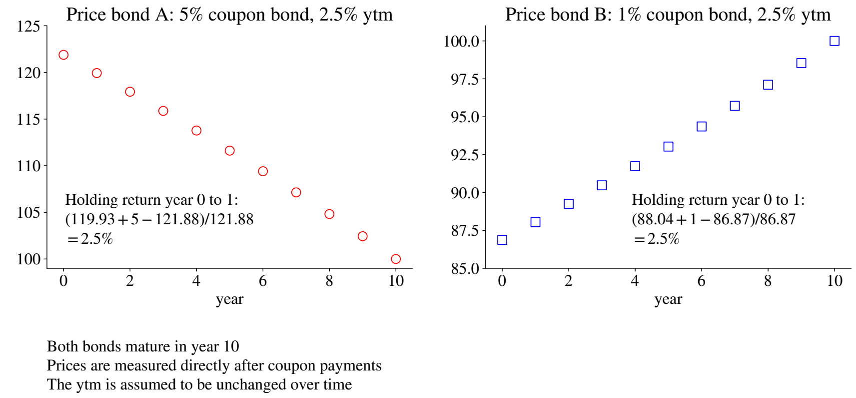


Figure 16.20: Bond price and yield to maturity

Chapter 17

Hedging Bonds

This chapter shows how we can hedge a liability stream against interest rate changes (*immunization*). The basic idea is to invest into a new portfolio that offsets the price changes of the liability. How to do that depends on the relative interest rate sentisitivity (duration) of the liability and the bond portfolio used as a hedge.

17.1 Bond Hedging

To simplify, the analysis is focused on changes over a short time period, and we often make strong assumptions about how the yield curve changes (for instance, only parallel movements).

Example 17.1 *(Why a liability is not hedged by putting its present value on a bank account) Suppose our liability is an annuity that pays 0.2 every year (starting a year from now) for 10 years. At a 5% interest rate for all maturities, the present value is*

$$\textstyle\sum_{k=1}^{10} 0.2/1.05^{k} = 1.54.$$

Instead, with an interest rate of 3%, the present value is

$$\textstyle\sum_{k=1}^{10} 0.2/1.03^{k} = 1.71.$$

Putting 1.54 on a bank account will not cover the liability payments if we only get a 3% interest rate.

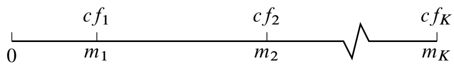


Figure 17.1: Timing convention of bond portfolio

17.2 Duration: Definitions

The *duration* of a bond portfolio is used to analyse how the price of the portfolio will change in response to changes in the yield curve. This section gives the definitions of the most commonly used duration measures.

Consider a bond portfolio with the cash flow cf_k in m_k years from now (for $k = 1...K$) as illustrated in Figure 17.1. Recall that the price P and the yield to maturity θ are related according to

$$P = \sum_{k=1}^{K} \frac{cf_k}{(1+\theta)^{m_k}}. \tag{17.1}$$

The change of the price, ΔP, due to a small change in the yield, $\Delta\theta$, is approximately

$$\Delta P \approx \frac{dP(\theta)}{d\theta} \times \Delta\theta, \tag{17.2}$$

where the derivative will be calculated below. We will later also discuss how/when it makes sense to think of changes in the yield to maturity as driving bond prices.

The *dollar duration*, $D^{\$}$, is defined as the *negative* of the derivative

$$D^{\$} = -\frac{dP(\theta)}{d\theta} \tag{17.3}$$

$$= \frac{1}{1+\theta} \sum_{k=1}^{K} m_k \frac{cf_k}{(1+\theta)^{m_k}}. \tag{17.4}$$

To calculate the dollar duration $D^{\$}$, we need all the cash flows and the times at which they occur (cf_k and m_k for $k = 1$ to K), as well as the yield to maturity (θ). The latter is typically calculated by (numerically) solving (17.1) for θ.

Remark 17.2 *(Sign of $D^{\$}$) Warning: some authors do not switch the sign in defining $D^{\$}$, only when defining the other durations (see below).*

The approximate change of the price in (17.2) can then be written

$$\Delta P \approx -D^{\$} \times \Delta\theta. \tag{17.5}$$

This expression says that an increase in the interest rate (more precisely, the yield to maturity, θ) translates into a decrease in the price—and more so if the duration ($D^{\$}$) is long.

Multiply the dollar duration with $(1+\theta)/P$ to get *Macaulay's duration* (D^M)

$$D^M = D^{\$}(1+\theta)/P \tag{17.6}$$

$$= \sum_{k=1}^{K} w_k m_k \text{, where } w_k = \frac{cf_k}{(1+\theta)^{m_k} P}. \tag{17.7}$$

This is a weighted average of the times to the cash flows ($m_1, m_2, ..., m_K$), where the weight w_k is the fraction of the bond price accounted for by the payment in m_k ($cf_k/[(1+\theta)^{m_k} P]$). The weights sum to unity. See Elton, Gruber, Brown, and Goetzmann (2014) 21–22, Hull (2022) 4 and McDonald (2014) 9 for more detailed discussions.

Macaulay's duration is an average "time to payment" of the bond portfolio. For bond portfolios with coupons or other intermediate payments (payment of the face value of some of the bonds in the portfolio) before the last one, Macaulay's duration is less than the time to maturity, and this effect is more pronounced at large intermediate payments and at high yield to maturity. In contrast, for zero coupon bonds, Macaulay's duration equals the time to maturity. This is illustrated in Figure 17.2.

For hedging purposes, it is often more convenient to work with the *modified duration* (D, sometimes also called the adjusted duration) which is

$$D = D^M/(1+\theta). \tag{17.8}$$

(Or, equivalently, $D = D^{\$}/P$.) This allows us to express the (approximate) relative change of the price (return) due to a small change in the yield as

$$\Delta P/P \approx -D \times \Delta\theta. \tag{17.9}$$

(To see this, use (17.5), divide both sides by P.) Clearly, the modified and Macaulay's durations are similar, only differing with the factor $1+\theta$, which is typically close to 1.

Remark 17.3 *(Duration of a zero coupon bond) For a zero-coupon bond with a face value of unity and time to maturity m, the price is $B = 1/(1+\theta)^m$, where θ is* *the yield to maturity. The duration measures are*

$$D^{\$} = mB/(1+\theta),\ D^{M} = m, \text{ and } D = m/(1+\theta).$$

In particular, Macaulay's duration is the same as the maturity.

Remark 17.4 *(Duration of a portfolio*) The dollar duration of a portfolio of n_i units of bond i and n_j of bond j is $D^{\$} = n_i D_i^{\$} + n_j D_j^{\$}$. Divide by the value of the portfolio, $P = n_i P_i + n_j P_j$, and rearrange to get he modified duration $D = w_i D_i + w_j D_j$, where $w_i = n_i P_i / P$. This can be used in (17.9) if $\Delta\theta_i = \Delta\theta_j$. Otherwise, we need to calculate the ytm of the portfolio and consider a change of that.*

Remark 17.5 *(Continuously compounded ytm*) The price in (17.1) can also be written $P = \Sigma_{k=1}^{K} cf_k e^{-m_k \gamma}$, where γ is the continuously compounded ytm. The derivative is $dP/d\gamma = \Sigma_{k=1}^{K} m_k cf_k e^{-m_k \gamma}$, so the duration is $D = -(dP/d\gamma)/P$, which can be used in an expression similar to (17.9), $\Delta P/P \approx -D \times \Delta\gamma$. This means that, with a continuously compounded ytm, the modified and Macaulay's durations are the same.*

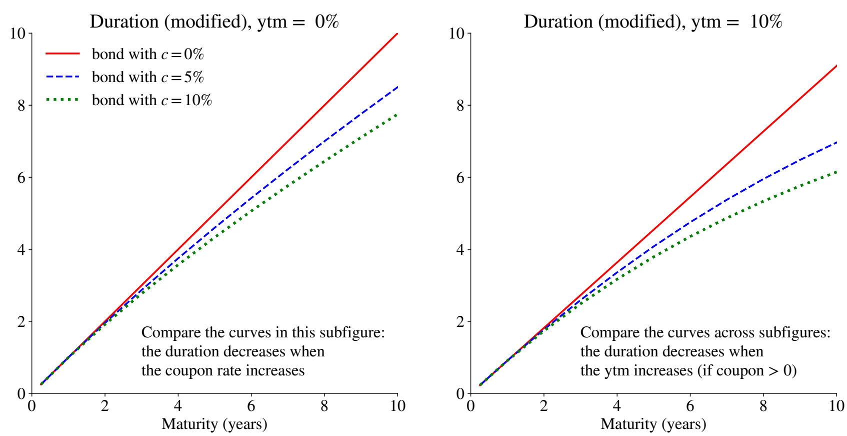


Figure 17.2: Macaulay's duration

Example 17.6 *(Duration) The liability in Example 17.1 has a yield to maturity (ytm) of 5% under the assumption that all interest rates are 5%. Macaulay's and the* *modified durations are*

$$D^M = \sum_{k=1}^{10} k \frac{0.2}{1.05^k \times 1.54} \approx 5.1 \textit{ and } D \approx 4.86.$$

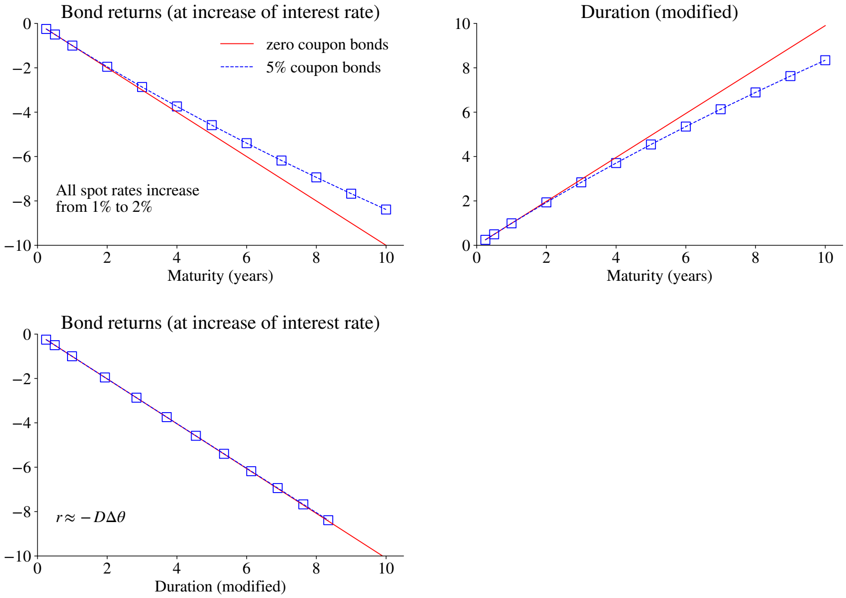


Figure 17.3: Returns after interest rate changes

Example 17.7 *(Approximate price change) When the ytm changes from 5% to 3%, then (17.9) says that the liability in Examples 17.1 and 17.6 has a relative value change of*

$$\Delta P/P = -4.86 \times -0.02 \approx 0.097.$$

From Example 17.1, we know that the exact change is $(1.71 - 1.54)/1.54 = 0.105$.

17.3 Hedging a Bond Portfolio

17.3.1 Basic Setup

This sections considers how to hedge a liability. A liability is the same as being short a bond portfolio with price P_L. We will hedge this portfolio against interest rate changes by buying v units of another bond portfolio, denoted H, with price P_H.

The value of the overall position is then

$$V = vP_H + M - P_L, \tag{17.10}$$

where M is a short-term money market account. The choice of M is typically such that the initial value of V is zero, that is, on the first day of the hedge. The subsequent amount on the money market account will change as payments are made and received and the valuation of the bonds change (the positions are marked-to-market, perhaps every day). The portfolio will typically have to be rebalanced over time in order to stay hedged.

In a first step, we choose the hedge (H) bond portfolio. Choosing a bond portfolio with a duration similar to the liability is typically a good idea. In a second step, we find v so that vP_H and P_L are equally sensitive to changes in interest rates.

One way of hedging is to hold a bond portfolio so as to *match every cash flow* of the liability, so portfolios L and H are identical ($v = 1$ and $M = 0$). However, that may be both difficult and costly because of transaction costs. The subsequent analysis will therefore focus on a case where we buy some other bond portfolio H to use as a hedge.

Using the approximate relation of the (bond portfolio) price change (17.9), we have that the change of value, due to a sudden change in the interest rates, of the overall position is

$$\Delta V = v\Delta P_H - \Delta P_L \tag{17.11}$$

$$\approx -vD_H P_H \times \Delta\theta_H + D_L P_L \times \Delta\theta_L, \tag{17.12}$$

where the durations are modified durations. See Figure 17.3 for an illustration of the fact that bonds portfolios with the same *duration* (not maturity) react similarly to interest rate changes.

17.3.2 Yield Curve Shifts and Yield to Maturity

The yield to maturity θ depends on the yield curve, so $\Delta\theta_H$ and $\Delta\theta_L$ in (17.12) may be different. For certain yield curve changes, the effect on θ is fairly straightforward. In particular, several of the hedging approaches discussed below assume that $\Delta\theta_L = \Delta\theta_H$, that is, a parallel shift of the yield curve. This section discusses how that is related to general yield curve changes.

First, the simplest case is when the yield curve is flat, meaning that spot interest

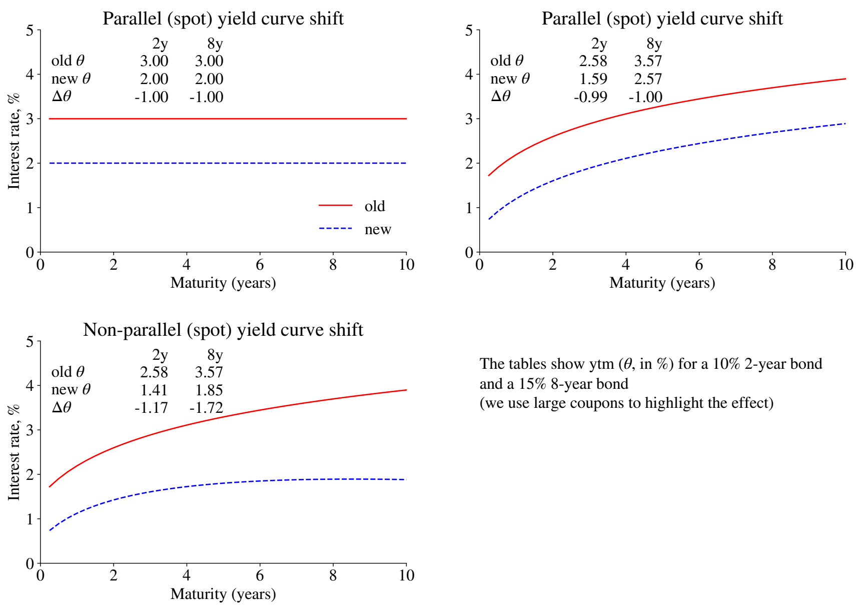


Figure 17.4: Yield curve shifts

rates (that is, interest rates on zero-coupon bonds) are the same across all maturities, and shift in parallel. Then all ytms will change equally much, as they are equal to *the* interest rate. See Figure 17.4, upper left subfigure, for an illustration. Second, when the yield curve is not flat but the shift is parallel, then ytms will change approximately the same (see upper right subfigure of Figure 17.4). (See Sundaresan (2009) 4 for a formal argument.) Thirdly, when the yield curve shift is non-parallel, then ytms will *not* change the same (see lower left subfigure of Figure 17.4).

Example 17.8 *(ytm changes) Suppose bond A pays 1 in one year and another 1 in two years, while bond B is a 2-year zero coupon bond. In the parallel yield curve shift (from the base case to scenario 1, see the 1- and 2-year spot rates $y(1)$ and $y(2)$), both ytms change by approximately 1 percentage point. In the nonparallel shift (from the base case to scenario 2), only the ytm for bond A changes.*

	$y(1)$	$y(2)$	θ_A	$\Delta\theta_A$	θ_B	$\Delta\theta_B$
Base case:	3%	2%	2.34		2%	
Scenario 1:	2%	1%	1.33%	−1.01%	1%	−1%
Scenario 2:	2%	2%	2%	−0.34%	2%	0%

17.3.3 Duration Matching

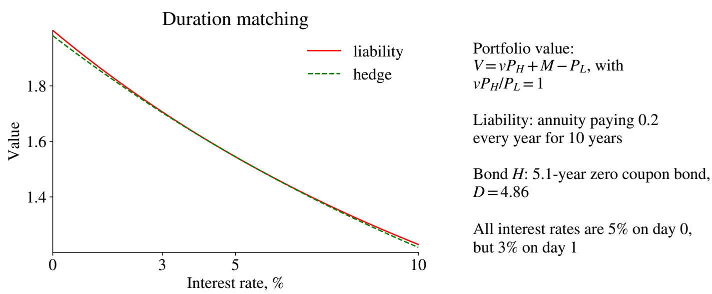
In this case, we choose a hedge bond (portfolio) with the same duration as the liability ($D_H = D_L$), and invest the same amount in the hedge bond as the value of the liability ($vP_H = P_L$). This means that the initial position on the money market account is zero. While the two bonds have the same durations, their cash flow streams might differ.

If the yield curve shifts up in an almost parallel fashion, so $\Delta\theta_L = \Delta\theta_H$, then (17.12) gives

$$\Delta V / P_L \approx 0, \tag{17.13}$$

so the duration hedge makes the overall portfolio approximately immune to interest rate changes.

As interest rates change, the durations do too. This means that a hedge bond that had the same duration as the liability in t may not be a duration match in a later period. This requires either switching hedge bond or to move over to a duration hedging (discussed below).



	θ	D_L	P_H	v	vP_H	M	$-P_L$	V	vP_H/P_L
Day 0	0.05	4.86	0.78	1.98	1.54	0.00	-1.54	0.00	1.00
Day 1 (m-to-m)	0.03	5.10	0.86	1.98	1.70	0.00	-1.71	-0.00	1.00

The duration of the hedge bond on day 1 is 4.95, the hedging has to be redone

Figure 17.5: Example of duration matching. "m-to-m" stands for the marking-to-market stage

Example 17.9 *(Duration matching) Figure 17.5 illustrates a case where a liability stream is hedged by a (here, zero-coupon) bond with the same duration. The interest rate (same for all maturities) decreases from 5% to 3%, so both the liability and the hedge increase in value, by very similar amounts, so the value of V stays close to* *zero. Notice that, after the ytm change, the liability and hedge bond have different durations, so the portfolio needs to be changed to stay hedged: another hedge bond or apply a duration hedge (see below).*

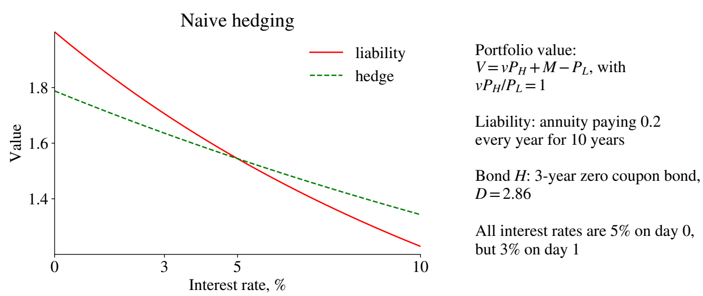
17.3.4 Naive Hedging

Suppose we again invest the same amount in the hedge bond as the value of the liability ($vP_H = P_L$), but this time we do not pay any attention to the durations.

This will typically make the overall portfolio vulnerable to interest rate changes. To illustrate that, assume, for simplicity, that the yield curve shifts up in a parallel fashion, so $\Delta\theta_L = \Delta\theta_H$. Then (17.12) gives

$$\Delta V/P_L \approx (D_L - D_H) \times \Delta\theta, \tag{17.14}$$

which shows that the portfolio value is sensitive to interest changes if there is a duration mismatch. See Figure 17.6 for an example.



	θ	D_L	P_H	v	vP_H	M	$-P_L$	V	vP_H/P_L
Day 0	0.05	4.86	0.86	1.79	1.54	0.00	-1.54	0.00	1.00
Day 1 (m-to-m)	0.03	5.10	0.92	1.79	1.64	0.00	-1.71	-0.07	0.96

Figure 17.6: Example of naive hedging

Example 17.10 *(Naive hedging) Figure 17.6 shows a case of naive hedging when we have a duration mismatch. This makes the overall portfolio (V) sensitive to interest rate changes. In this case, interest rates decrease (from 5% to 3%) so the liability (with a long duration) increases more in value than the hedge bond (with a short duration). In terms of (17.14), we have $D_L > D_H$ and $\Delta\theta < 0$.*

Remark 17.11 *(Effect of yield curve shift on a bank) A bank typically has liabilities with short duration (deposits, inter-bank lending) and assets (plays the same role* *as the "hedge" above) with long duration (loans to companies and households), so $D_L < D_H$. Equation (17.14) then shows that an increase in the interest rate level ($\Delta\theta > 0$) will hurt the bank, since the assets decrease more than the liabilities. This can also be phrased as follows: the bank has fixed incomes from the loans it has made, but it now needs to refinance itself (deposits and inter-bank loans) at a higher cost.*

Remark 17.12 *(Effect of yield curve shift on a pension fund) A pension fund (with defined benefits) typically has liabilities with very long durations and assets with somewhat shorter duration ($D_L > D_H$). This is similar to Example 17.10, although the durations are typically much longer. For instance, $(D_L, D_H) = (25, 15)$, $\Delta\theta = -1\%$ would give a return of -10%, which highlights the need for careful hedging (or moving to a defined contributions setup).*

17.3.5 Duration Hedging

Instead of the naive hedge, suppose we instead choose to offset the duration differences by the size of the position (v), so that

$$v = \frac{D_L P_L}{D_H P_H}, \text{ that is, } \frac{v P_H}{P_L} = \frac{D_L}{D_H}. \tag{17.15}$$

For instance, with $D_H < D_L$ this suggests that the *amount* invested into the hedge bond ($v P_H$) should exceed the value of the liability (P_L). In this way, by having a larger position, we offset the hedge bond's lower interest rate sensitivity. The initial position on the money market account is typically nonzero. As in the other cases, the portfolio needs to be rebalanced over time.

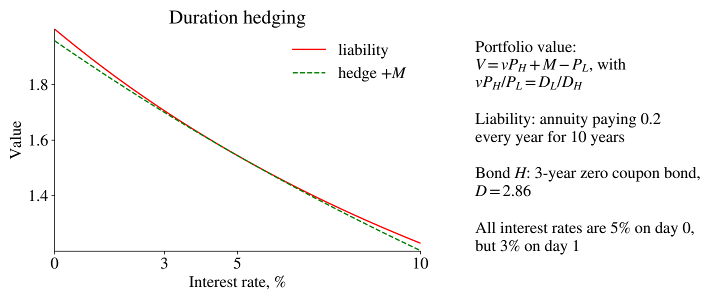
Combine (17.12) and the hedge ratio (17.15) to get

$$\Delta V / P_L \approx D_L \times (\Delta\theta_L - \Delta\theta_H). \tag{17.16}$$

Suppose again that the yield curve shifts in a parallel fashion. Then, (17.16) shows that the overall portfolio value will not change ($\Delta V / P_L \approx 0$). See Figure 17.7 for an example how the duration hedging works.

Example 17.13 *(Duration hedging) Figure 17.7 illustrates a case where we have a duration mismatch (similar to the case of naive hedging), but where this compensated for by a hedge ratio that takes the mismatch into account. The hedge bond has a too short duration, we therefore take a larger position in it (the amount invested into the* *hedge bond, vP_H, is much larger than the value of the liability)—so as to increase the interest rate sensitivity of the position.*

Empirical Example 17.14 *Figure 17.8 shows an example based on the German yield. The value of the (artificial) liability is calculated by using estimated yield curves for each trading day. In contrast, the hedge bond is one of the bonds in the data set. Notice that the duration of the liability jumps up just after a cash flow has been made. (The average time to future cash flows is then longer.)*



	θ	D_L	P_H	v	vP_H	M	$-P_L$	V	vP_H/P_L
Day 0	0.05	4.86	0.86	3.04	2.62	-1.08	-1.54	0.00	1.70
Day 1 (m-to-m)	0.03	5.10	0.92	3.04	2.78	-1.08	-1.71	-0.01	1.63
Day 1 (rebal)	0.03	5.10	0.92	3.27	2.99	-1.29	-1.71	-0.01	1.75

Figure 17.7: Example of duration hedging

Remark 17.15 *(Using the dollar duration instead*) Recall that $DP = D^{\$}$, the hedge ratio in (17.15) can also be written $v = D_L^{\$}/D_H^{\$}$.*

17.4 Addressing Issues in Duration Hedging

This section discusses potential problems with the duration hedging.

17.4.1 Problem 1: Approximation Error

The formula for the price change (17.2) is a first-order Taylor approximation of the form

$$\Delta P \approx \frac{dP}{d\theta} \times \Delta\theta. \tag{17.17}$$

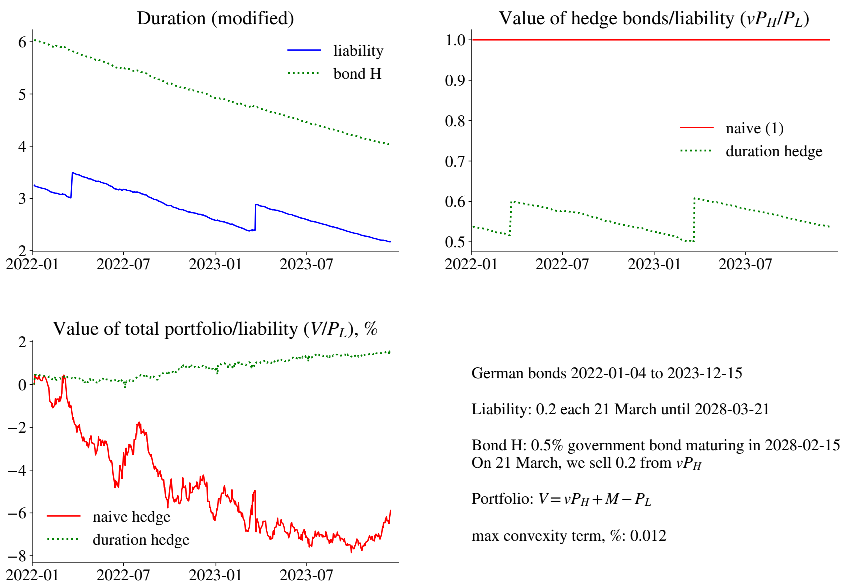


Figure 17.8: Duration hedging

Obviously, a second-order Taylor approximation is more precise. It would be

$$\Delta P \approx \frac{dP}{d\theta} \times \Delta\theta + \frac{1}{2}\frac{d^2P}{d\theta^2} \times (\Delta\theta)^2, \tag{17.18}$$

where the last term includes the second derivative of the bond price with respect to the yield to maturity. In Figure 17.7 the non-linear effect appears to be important only for very large interest changes, but that might be different for bonds with very long durations.

Dividing (17.18) by the bond price and using (17.9) gives

$$\Delta P/P \approx -D \times \Delta\theta + \frac{1}{2}C \times (\Delta\theta)^2, \tag{17.19}$$

where C (often called "convexity") is the second derivative in (17.18) divided by the bond price.

The convexity is easily calculated as

$$C = \frac{1}{P}\sum_{k=1}^{K} m_k(m_k+1)\frac{cf_k}{(1+\theta)^{m_k+2}}. \tag{17.20}$$

It is clear that the convexity is positive (since $cf_k \geq 0$), but tends to be lower if much of the cash flow comes early, similarly to the duration. Often, the convexity effect is modest compared to the duration effect, at least for bonds with short duration, see Figure 17.9. Still, choosing the hedging bond (portfolio) so that it has a similar convexity to the bond portfolio to be hedged may reduce the approximation error.

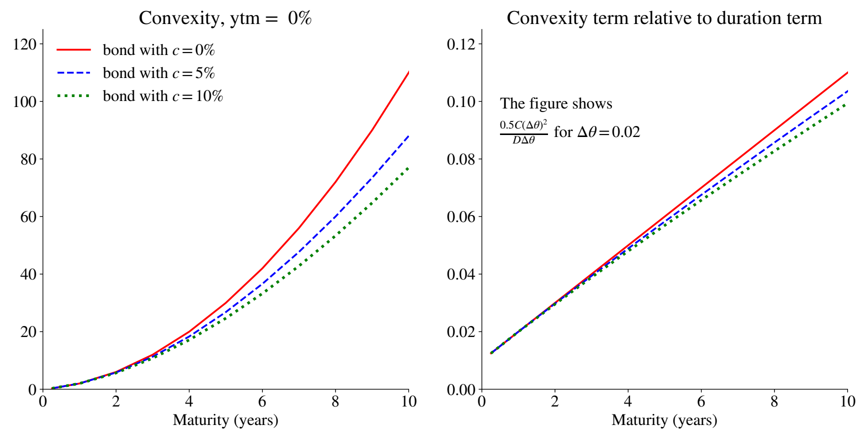


Figure 17.9: Convexity

Example 17.16 *(Convexity) The convexity of the 10-year bond in Example 17.1 is (when interest rates are 5%)*

$$C = \frac{1}{1.54} \sum_{k=1}^{10} k(k+1) \frac{0.2}{1.05^{k+2}} \approx 35.6.$$

If interest rates decrease from 5% to 3%, then the second-order term in (17.19) is

$$\frac{35.6}{2} \times 0.02^2 = 0.007,$$

which is fairly small compared to the duration effect (see Example 17.7).

17.4.2 Problem 2: Changing Cash Flows

The duration measures assume that the cash flow is unaffected by the yield change. That is true for many instruments, like most government bonds, but not for callable bonds and effectively not for bonds with (time varying) default risk. In such cases, another approach is needed.

17.4.3 Problem 3: Yield Curve Changes vs. Changes in Yields to Maturity

An important problem with using duration for hedging is that the hedge ratio in (17.21) depends on how the yields change.

The ideal case for duration hedging is when the yields to maturity move in parallel. In reality, level shifts of the entire yield curve make up a sizeable fraction of the overall variability of the curve. However, there are also other important aspects, for instance, changes in the slope of the curve.

Equation (17.16) shows how the value of the overall portfolio depends on the yields of the liability and the hedge bond. For instance, suppose the yield curve changes from being flat to being downward sloping and the hedging bond has shorter duration than the liability. In this case, the overall portfolio loses value. The reason is that the value of the hedging portfolio increases less, as the yield decreases less, in price than the liability. See Figure 17.4 for an illustration.

To overcome this problem, the hedge ratio should be (set $\Delta V = 0$ in (17.12))

$$v = \frac{D_L}{D_H} \times \frac{P_L}{P_H} \times \frac{\Delta\theta_L}{\Delta\theta_H}. \tag{17.21}$$

This is consistent with the duration hedging equation (17.15) when all changes of the yield curve are parallel shifts, rendering the last term in (17.21) equal to unity. Otherwise, we need to model how the yield curve changes (level, slope, curvature) in response to the overall economic situation.

Chapter 18

Interest Rate Models

This chapter presents a factor model that aims to describe movements in the entire yield curve (a simplified version of the Vasicek model). As an application, it is used to improve the hedging of a liability stream.

18.1 Empirical Properties of Yield Curves

Yield curves in the US and most other developed countries tend to exhibit the following features: first, the yield curve is usually upward sloping; second, it changes over time, primarily due to general level shifts but occasionally due to changes in its slope.

Empirical Example 18.1 *Figure 18.1 show U.S. yield curves at the beginning of some selected years. There are important shifts in both the level and slope.*

Yield curve movements are commonly described in terms of three factors: level, slope, and curvature. One way of measuring these factors is by defining

$$\begin{aligned}
\text{Level} &= y(\text{10-year}) \\
\text{Slope} &= y(\text{10-year}) - y(\text{3-month}) \\
\text{Curvature} &= [y(\text{2-year}) - y(\text{3-month})] - [y(\text{10-year}) - y(\text{2-year})]. \qquad (18.1)
\end{aligned}$$

This means that we measure the level by a long rate, the slope by the difference between a long (maturity) and a short (maturity) rate—and the curvature (or rather, concavity) by how much the medium/short spread exceeds the long/medium spread.

Most evidence from US data suggests that changes in the level factor dominate, and the slope ranks second, while the curvature is fairly unimportant.

Empirical Example 18.2 *Figure 18.2 shows the U.S. yield curve factors over time. The level shifts seem to dominate, as indicated by the standard deviations. (Notice that the subfigures use different scales of the vertical axis.) It can also be noticed that the slope factor is often very small or even negative at the beginning of recessions (shaded areas) and then increases towards the end.*

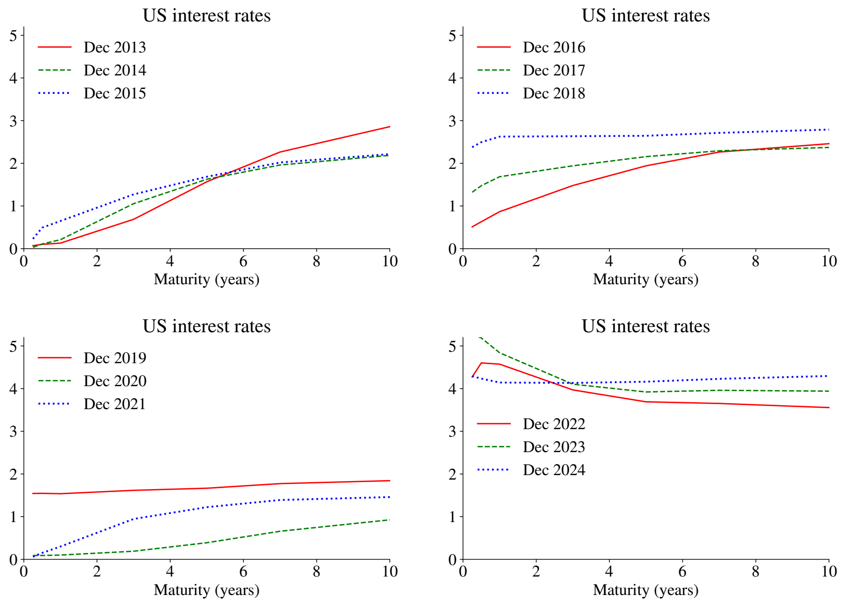


Figure 18.1: Estimated yield curves

18.2 Yield Curve Models

Yield curve models aim to describe the dynamics of the yield curve. Such models can, among other things, improve the hedging of bond portfolios. The previous empirical evidence suggests that accounting for a level (parallel) shift is important, but that we should also try to model changes in the slope. The curvature and further factors might be less important.

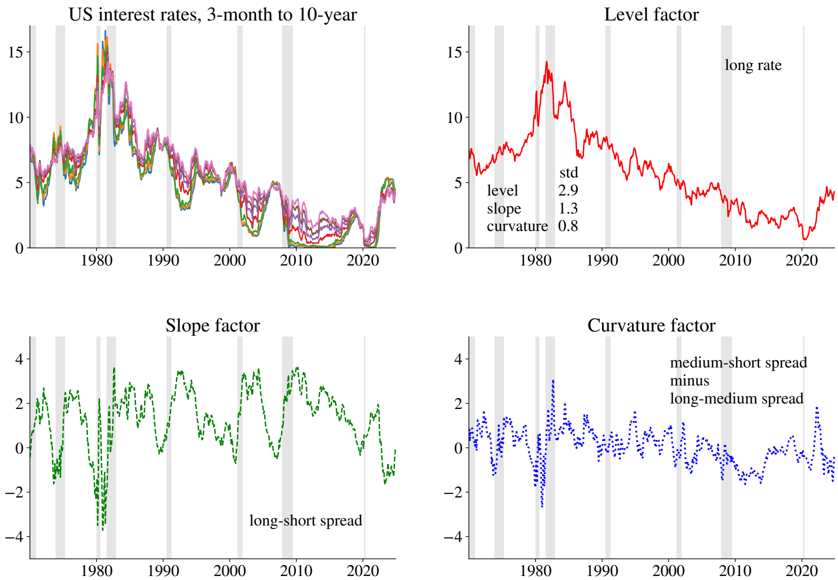


Figure 18.2: US yield curves: level, slope and curvature. NBER recessions are marked by shaded areas.

18.2.1 The Expectations Hypothesis of Interest Rates

The expectations hypothesis (EH) of interest rates posits that long bonds either have no risk premia or only a constant risk premium. The empirical evidence is mixed, so the expectations hypothesis is best thought of as an approximation.

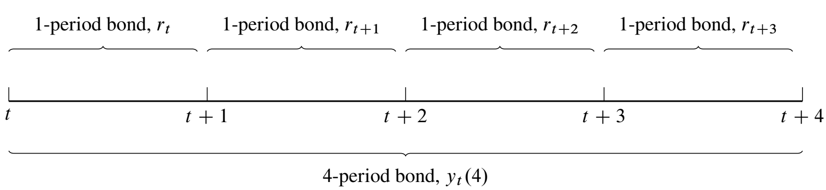


Figure 18.3: Timing for expectations hypothesis

EH implies that the n-period interest rate $y(n)$ equals the average of the 1-period (the shortest maturity) rates over t to $t+n$

$$y_t(n) = \lambda(n) + \textstyle\sum_{s=0}^{n-1} \mathrm{E}_t\, r_{t+s}/n, \tag{18.2}$$

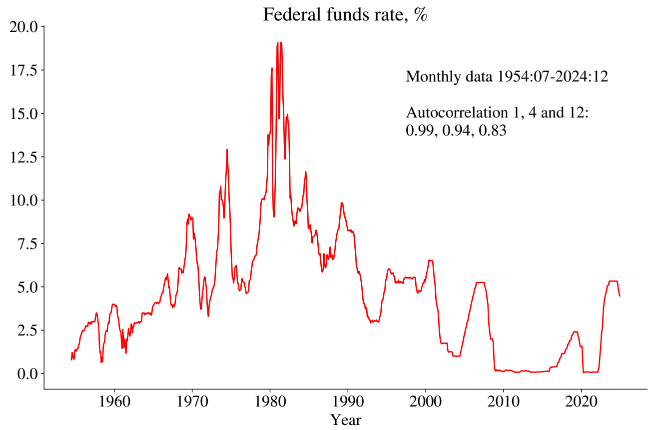


Figure 18.4: Federal funds rate, monthly data

where r_t is short hand notation for the 1-period rate and where $\lambda(n)$ may vary with n but is assumed constant over time. If $\lambda(n) = 0$, then the *pure* expectations hypothesis is said to hold. See Figure 18.3 for an illustration.

In (18.2), the period length (from t to $t+1$) corresponds to the maturity of the short interest rate. For instance, if r_t is a 1-month rate today, then r_{t+1} is the 1-month rate a month later and $y_t(120)$ is today's 120-month (10 year) interest rate. As usual, all interest rates are annualized rates of returns of keeping the bond until maturity. These features require some care when using $y_t(n)$ in bond pricing formulas.

Example 18.3 *(The expectations hypothesis) Suppose* $(r_t, \mathrm{E}_t\, r_{t+1}, \mathrm{E}_t\, r_{t+2}, \mathrm{E}_t\, r_{t+3})$ *equal* $(3\%, 2\%, 2\%, 1\%)$*, then the 4-month rate is* $\lambda(4) + 2\%$*.*

18.2.2 Risk Premia

There are several reasons for why bonds should have risk premia. First, long bonds are risky for investors who do not intend to hold them until maturity and therefore carry term premia. Second, some bonds are infrequently traded (for instance, off-the-run bonds and many index-linked bonds) and are likely to have liquidity risk premia. Third, the real return of a long bond is very sensitive to inflation changes, likely more so than equities. Bonds are therefore likely to have inflation risk premia. In general, the typical upward sloping yield curve observed in data is consistent with the view that long-maturity bonds are considered risky.

18.2.3 A Simple One-Factor Model: The Vasicek Model

The Vasicek (1977) model uses a single factor to model the entire yield curve: the short interest rate, which is assumed to follow a first-order autoregression, AR(1).

To present a simplified version of the model, the current section applies some unspecified constant risk premia. The more general formulation (discussed in an appendix) derives the risk in terms of the mean reversion and volatility of the short rate.

To simplify the notation, let the short interest rate, r_t, follow an AR(1)

$$r_{t+1} - \mu = \rho\,(r_t - \mu) + \varepsilon_{t+1}, \tag{18.3}$$

where μ is the average short interest rate, and ρ describes the dynamics. Typically, we consider the mean-reverting (stationary) case when $0 < \rho < 1$, but we will also discuss the borderline case of $\rho = 1$.

Empirical Example 18.4 *Figure 18.4 shows how the U.S. Federal Funds rate has developed over time. It shows significant persistence.*

Remark 18.5 *(Alternative formulation of (18.3)*) The process is sometimes specified in terms of changes as $r_{t+1} - r_t = a\,(r_t - \mu) + \varepsilon_{t+1}$. Clearly, this can be written $r_{t+1} - \mu = (1 + a)\,(r_t - \mu) + \varepsilon_{t+1}$, where $1 + a$ corresponds to ρ in (18.3). With $-1 < a < 0$ (that is, with $0 < \rho < 1$) the process is mean reverting.*

The forecast made in t of r_{t+s} is

$$\mathrm{E}_t\, r_{t+s} = (1 - \rho^s)\,\mu + \rho^s r_t, \tag{18.4}$$

where E_t denotes expectations formed in t. Notice that when r_t is a 1-month rate, then (18.4) is today's expectation of the 1-month rate in s months. See Figure 18.5 for an illustration of how these expectations depend on the starting value r_t and the horizon (for some specific (μ, ρ) parameters).

Example 18.6 *(Predictions from an AR(1)) With $\mu = 0.05$, $\rho = 0.975$ and $r_t = 0.07$, then $\mathrm{E}_t\, r_{t+50} = 0.72 \times 0.05 + 0.28 \times 0.07 = 0.056$.*

Remark 18.7 *(Calibrating the AR(1) to data*) Notice that (18.3) implies that the correlation is $\operatorname{Corr}(r_t, r_{t-s}) = \rho^s$, so we could estimate ρ by $\operatorname{Corr}(r_t, r_{t-s})^{1/s}$. If the AR(1) is a very good fit to data, then it should not matter (much) if you use*

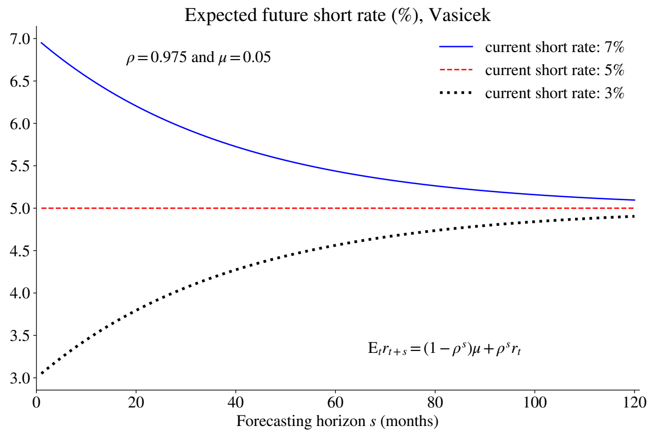


Figure 18.5: Expected future short rate in Vasicek model, for different initial short rates

$s = 1$ *or* $s = 12$ *(say). In practice, the results may differ. For instance, suppose monthly data gives* $\text{Corr}(r_t, r_{t-1}) = 0.99$ *but* $\text{Corr}(r_t, r_{t-12}) = 0.80$*, which imply* $\rho = 0.99$ *and* $\rho = 0.982$ *respectively. This matters for the pricing of long-maturity bonds: with 120 months (10 years) we get* $0.99^{120} \approx 0.3$ *while* $0.982^{120} \approx 0.11$*. Which value we choose to use depends on whether we are most interested in short maturities (use* $\rho = 0.99$*) or long maturities (use* $\rho = 0.982$*).*

We now assume that the expectations hypothesis holds for continuously compounded rates. Using this in (18.2) gives the long interest rate. For instance, the two-period (annualized, continuously compounded) rate is

$$\begin{aligned} y_t(2) &= \lambda(2) + [r_t + (1-\rho)\,\mu + \rho r_t]/2 \\ &= \lambda(2) + \mu\,(1-\rho)/2 + r_t\,(1+\rho)/2, \end{aligned} \tag{18.5}$$

where we have collected the terms that are constant first and those that involve r_t last. The general expression for a maturity of n periods is

$$\begin{aligned} y_t(n) &= a(n) + b(n) r_t, \text{ where} \\ a(n) &= \lambda(n) + \mu\,[1 - b(n)] \text{ and} \\ b(n) &= (1 + \rho + \ldots + \rho^{n-1})/n = (1-\rho^n)/[(1-\rho)n]. \end{aligned} \tag{18.6}$$

Again, notice that the period length is defined by the maturity of the short rate. For instance, when r_t is a 1-month rate, $y_t(120)$ is a 120-month (10 year) rate.

Remark 18.8 *(*A recursive expression for $b(n)$) Equation (18.6) implies $b(n) = [1 + \rho(n-1)b(n-1)]/n$, where the recursion starts at $b(1) = 1$.*

In this model, all movements of the yield curve are driven by the short rate, so it is a *one-factor model*. The shifts of the yield curve are parallel if $\rho = 1$ (the random walk model) since then $b(n) = 1$ in (18.6), so we get

$$y_t(n) = \lambda(n) + r_t, \text{ if } \rho = 1. \tag{18.7}$$

For lower values of ρ, the short rate process r_t is mean-reverting, so the expected future short rates (and therefore the current long rates) are always closer to the mean than the current short rate. This means that the level and slope are related: at high short rates, the slope is negative (disregarding any constant risk premia) and vice versa. See Figures 18.6–18.7 for an illustration. Also, see Hull (2022) 31 for a more detailed discussion.

Example 18.9 *(Vasicek model) For $\rho = 0.975$ and $\mu = 0.05$, (18.6) gives (assuming no risk premia)*

$$\begin{bmatrix} y_t(1) \\ y_t(2) \\ y_t(3) \\ y_t(4) \end{bmatrix} \approx \begin{bmatrix} 0 \\ 0.00062 \\ 0.00124 \\ 0.00184 \end{bmatrix} + \begin{bmatrix} 1 \\ 0.988 \\ 0.975 \\ 0.963 \end{bmatrix} r_t.$$

Proof (of (18.6)) The general expression (for the case when $\lambda = 0$) is $\mathrm{E}_t\, r_{t+s} = \Sigma_{s=0}^{n-1}[(1-\rho^s)\,\mu + \rho^s r_t]/n$. Notice that $\Sigma_{s=0}^{n-1}\rho^s = (1-\rho^n)/(1-\rho)$, so (18.6) follows directly. □

18.3 Hedging a Bond using the Vasicek Model

The Vasicek model allows us to apply a potentially better way of *hedging a bond portfolio* than the traditional duration hedging. The model can account for both level and slope changes of the yield curve, while the duration hedging was based on the assumption of only level (parallel) shifts.

Recall that we have a liability worth P_L, and we buy v units of a bond portfolio (denoted H) with price P_H. The value of the overall position is

$$V = vP_H + M - P_L, \tag{18.8}$$

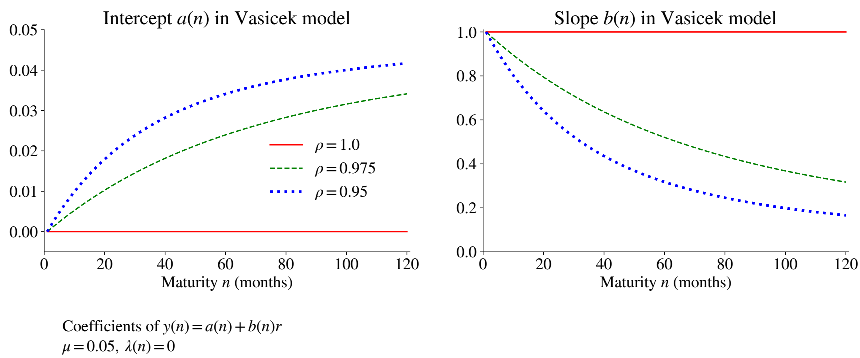


Figure 18.6: Intercept and slope in the Vasicek model

where M is a short-term money market account.

The change of the hedge portfolio (over a short time interval) is

$$\Delta V = v\Delta P_H - \Delta P_L, \tag{18.9}$$

and a bond price can be calculated as

$$P = \textstyle\sum_{k=1}^{K} cf_k e^{-m_k y(m_k)}, \tag{18.10}$$

where cf_k is the cash flow at time m_k, $y(m_k)$ is the continuously compounded interest between now and m_k years later. Notice that time (m_k) is here measured in *years* since the interest rates $y(m_k)$ are annualized rates.

Once we know the parameters of the Vasicek model, it is straightforward to numerically calculate what ΔP_H and ΔP_L are, as functions of the change in the current short rate interest rate (Δr_t). We then set v so that $\Delta V = 0$ in (18.9), that is,

$$v = \Delta P_L / \Delta P_H. \tag{18.11}$$

This can be done by either using spot rate durations or a numerical approach.

For a *spot rate duration* approach, notice that a Taylor approximation of (18.10) in terms of all (continuously compounded) spot rates gives

$$\Delta P/P = -\textstyle\sum_{k=1}^{K} D_k \times \Delta y(m_k), \text{ with } D_k = m_k cf_k e^{-m_k y(m_k)}/P, \tag{18.12}$$

and where D_k are the spot or rate durations (as many as there are cash flows). (By definition, $D_k = [-\partial P/\partial y(m_k)]/P$.) When $\Delta y(m_k)$ is the same for all k, then this

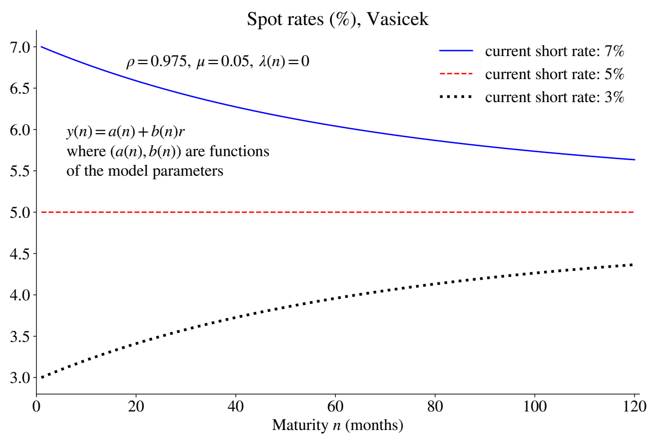


Figure 18.7: Vasicek model, spot rates for different initial short rates

is the same as modified duration approach. Otherwise, this can be combined with a model which describes all $\Delta y(m_k)$, often in terms of a low number of factors.

Remark 18.10 *(Spot rate duration, effective rates*) Recall that the bond price is* $P = \Sigma_{k=1}^{K} cf_k \left[1 + Y(m_k)\right]^{-m_k}$*, where* $Y(m_k)$ *is the effective spot rate. A Taylor approximation in terms of all (effective) spot rates gives*

$$\Delta P/P = -\Sigma_{k=1}^{K} D_k \times \Delta Y(m_k), \text{ with } D_k = m_k cf_k \left[1 + Y(m_k)\right]^{-m_k-1}/P.$$

The Vasicek model has a single factor: the short rate (r). From (18.6) we have $\Delta y(m_k) = b(m_k)\Delta r$, so combining with (18.12) gives

$$\Delta P = -P[\textstyle\sum_{k=1}^{K} D_k b(m_k)]\Delta r. \tag{18.13}$$

Apply this to both the liability and the hedge bond and then calculate the hedge ratio as in (18.11).

For a more direct *numerical approach*, see Figure 18.8 and the details in Remark 18.11.

Remark 18.11 *(Numerical calculation of* ΔP*) (1) For an initial value of the short interest rate* r*, use (18.6) to calculate all spot rates* $y(m_k)$ *needed in (18.10). (2) Redo, but starting from another short rate, say, the earlier* r *plus 1%.* (3) *Calculate* ΔP.

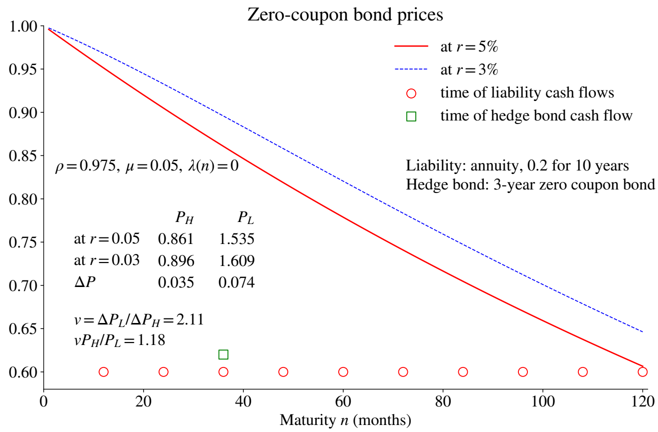


Figure 18.8: Bond prices in the Vasicek model

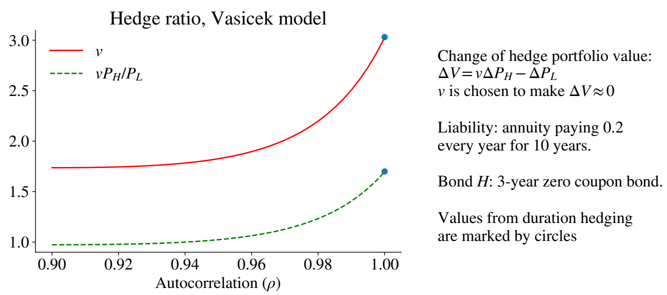


Figure 18.9: Hedge ratios in the Vasicek model

Figure 18.9 gives an illustration, for different values of the autocorrelation ρ in the short rate process (18.3). The hedge ratio v converges to the duration hedge ratio as the autocorrelation (ρ) increases towards unity: in that limiting case all yield curve movements are indeed parallel. For lower values of the autocorrelation, the hedge ratio is lower. The main reason is that a low autocorrelation makes interest rates on long-maturity bonds (here, the liability) move less than interest rates on short-maturity bonds (here, the hedge bonds), see Figure 18.7. This dampens the movements in the bond prices, as illustrated in Figure 18.10.

Notice, however, that all one-factor models (including the Vasicek model) imply that all yields are perfectly correlated (there is a common single driving force) and only fairly limited yield curve movements are possible. For instance, if the current

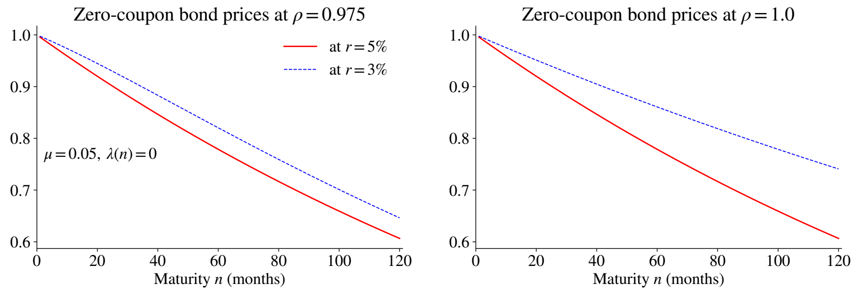


Figure 18.10: Bond price changes in the Vasicek model

short rate is low, then the yield curve must be upward-sloping. *Multi-factor models* overcome most of those limitations, for instance, the two-factor Nelson and Siegel (1987) model.

18.4 Interest Rates and Macroeconomics*

This section outlines several (not mutually exclusive) macroeconomic approaches to modelling the yield curve.

18.4.1 The Fisher Equation and Index-Linked Bonds

Let π_{t+n} be the annualised inflation rate over t to $t+n$, and $y_t^r(n)$ the *real interest rate* for the same period. The real interest rate is a return in terms of real purchasing power. Note the difference between a real interest rate and a traditional interest rate, where the latter (also called a nominal interest rate) is in terms of monetary units (dollars, say).

The *Fisher equation* says that the nominal interest rate includes compensation both for inflation expectations, $\mathrm{E}_t\,\pi_{t+n}$, the real interest rate, $y_t^r(n)$, and possibly a constant risk premium, $\psi(n)$,

$$y_t(n) = \mathrm{E}_t\,\pi_{t+n} + y_t^r(n) + \psi(n). \tag{18.14}$$

Example 18.12 *(Fisher equation) Suppose the nominal interest rate is* $y(n) = 0.07$*, the real interest rate is* $y^r(n) = 0.03$*, and the nominal bond has no risk premium* ($\psi = 0$)*, then the expected inflation is* $\mathrm{E}_t\,\pi_{t+n} = 0.04$.

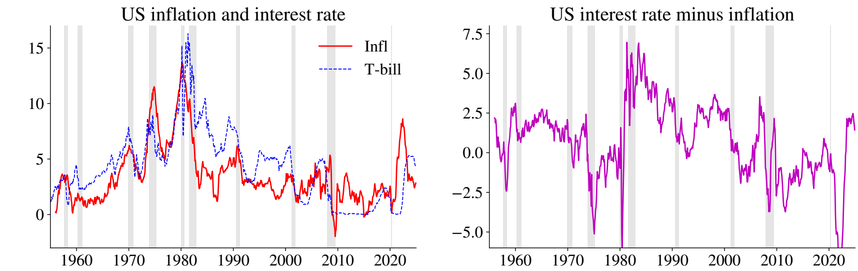


Figure 18.11: US inflation and 3-month interest rate

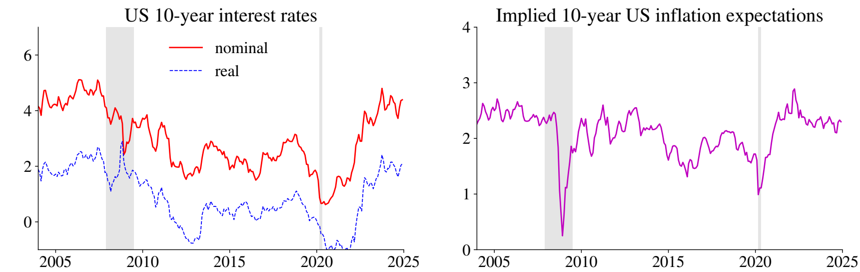


Figure 18.12: US nominal and real interest rates

The Fisher equation suggests a framework for analysing nominal interest rates in terms of real interest rates and inflation expectations. Information about real interest rates could possibly be elicited from *index-linked bonds*, that is, bonds which give automatic compensation for actual inflation.

Empirical results typically indicate non-trivial fluctuations in the real interest rate and risk premia (possibly driven by liquidity risk). This holds also when inflation expectations as measured by surveys. It is therefore not straightforward to extract inflation expectations from nominal interest rates.

Empirical Example 18.13 *Figures 18.11–18.13 illustrate the relation between U.S. nominal and real interest rates, as well as inflation. A potential conclusion is that there are considerable movements in real interest rates (and/or liquidity premia on index linked bonds).*

The Fisher equation is sometimes embedded in a macro model to construct a sophisticated model of the yield curve. This involves using macro theory/empirics

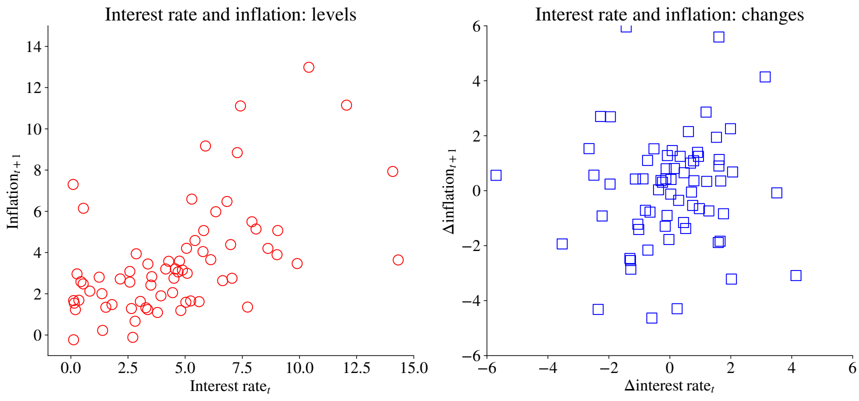


Figure 18.13: US nominal interest rates and subsequent inflation

to model how real interest rates and inflation expectations depend on the state of the economy.

18.4.2 The Expectations Hypothesis of Interest Rates

The expectations hypothesis of interest rates says that long interest rates equal an average of expected future short rates, possibly with a constant (across time, not maturities) risk premium as in (18.2). This can help interpreting yield curve changes around, for instance, interest rate hikes by a central bank. Suppose the central bank increases its policy rate, a short-maturity rate. The impact on longer rates depends of several factors.

First, one possibility is that only the very short interest rates change, and that all longer interest rates stay unchanged. This would happen if the policy move was well anticipated.

Second, another possibility is that long interest rates increase. Under the expectations hypothesis of interest rates, the interpretation is that the market now expects high short interest rates also in the future. If we are willing to assume that the real interest rate was not affected by the policy move, then one possible interpretation is that the central bank has received information about a long-lasting inflation pressure.

Third, and finally, short rates may increase, but long interest rates decrease. A common interpretation of this scenario is that the central bank has become more

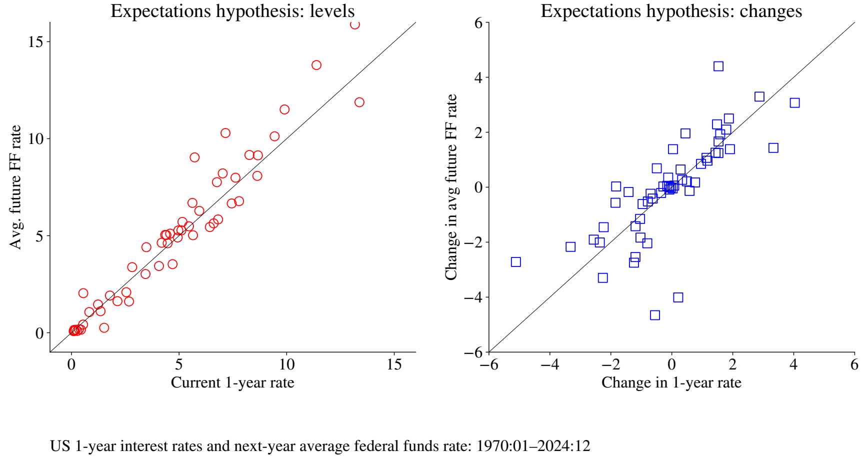


Figure 18.14: US 12-month interest and average federal funds rate (next 12 months)

inflation averse. It therefore raises the policy rate to bring down inflation. If the market believes that it will succeed, then it follows that it will eventually be possible to lower interest rates (when inflation and inflation expectations are lower).

The expectations hypothesis has been tested many times, typically by an ex post linear regression (realized interest rates regressed on lagged forward rates). The results often give mild support to the hypothesis.

Empirical Example 18.14 *Figure 18.14 shows scatter plots of long interest rates and average future short rates—in levels and in changes. The evidence suggests some mild support of the expectations hypothesis.*

18.4.3 A New-Keynesian Model of Monetary Policy

Monetary policy is a crucial part of the macroeconomic setting, so it is important to understand how the policy is formed. It has not always been this way: there are long periods when many countries adopted a very simple (or so it seemed) monetary policy by pegging the currency to another currency. Macroeconomic policy was then synonymous with fiscal policy.

Modern macro models are often smaller than the older macroeconometric models and they pay more attention to theory, the supply side of the economy and the role of expectations. These models try to capture the key elements in the way central banks (and most other observers) reason about the interaction between inflation, output, and monetary policy.

In these models, inflation depends on expected future inflation (some prices are set today for a long period and will therefore be affected by expectations about future costs and competitors' prices), lagged inflation, and a "Phillips effect" where an *output gap* (output less trend output) affects price setting via demand pressure. For instance, inflation (π_t) is often modelled as

$$\pi_t = \alpha \operatorname{E}_t \pi_{t+1} + \beta \pi_{t-1} + \phi x_t + \varepsilon_{\pi t}, \tag{18.15}$$

where x_t is the output gap and $\varepsilon_{\pi t}$ can be interpreted as "cost push" shocks (wage demands, commodity price shocks). This equation can be said to represent the supply side of the economy and it is typically derived from a model where firms with some market power want to equate marginal revenues and marginal costs, but choose to change prices only gradually.

The demand side of the economy is modelled from consumers' savings decision, where the trade-off between consumption today and tomorrow depends on the real interest rates. Simplifying by setting consumption equal to output we get something like the following equation for the output gap

$$x_t = x_{t-1} - \gamma (i_t - \operatorname{E}_t \pi_{t+1}) + u_t, \tag{18.16}$$

where i_t is the nominal interest rate (set by the central bank) and u_t is a shock to demand. Note that the expected *real* interest rate affects demand (negatively).

In some cases, the real exchange rate is added to both (18.15) and (18.16), capturing price increases on imported goods and foreign demand for exports, respectively. The exchange rate is then often linked to the rest of the model via an assumption of uncovered interest rate parity.

Some of the important features of this simple model are: *(i)* inflation expectations matter for today's inflation (think about wage inflation), *(ii)* the instrument for monetary policy, the short interest rate i_t, can ultimately affect inflation only via the output gap; *(iii)* it is the real, not the nominal, interest rate that matters for demand.

To make the model operational, two more things must be added: the monetary policy rule and a formalization of how expectations in (18.15)–(18.16) are formed.

It is common to assume that the central bank has some instrument rule like the "Taylor rule"

$$i_t = \theta_0 + 0.5 x_t + 1.5 \pi_t + v_t. \tag{18.17}$$

The residual v_t is a "monetary policy shock," which picks up factors left out of the model, for instance, the central bank's concern for the banking sector or simply changes in the central bank's preferences.

Another approach to find a policy rule is to assume that the central bank has some loss function that it minimizes by choosing a policy rule. This loss function is often a weighted average of the variance of inflation and the variance of the output gap.

The expectations in (18.15)–(18.16) can be handled in many ways. The perhaps most straightforward way is to assume that the expectations about the future equal the current value of the same variable (a "random walk"). A more satisfactory way is to use survey data on inflation expectations. Finally, many model builders assume that expectations are "rational" (or "model consistent") in the sense that the expectation equals the best guess we could do under the assumption that the model is correct. This latter approach typically requires a sophisticated way of solving the model, as the model both generates the best guesses and depends on them.

18.5 Risk Premia on Fixed Income Markets*

There are many different types of risk premia on fixed income markets.

Nominal bonds are risky in real terms, and are therefore likely to carry *inflation risk premia*. Long bonds are risky because their market values fluctuate over time, so they probably have *term premia*. Corporate bonds and some government bonds (in particular, from developing countries) have *default risk premia*, depending on the risk for default. Interbank rates may be higher than T-bill of the same maturity for the same reason (see the TED spread, the spread between 3-month Libor and T-bill rates) and illiquid bonds may carry *liquidity premia* (see the spread between off-the run and on-the-run bonds).

Empirical Example 18.15 *Figures 18.15–18.17 illustrate some U.S. data. In particular, there seems to be considerable (and business cycle related) default risk premia in the corporate sector, and also within the banking sector. in addition, the evidence on the on/off-the run interest rates suggests important liquidity risk premia, even across comparable bonds with the same issuer (the U.S. Treasury).*

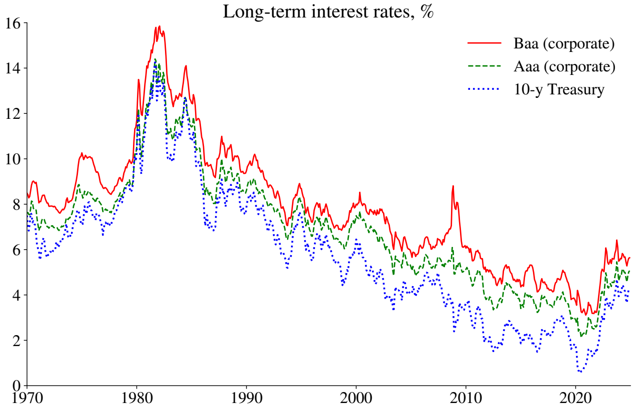


Figure 18.15: US interest rates

18.6 Appendix – Formal Derivation of the Vasicek Model*

Remark 18.16 *This section uses a slightly different notation, namely a subscript$_n$ to indicate the maturity and P to indicate a zero coupon bond price. For instance, y_{nt} for the n-period interest rate in t (same as $y_t(n)$ in the rest of this chapter) and P_{nt} and $P_{n-1,t+1}$ to indicate a bond price.*

Write (18.6) as

$$y_{nt} = a_n + b_n r_t, \text{ where } a_n = A_n/n \text{ and } b_n = B_n/n. \tag{18.18}$$

The expressions for A_n and B_n will be derived below.

The price of an n-period zero coupon bond equals the cross-moment between the stochastic discount factor (SDF) and the value of the same bond next period (when it's an $n-1$-period bond)

$$P_{nt} = \mathrm{E}_t\, e^{m_{t+1}} P_{n-1,t+1}, \tag{18.19}$$

where m_{t+1} is the *logarithm* of the stochastic discount factor $e^{m_{t+1}}$. Notice that this notation differs from some other chapters.

The *Vasicek model* assumes that the log SDF (m_{t+1}) is a linear function of r_t

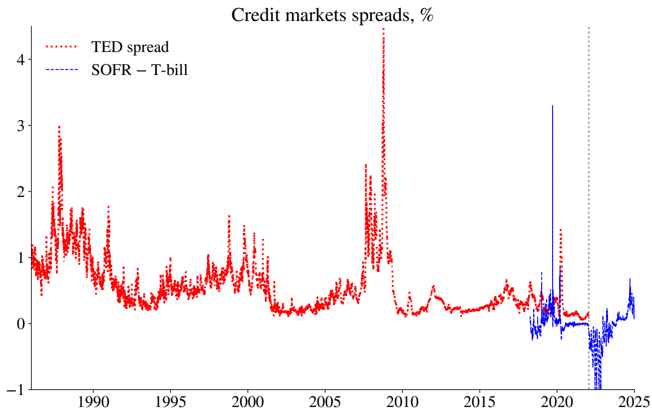


Figure 18.16: TED spread

and an iid shock

$$-m_{t+1} = r_t + \gamma\varepsilon_{t+1}, \text{ where } \varepsilon_{t+1} \text{ is iid } N(0,\sigma^2) \text{ and} \tag{18.20}$$

$$r_{t+1} = (1-\rho)\,\mu + \rho r_t + \varepsilon_{t+1}. \tag{18.21}$$

The short rate process is the same as in (18.3).

Remark 18.17 *If* $x \sim N(\mu,\sigma^2)$ *, then* $\mathrm{E}\, e^x = e^{\mu+\sigma^2/2}$*. Take logs to get* $\ln \mathrm{E}\, e^x = \mu + \sigma^2/2$.

The model values of (A_n, B_n) are found by using *(a)* $P_n = e^{-ny_n}$; *(b)* the proposed model (18.18); *(c)* the dynamics in (18.20)–(18.21) to calculate the logarithm of (18.19) as

$$p_{nt} = \mathrm{E}_t(m_{t+1} + p_{n-1,t+1}) + \mathrm{Var}_t(m_{t+1} + p_{n-1,t+1})/2, \tag{18.22}$$

where p_{nt} is the log bond price ($\ln P_{nt}$). This is an application of Remark 18.17 with $m+p$ playing the role of x. The result (see below for a proof) is that

$$B_n = 1 + \rho B_{n-1} \text{ and} \tag{18.23}$$

$$A_n = A_{n-1} + B_{n-1}\,(1-\rho)\,\mu - (\gamma + B_{n-1})^2\,\sigma^2/2, \tag{18.24}$$

where the recursion starts at $B_0 = 0$ and $A_0 = 0$. Notice that the expression for B_n is the same as in Remark 18.8, but that we have another expression for the A_n which

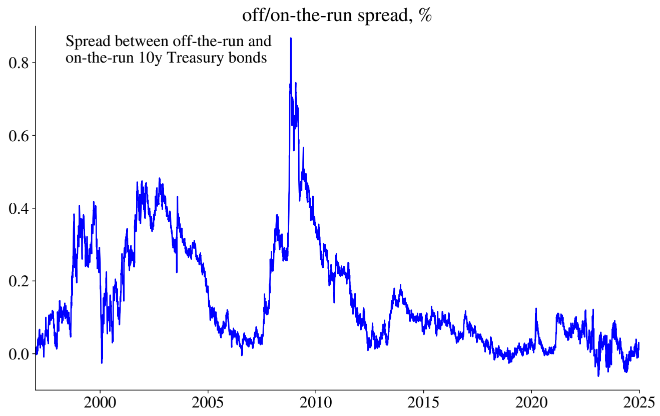


Figure 18.17: Off-the-run liquidity premium

involves both the mean μ and the volatility (risk) σ^2.

Example 18.18 *(A_n and B_n in the Vasicek model) (18.23 –(18.24) give*

$$B_0 = 0 \text{ and } A_0 = 0$$
$$B_1 = 1 \text{ and } A_1 = -\gamma^2\sigma^2/2$$
$$B_2 = 1 + \rho \text{ and } A_2 = (1-\rho)\,\mu - \left[\gamma^2 + (1+\gamma)^2\right]\sigma^2/2.$$

Proof (of (18.23)–(18.24)) First, rewrite

$$m_{t+1} + p_{n-1,t+1} = \underbrace{-r_t - \gamma\varepsilon_{t+1}}_{m_{t+1}}\underbrace{-A_{n-1} - B_{n-1}r_{t+1}}_{p_{n-1,t+1}}$$
$$= -(1 + B_{n-1}\rho)\,r_t - (\gamma + B_{n-1})\,\varepsilon_{t+1} - A_{n-1} - B_{n-1}(1-\rho)\,\mu,$$

where we use (18.21) to substitute for r_{t+1}. The conditional moments in (18.22) can then be calculated as

$$\mathrm{E}_t\,(m_{t+1} + p_{n-1,t+1}) = -(1 + B_{n-1}\rho)\,r_t - A_{n-1} - B_{n-1}(1-\rho)\,\mu$$
$$\mathrm{Var}_t\,(m_{t+1} + p_{n-1,t+1}) = (\gamma + B_{n-1})^2\,\sigma^2.$$

Second, substitute $p_{nt} = -A_n - B_n r_t$ on the LHS of (18.22) and plug in the conditional moments from above on the RHS

$$-A_n - B_n r_t = -(1 + B_{n-1}\rho)\,r_t - A_{n-1} - B_{n-1}(1-\rho)\,\mu + (\gamma + B_{n-1})^2\,\sigma^2/2.$$

This equation must always hold (for any value of r_t: match coefficients of r_t and the "constant" to get (18.23)–(18.24). □

Chapter 19

Basic Properties of Options

This chapter presents different types of options (calls/puts, European/American), and the payoff from option portfolios, for instance, straddles and protective puts. It also discusses the put-call partity and as well as pricing bounds for options. In general, this provides an overview of basic properties of options. However, a more detailed discussion of option pricing is postponed to later chapters.

19.1 Derivatives

Derivatives are assets whose payoffs depend on an underlying asset (for instance, shares of a company). The most common derivatives are futures contracts (or similarly, forward contracts) and options. Options are sometimes written on (depend on) the price of a futures contract, not the underlying directly. See Figure 19.1.

Derivatives have zero net supply, so a contract must be issued (a short position) by someone for an investor to be able to buy it (a long position). For that reason, gains and losses on derivatives markets sum to zero.

19.2 Introduction to Options

Remark 19.1 *(On the notation) The notation here is kept short. The current period is assumed to be* 0 *and the derivative expires m years later. The current price of the underlying is denoted S (rather than S_0), the forward price according to a contract agreed on now and expiring in m is F (rather than $F_0(m)$) and the continuously compounded interest between* 0 *and m is y (rather than $y_0(m)$). However, to avoid confusion, the price of the underlying asset at expiration is denoted S_m. The more*

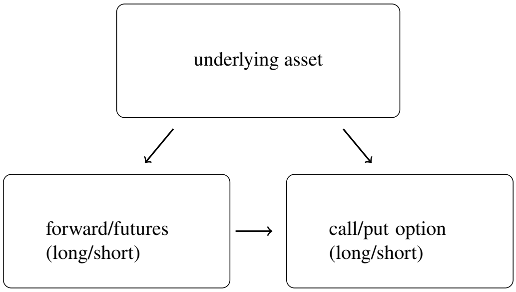


Figure 19.1: Derivatives on an underlying asset

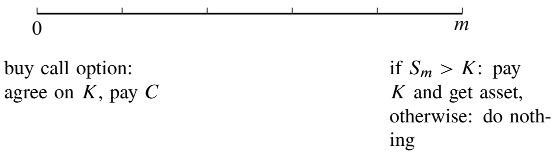


Figure 19.2: Timing convention of a European call option contract

precise notation is used only when needed in a particular context.

19.2.1 Definition of European Calls and Puts

A European *call* option contract traded today (period 0) stipulates that the owner of the contract has the *right* (but not the obligation) to *buy* one unit (a normalization) of the underlying asset ("exercise the option") from the issuer of the option on the expiration date m at the strike price K. This is different from a forward contract where the owner *must* exercise. See Figure 19.2 for the timing convention.

The analysis here normalizes all contracts to one unit of the underlying. A simple rescaling is needed for an application to typical contracts, which may be for many more units.

To the owner of a call option, the payoff at expiration is either zero (if the owner does not exercise) or the value the underlying asset S_m minus the strike price K (if the owner exercises). For a rational investor (who only exercises if $S_m \geq K$), the *payoff* is thus

$$\text{call payoff}_m = \max\left(0, S_m - K\right). \tag{19.1}$$

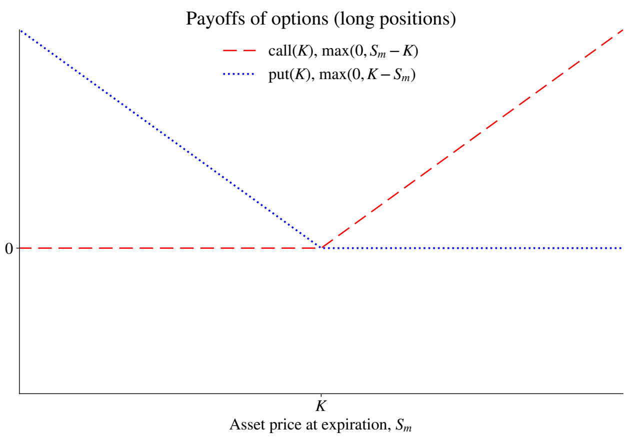


Figure 19.3: Payoffs of options, long positions

Clearly, this is (weakly) increasing in the price of the underlying asset.

Example 19.2 *(Call option payoffs) With $K = 5$ we have*

S_m	*Exercise*	*Payoff*
4.5	*no*	0
5.5	*yes*	$5.5 - 5 = 0.5$

The *profit at expiration* is thus

$$\text{call profit}_m = \text{call payoff}_m - e^{my}C, \tag{19.2}$$

where C is the call price, typically on the trade day. (To simplify the notation, the time subscript on C is suppressed, but we could write C_0 when required.) The e^{my} factor captures the capital cost of paying the option price already on the trade date (think: borrow C in period 0 and repay with interest, $e^{my}C$, on the expiration date to calculate the final profit). Time to expiration m is measured in years, since interest rates are annualized rates. See Figure 19.4 for an illustration. Notice that the price of the option (C) is always paid, irrespective of whether the option is later exercised or not.

Remark 19.3 *(In-the-money*) An option that would be profitable to exercise is called* in-the-money*; an option that would be unprofitable to exercise is called* out-of-the-money*—and an option that would just break even is called at-the-money.*

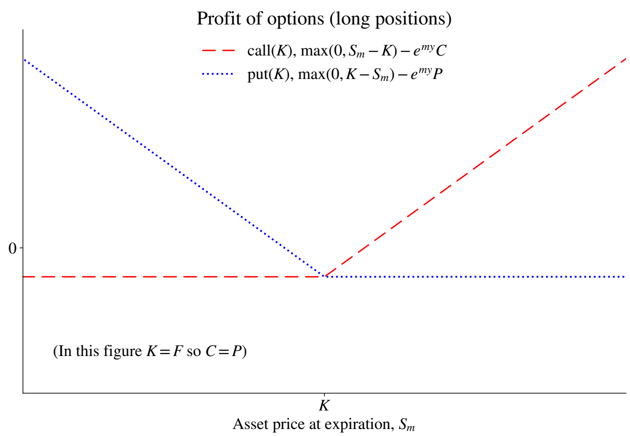


Figure 19.4: Profit of options, long positions

The payoff (profit) of the issuer is the mirror image of the owner's payoff (profit): the owner's gain is the issuer's loss: a *zero sum game*. See Figure 19.5 for an illustration.

Remark 19.4 *(Margin requirements*) On an options exchange, a buyer of an option does not have to post any margin, but an issuer typically does. The reason is that a default of the issuer could create a loss for the option owner (if the option is worth exercising). In contrast, a default of the owner cannot create a loss for the issuer. On the OTC market, collateral is more common.*

A *put* option instead gives the owner of the contract the right to *sell* one unit of the underlying asset at the strike price K. The put price is here denoted by P. An owner of a put option benefits from a low price of the underlying asset (buy the asset cheaply and exercise the right to sell it for K). The payoff is

$$\text{put payoff}_m = \max(0, K - S_m). \qquad (19.3)$$

See Figures 19.3 and 19.4 for the payoff and profit of a long position, and Figure 19.5 for the profit of a short position.

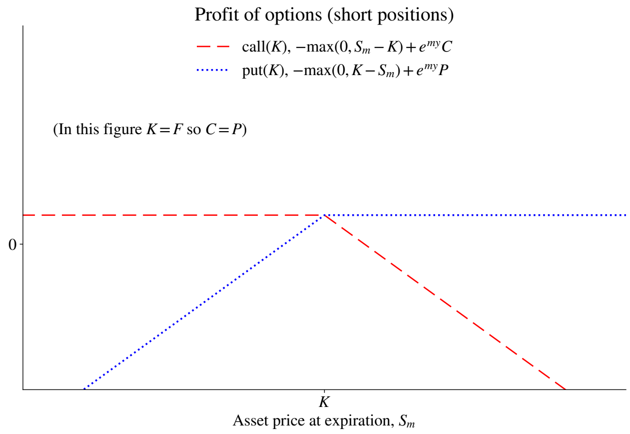


Figure 19.5: Profit of options, short positions

Example 19.5 *(Put option payoffs) With* $K = 5$ *we have*

S_m	*Exercise*	*Payoff*
4.5	*yes*	$5 - 4.5 = 0.5$
5.5	*no*	0

Remark 19.6 *(Which options are traded?) Most of the trade is in out-of-the-money options (high strike prices for the calls and low strike prices for the puts). Also, most of the trade happens close to the expiration date, and there is a seasonality pattern related to rolling over the investment from other (expired) options. Figure 19.6 shows how the trading volume at CBOE has developed over time. The volume seems to correlate with the general business cycle movements. The ratio of traded put contracts to traded call contracts in Figure 19.6 is sometimes used to gauge market nervousness. The idea is that investors will demand put contracts if they want to insure against a stock market decline.*

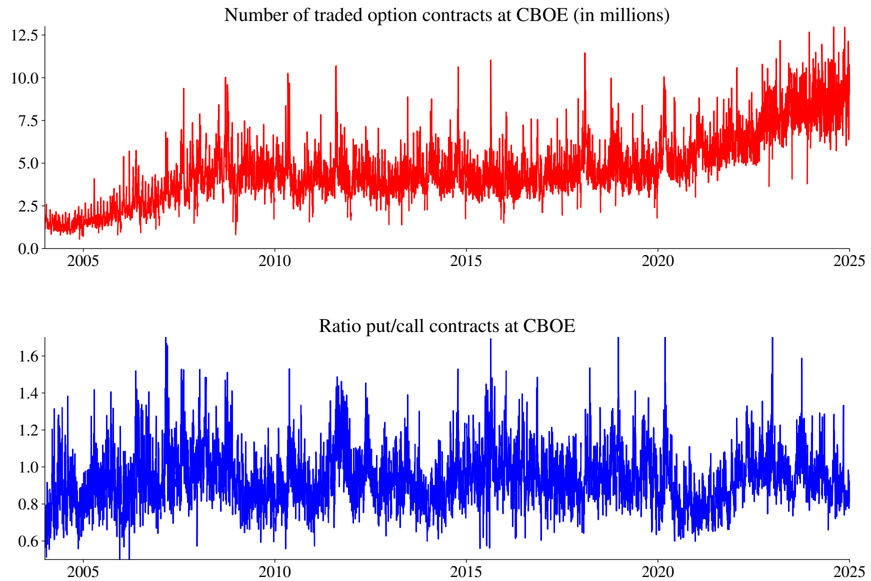


Figure 19.6: Option trade volume

19.2.2 Options Are Risky Assets

The net return on a long position in a European call option is

$$\text{return on call}_m = \frac{\max(0, S_m - K)}{C} - 1, \tag{19.4}$$

where C is the call option price. Whenever the option is not exercised ($S_m < K$), the whole investment is lost (and the return is -100%). In contrast, when the option is exercised ($S_m > K$), then the return can potentially become very large.

It is clear that the option return (19.4) cannot be normally (or even lognormally) distributed: the density function has a spike at -100% (whose probability mass is the same as the probability of $S_m \leq K$). This means that we cannot motivate "mean-variance" pricing of options by referring to a normal distribution of the return. (This does not rule out mean-variance pricing, which could be motivated by, for instance, mean-variance preferences.)

19.3 Financial Engineering

This section discusses the properties of some specific portfolios of options, forwards and the underlying asset.

19.3.1 Replicating a Forward

Options markets are often very liquid—and are therefore useful for constructing replicating portfolios. Consider a portfolio which is long one call option with strike price K and short one put option with the same strike price. When $K = F$, then this portfolio replicates a forward contract, so it is a synthetic forward. (The proof is at the end of the section.) See Figure 19.7, which plots profit functions. Clearly, we can then replicate a short position in a forward contract by selling such a portfolio. To see this, sum up the two positions for any outcome of S_m, as in the next example.

Example 19.7 *(Payoff of a synthetic forward) With $K = 5$, we have the differences of the payoffs in Examples 19.2 and 19.5, that is,*

S_m	*Ex. call*	*Call payoff*	*Ex. put*	*Put payoff, short*	*Total Payoff*
4.5	*no*	0	*yes*	$-(5-4.5)$	-0.5
5.5	*yes*	$5.5-5$	*no*	0	0.5

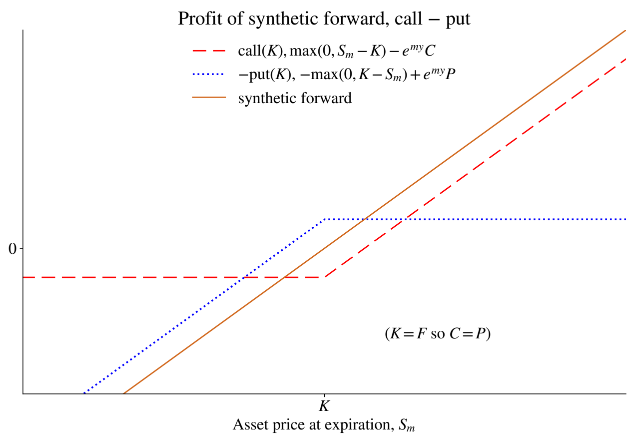


Figure 19.7: Profit of an option portfolio that replicates a forward contract

The payoff is indeed linear in S_m, similar to the forward ($S_m - F$). To get the profit, subtract the capitalized difference of the call and put prices from the total payoff, which is zero if $K = F$.

19.3.2 Portfolio Insurance

A *protective put* is a combination of a put and a position in the underlying asset. The latter has the profit $S_m - e^{ym}S$, where the second term is the capitalized value of the purchase price. This allows the owner to capture the upside of the price movement (of the underlying), at the same time as insuring against the downside. This is very similar to just buying a call option. See Figure 19.8 (and below for a proof/details).

Example 19.8 *(Payoff of portfolio insurance) With $K = 5$, we have the sum of the payoffs in Examples 19.2 and 19.5, that is,*

S_m	*Payoff from underlying*	*Ex. put*	*Payoff from put*	*Total Payoff*
4.5	4.5	*yes*	$5 - 4.5$	5
5.5	5.5	*no*	0	5.5

Suppose $e^{ym}S = 5$, then the profits are $0 - e^{ym}P$ and $0.5 - e^{ym}P$, which is the same pattern as for the call option in Example 19.2.

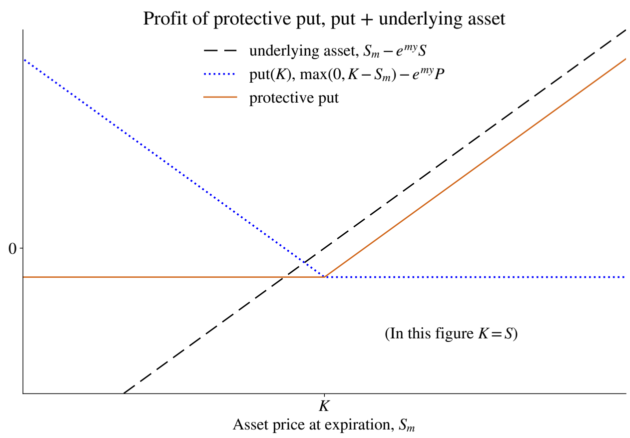


Figure 19.8: Profit of an option portfolio that insures the underlying asset

19.3.3 Betting on Large Changes (in Any Direction)

An option is a bet on a change in a specific direction, but option portfolios can be constructed to instead make a bet on a large change in either direction (that is, high volatility). A *straddle* is long one call and one put with the same strike price. It pays off whenever the underlying moves away from the strike price, irrespective of direction. See Figure 19.9 (and also below a proof/details).

Example 19.9 *(Payoff of a straddle) With $K = 5$, we have the sum of the payoffs in Examples 19.2 and 19.5, that is,*

S_m	*Ex. call*	*Payoff from call*	*Ex. put*	*Payoff from put*	*Total Payoff*
4.5	*no*	0	*yes*	$5 - 4.5$	0.5
5.5	*yes*	$5.5 - 5$	*no*	0	0.5

To get the profit, subtract the capitalized sum of the call and put prices.

A *strangle* is similar, but the put has a lower strike price (K_1) than the call (K_2). The profit figure is similar to that of a straddle, except that it has a flat portion between the two strike prices, see Figure 19.10.

Proof (Profits of selected option portfolios) These proofs present results for European style options for which the put-call parity in Proposition 19.10 hold. (a) The synthetic forward has the payoff/profit $\max(0, S_m - F) - \max(0, F - S_m)$ since

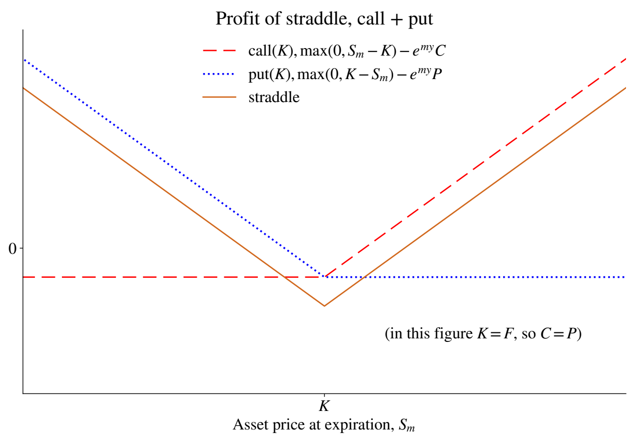


Figure 19.9: Profit of a straddle

$K = F$ and the two option prices cancel due to the put-call parity. This directly gives $S_m - F$. (b) The protective put has the profit $S_m + \max(0, K - S_m) - e^{ym}(P + S)$, which equals $\max(0, S_m - K) + K - e^{ym}(P + S)$. Assuming no dividends (so $F = Se^{ym}$), the put-call parity gives $e^{my}(P + S) = e^{my}C + K$. Combine to get the call profit, $\max(0, S_m - K) - e^{ym}C$. (c) The straddle directly gives $|S - K| - e^{my}(C + P)$. $\square$

19.3.4 Betting on a Large Price Decrease

A variation on a short synthetic short forward is the *collar*: long a put with a low strike price (K_1) and short a call with a high strike price (K_2). This looks like a short position in a forward contract, except that the payoff is flat between the strike prices. Clearly, this is betting on a large price decrease. See Figure 19.10. Selling a collar (also called *reversal*) is instead a bet on a large price increase.

19.3.5 Betting On a Small Price Increase

To bet on a small increase in the price of the underlying asset we can use a *bull spread*: long a call option with a low strike price (K_1) and short a call option with a high strike price (K_2). This portfolio has flat payoffs outside the strike prices, but a payoff that increases with the underlying asset between them. See Figure 19.10. Selling a bull spread creates a *bear spread*, which is a bet on a small decrease of the

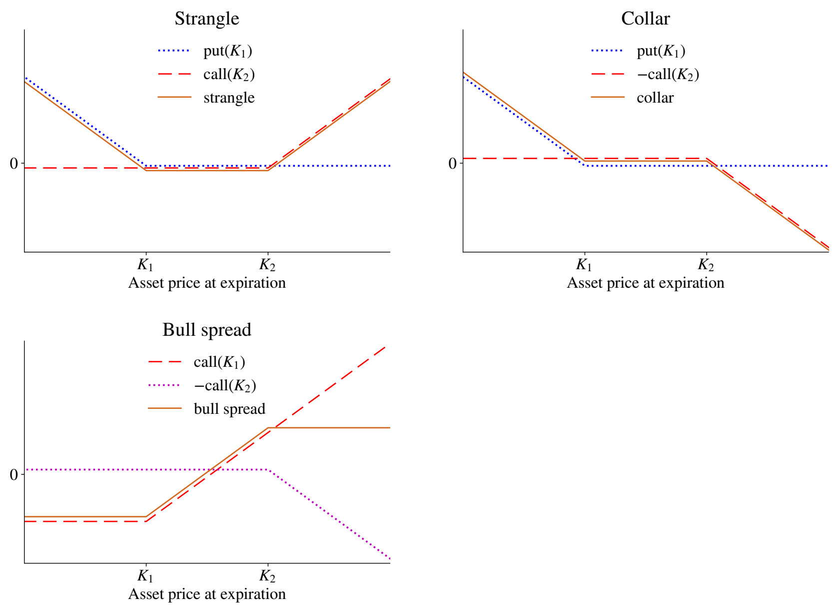


Figure 19.10: Profits of option portfolios

underlying price. (These spreads can also be constructed by combing puts.)

See, for instance, Sundaram and Das (2015) 8 for further strategies and discussion.

19.4 Prices of Options

Much of the subsequent analysis will focus on understanding how options are priced (before the expiration date), that is, how the C and P are determined.

As an example, Figure 19.11 shows results from a particular model for option pricing (the Black-Scholes model) to illustrate some properties that are actually shared among many option pricing models. The general pattern is that a call option is increasing in today's underlying price and decreasing in the strike price—and that these effects are smoother when the time to expiration is large. Also, the option price is increasing in the volatility of the underlying asset.

However, before creating such pricing models, we will first discuss *(a)* how put and call prices are related; *(b)* the effects of the strike price K and volatility σ; *(c)* and also derive (no-arbitrage) bounds that option prices have to obey.

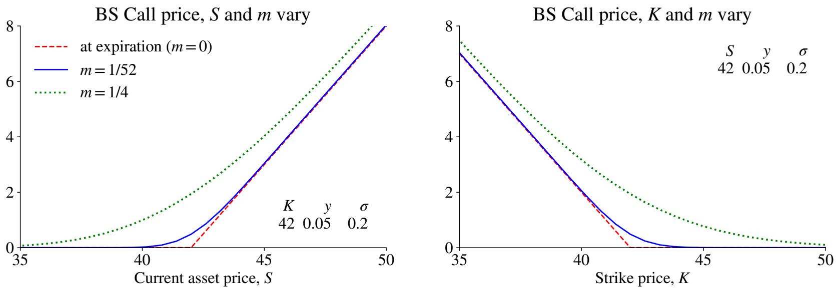


Figure 19.11: Call option prices (from the Black-Scholes model)

19.5 Put-Call Parity for European Options

There is a tight link between European call and put prices. If you know one of them (and the forward price), then you can easily calculate what the other must be. The following proposition is more precise.

Proposition 19.10 *(Put-call parity for European options) The put-call parity for European options is*

$$C - P = e^{-my}(F - K), \tag{19.5}$$

where $e^{-my}(F - K)$ is the present value of the forward price minus the strike price.

The parity holds irrespective of whether the underlying asset has dividends or not (since the expression uses the forward price).

Example 19.11 *(Put-call parity) Let $S = 42, m = 1/2, y = 5\%, K = 38$. If the underlying asset has no dividends, then $F = e^{0.5\times0.05}42 = 43.06$. With $C = 5.5$, (19.5) gives*

$$5.5 - P = e^{-0.5\times0.05}(43.06 - 38) \text{ or } P \approx 0.56.$$

Proof (of Proposition 19.10) Portfolio A: buy one call option and sell one put option, both with the strike price K, at the cost $C - P$. This will with certainty give $S_m - K$ at maturity (since the call *or* the put will be exercised). Portfolio B: enter a forward contract and put $e^{-my}(F - K)$ in the bank (your cost). At expiration, get $S_m - F$ from the forward contract plus the $F - K$ that you have in the bank: $S_m - K$. Since the two portfolios give the same at expiration, they must have the same costs today. See Figure 19.12 for an illustration. □

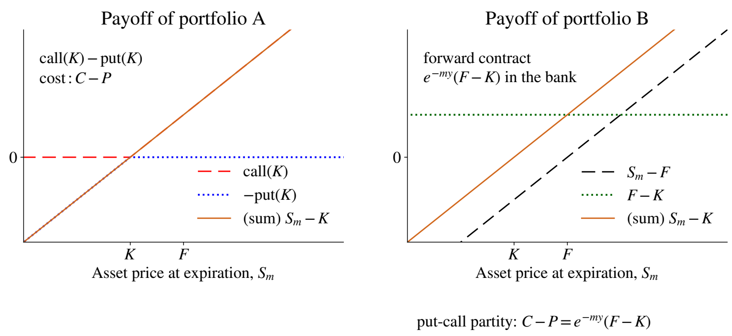


Figure 19.12: Put-call parity

Example 19.12 *(Trading on deviations from the put-call parity) Assume the same numbers as in Example 19.11, except that* $P = 1$. *Buying a call, selling a put and issuing a forward then costs* $C - P = 4.5$ *in* $t = 0$. *To finance this, we borrow and pay back* $e^{0.5\times0.05}4.5 = 4.61$ *at expiration. The options and forwards together give* $F - K = 43.06 - 38 = 5.06$ *for sure at expiration. There is thus a risk-free profit. (With* $P = 0.56$ *there is not.)*

A few special cases of (19.5) are of particular interest. First, when the underlying asset pays no dividends, then (19.5) together with the forward-spot parity ($F = e^{my}S$) give

$$C - P = S - e^{-my}K \text{ if no dividends.} \tag{19.6}$$

Second, with dividends (continuous or at discrete times) we get

$$C - P = Se^{-m\delta} - e^{-my}K \text{ if continuous dividend rate } \delta, \tag{19.7}$$

$$C - P = S - \textstyle\sum_{i=1}^{n} e^{-m_i y(m_i)} D_i - e^{-my}K \text{ if discrete dividends.} \tag{19.8}$$

See Hull (2022) 11 and McDonald (2014) 11–12 for more detailed treatments.

19.5.1 Put-Call Parity and Synthetic Replications*

The background to the put-call parity is that we use two assets to replicate a third. This can be used in different ways. For instance, we can combine a call option (with strike price $K = F$) and a forward contract to replicate a put option, or buy a call and sell a put (with strike price $K = F$) to replicate a forward contract. (Transaction costs can, of course, cause some deviations from the parity conditions). See Figure

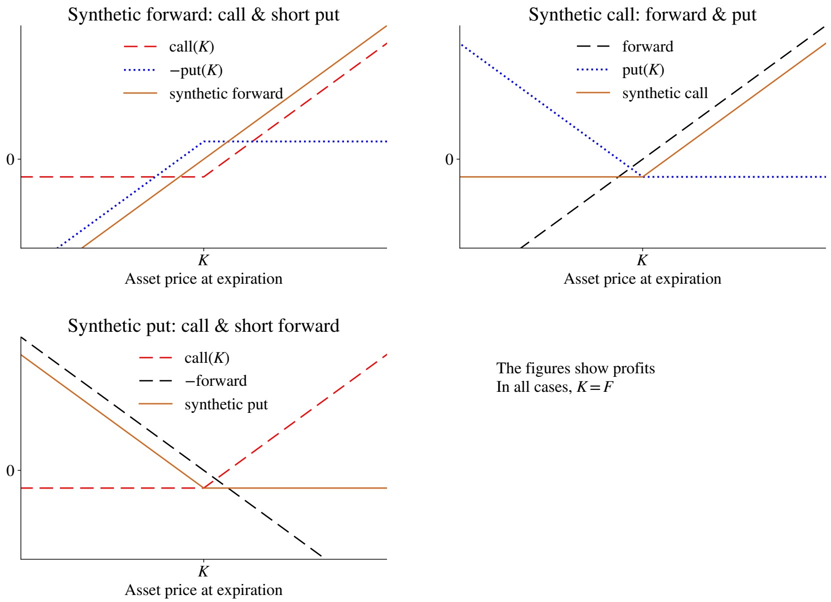


Figure 19.13: Synthetic replication

19.13 for illustrations. Further details are in the following remarks.

Remark 19.13 *(Synthetic forward*) Buy one call and sell one put at a strike price $K = F$. By (19.5), the cost of this portfolio is zero. At expiration, it will give one unit of the underlying, at the cost K. Just like a forward contract. See Figure 19.13.*

Remark 19.14 *(Synthetic call option*) Buy one forward and one put with strike price $K = F$. By (19.5) this has the price C. If $S_m < K$ (at expiration), then the forward pays off $S_m - F$ and the put option $K - S_m$. Since $K = F$, the sum is zero. Instead, if $S_m > K$, then the forward pays off $S_m - F$ and the put nothing. In either case, this is just like a call option with strike price K. See Figure 19.13.*

Remark 19.15 *(Synthetic put option*) Buy one call with strike price $K = F$ and sell one forward. By (19.5), this has the price P. If $S_m < K$ (at expiration) then the call pays off nothing and the short forward $-(S_m - F)$. Since $K = F$, the sum is $K - S_m$. Instead, if $S_m > K$, then the call pays off $S_m - K$ and the short forward $-(S_m - F)$, which sums to zero. In either case, this is just like a put option with strike price K. See Figure 19.13.*

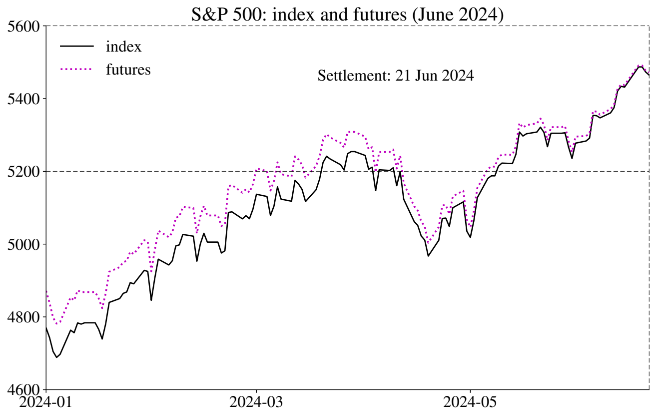


Figure 19.14: S&P 500 index level and futures

19.6 Definition of American Calls and Puts

An American option is similar to a European option, except that it *can be exercised on any day* before or on the expiration date. This means that an American option has more rights than a European option and is therefore worth at least as much

$$C_A \geq C_E \text{ and } P_A \geq P_E, \tag{19.9}$$

where we use subscripts to distinguish between American ($_A$) and European ($_E$) options.

You would only consider exercising an American call option if its profitable ($S > K$) so the immediate payoff is $S - K$, where S should be understood as the current price of the underlying. Instead, if you keep the option, then you know that it always worth 0 or more. A similar logic applies to an American put option. This means that the option prices must (at any point in time) obey

$$\begin{aligned} C_A &\geq \max(0, S - K) \\ P_A &\geq \max(0, K - S). \end{aligned} \tag{19.10}$$

The right hand sides are called the "intrinsic values," which can be thought of as what you get if you decide to get rid of the option today (exercise or burn it).

Empirical Example 19.16 *Figures 19.14 and 19.15 provide an example of how the futures price (on S&P 500), the intrinsic values of two options and their prices*

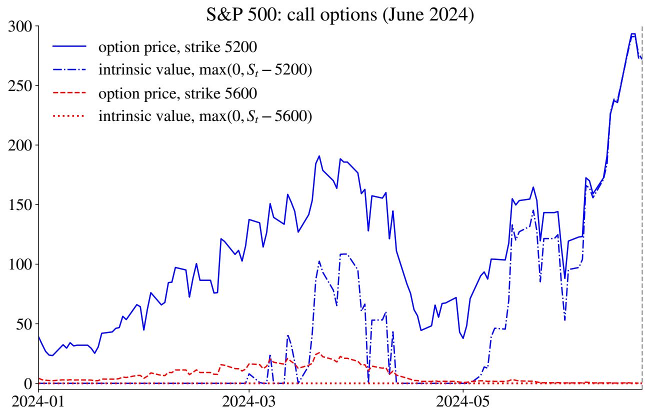


Figure 19.15: S&P 500 options

developed over six months. Notice that also options with zero intrinsic value can have a fairly high option price—at least if the time to expiration is long, but it converges to zero as the expiration date gets closer.

There is no strict put-call parity for American options. However, pricing bounds (based on the values of European options) can be derived.

Remark 19.17 *(Put-call, American option, no dividend) For an American option on an asset without dividends, the put price must be inside the interval*

$$\underbrace{C_A - S + e^{-my}K}_{P_E} \leq P_A \leq \underbrace{C_A}_{C_E} - S + K. \tag{19.11}$$

See Hull (2022) 11 and McDonald (2014) 11 Appendix A for details and proofs.

19.7 Basic Properties of Option Prices

Options prices depend on many things, but there are some fairly general results, which we discuss here.

First, *call option prices are decreasing in the strike price*, while put options prices are increasing in the strike price, see Figure 19.16. The intuition is illustrated in Figure 19.17 which illustrates the perceived (by the market) distribution of the asset price at expiration. Notice that a higher strike price means that an owner of

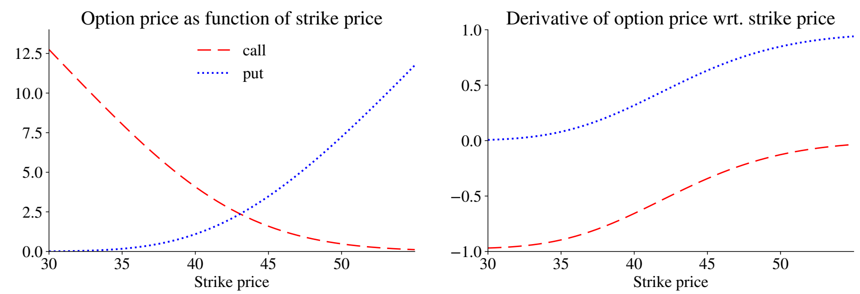


Figure 19.16: Option price as a function of the strike price

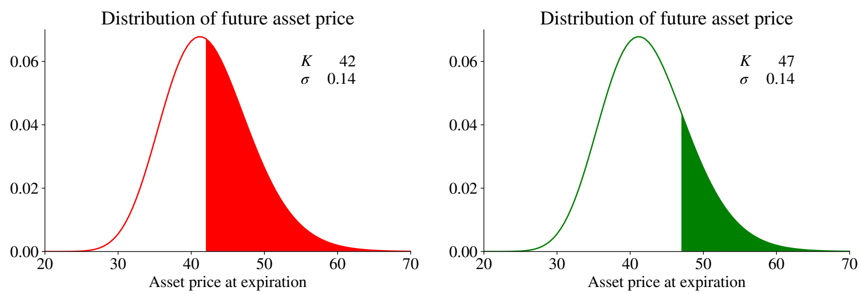


Figure 19.17: Distribution of future asset price

a call option will have to pay more in case of exercise—and there is also a lower chance of exercise.

Actually, it can be shown the call option price is always decreasing in the strike price, but slower than the strike price itself, but that the curve flattens out at high strike prices

$$-e^{-my} \leq dC(K)/dK \leq 0 \text{ and } dC^2(K)/dK^2 \geq 0. \tag{19.12}$$

That is, the first derivative goes from almost -1 to 0 as the strike price increases, see Figure 19.16. In contrast, a put option is increasing in the strike price but slower than the strike price ($0 \leq dP(K)/dK \leq e^{-my}$), but with a similar convexity (positive second derivative). That is, the first derivative goes from 0 to almost 1 as the strike price increases. (See, for instance, McDonald (2014) 11 for proofs.)

Second, both *call and put option prices are typically increasing in the (perceived) uncertainty* of the future price of the underlying asset, see Figure 19.18. The intuition

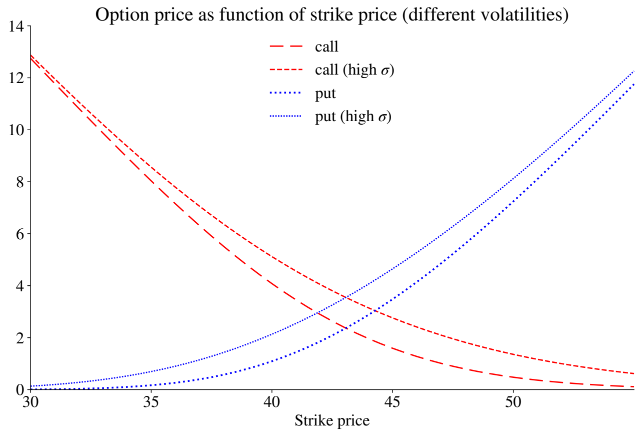


Figure 19.18: Option price as a function of the strike price

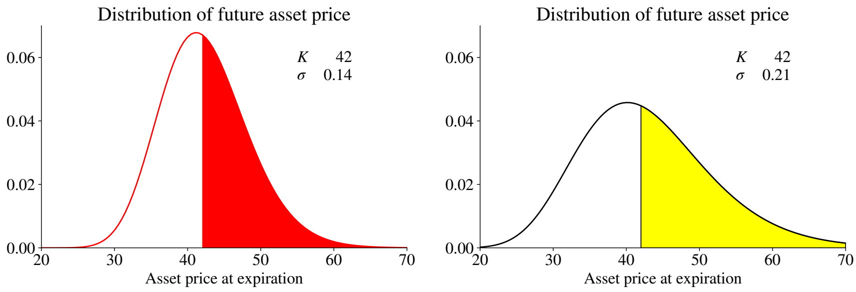


Figure 19.19: Distribution of future asset price

is illustrated in Figure 19.19, which shows that a wider dispersion of the distribution increases the probability of a really high price of the underlying asset (although the figure is constructed to have the same probability of exercise in the two cases). Of course, it also increases the probability of a really low asset price, but that is of no concern since the call option payoff is bounded from below (at zero).

19.8 Pricing Bounds and Convexity

19.8.1 Pricing Bounds for Call Options

The prices of call options must satisfy the following restrictions

$$C \leq e^{-my} F \leq S \tag{19.13}$$

$$0 \leq C \tag{19.14}$$

$$e^{-my}(F - K) \leq C. \tag{19.15}$$

These bounds hold for both American or European call options (we here use C to denote both of them.)

The motivations are basically as follows (the intuition is based on European options, but the results extend to American options as well). First, a call option with a zero strike price ($K = 0$) would be the same as owning a prepaid forward contract (which is worth as much or less than the underlying asset). Whenever the strike price is higher, the call price is lower. Second, the call option gives rights, not obligations: its price value cannot be negative. Third, the lowest possible value of a put option is zero, so the put-call parity (19.5) immediately gives that the call price must exceed the present value of $F - K$. Transaction costs can cause (relatively small) failures of the bounds.

Combining the bounds, we get

$$C \leq e^{-my} F \leq S \tag{19.16}$$

$$C \geq \max[0, e^{-my}(F - K)]. \tag{19.17}$$

For instance, for an asset without dividends until expiration of the option, we have $\max(0, S - e^{-my} K) \leq C \leq S$. See Figure 19.20 for an illustration.

The pricing bounds are typically very wide, so they are of little importance in determining a fair option price. However, they may be helpful in checking data and also as a sanity check of a pricing model.

Example 19.18 *(Pricing bounds for call option) Using the same parameters as in Example 19.11, we get* $C \leq 42$ *and* $C \geq \max[0, e^{-0.5\times 0.05}(43.06 - 38)] = 4.94$.

Empirical Example 19.19 *(The option price bounds in Figure 19.21) At very low strike prices, it is almost certain that the option will be exercised at expiration. Therefore, the present value of the cost,* $C + e^{-my} K$, *must be almost equal to the*

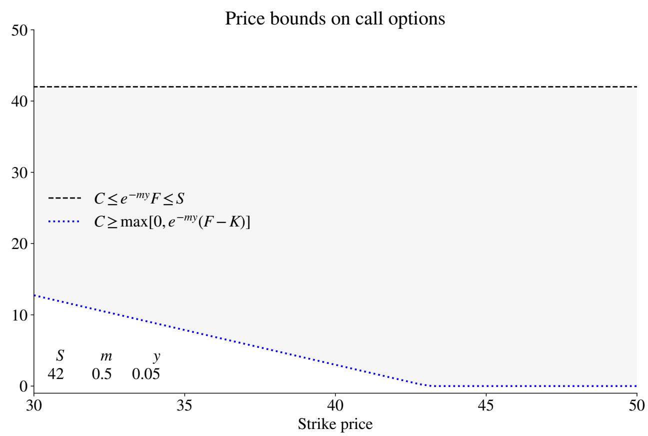


Figure 19.20: Call option price bounds as a function of the strike price

present value of a forward contract, $e^{-my}F$. Combining gives $C = e^{-my}(F - K)$. In contrast, at very high strike prices, the probability of exercise is almost zero—so the option price is too.

19.8.2 Pricing Bounds for Put Options

The prices of American and European put options must satisfy the following restrictions

$$P_E \leq e^{-my}K \text{ and } P_A \leq K \tag{19.18}$$

$$0 \leq P_E \text{ and } 0 \leq P_A \tag{19.19}$$

$$e^{-my}(K - F) \leq P_E \text{ and } K - S \leq P_A. \tag{19.20}$$

See Figure 19.22.

The motivations are as follows. First, the payoff from a put option is $\max(K - S, 0)$, so the maximum value is the strike price (when $S = 0$). For a European put, this payoff is received only at expiration, so the maximum value today is the present value of the strike price. Second, the put option gives rights, not obligations: its price value cannot be negative. Third, the lowest possible value of a call option is zero, so the put-call parity (19.5) immediately gives that the European put price must exceed the present value of $K - F$. In contrast, the American put can be exercised now so its value must be at least as high as the intrinsic value.

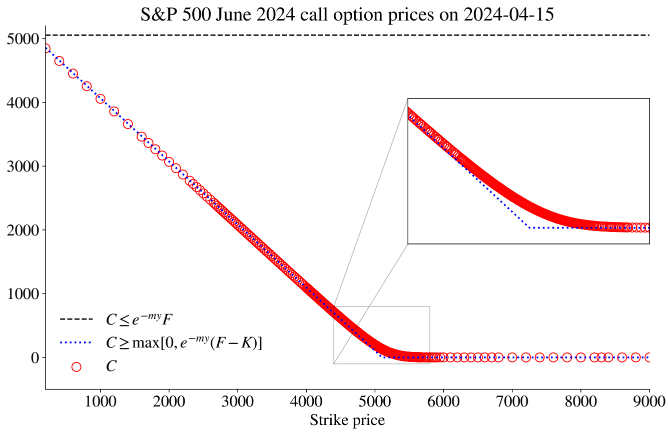


Figure 19.21: Prices and bounds for S&P 500 options

19.9 Early Exercise of American Options

This section discusses early exercise of American options. There are some cases where we can exclude early exercise, so the American option is priced as a European option. In other cases, we cannot exclude early exercise—but we may still be able to say something about when early exercise is likely. More precise answers will require building a model for the pricing. Clearly, the answer is then model dependent.

The key results are as follows (assuming interest rates are positive):

	no dividends	with dividends
Call	no early exercise	early exercise (at high S)
Put	early exercise (at low S)	early exercise

(Negative interest rates means that you could plausibly have early exercise in all four cases.)

Proposition 19.20 *(No early exercise, American call, no dividends) An American call option on an asset without dividends should never be exercised early (if the interest rate is positive). It therefore has the same price as a European call option.*

See Figure 19.23 for an illustration of the fact that early exercise is not profitable for a call on an underlying asset without dividends since $C_A \geq C_E > \max(0, S-K)$, so the market price of the American call option will always be higher (or equal) to

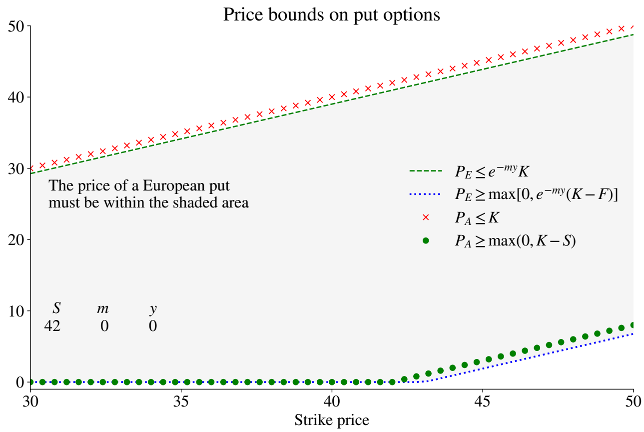


Figure 19.22: Put option price bounds as a function of the strike price

what you get by exercising. Rather, sell the option. This is potentially different for a put, also illustrated in the same figure. In particular, it shows that we cannot guarantee that $P_E > \max(0, K-S)$, which opens up the possibility of early exercise.

Proof (of Proposition 19.20) From the put-call parity for European options (19.5), $C_E = P_E + S - e^{-my}K$, we have $C_E \geq S - K$ as long as the interest rate is positive (since $P_E \geq 0$). Since $C_A \geq C_E$, selling the option gives more than exercising it. □

Example 19.21 *(Bankruptcy, American put, no dividends) Suppose the underlying asset goes bankrupt, then $S = 0$ and it is known that it will stay at $S = 0$. Exercising the American put option now gives K, whereas waiting until expiration has a present value of $e^{-my}K$ (which is lower): early exercise is optimal.*

See Figure 19.24 for an illustration, based on a numerical solution (of a specific model, so the precise results are not general) for the price on American options. In particular, it shows in which nodes early exercise is optimal for an American put option: at low asset prices. The Figure also illustrates that a numerical calculation agrees that an American call option is not exercised early.

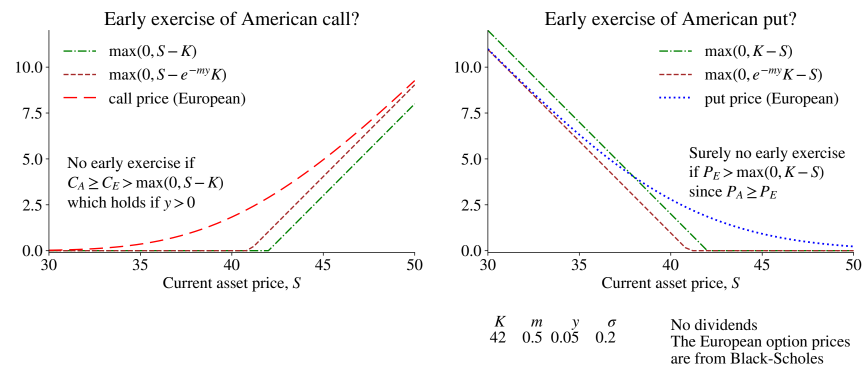


Figure 19.23: Early exercise of American call and put options (no dividends)

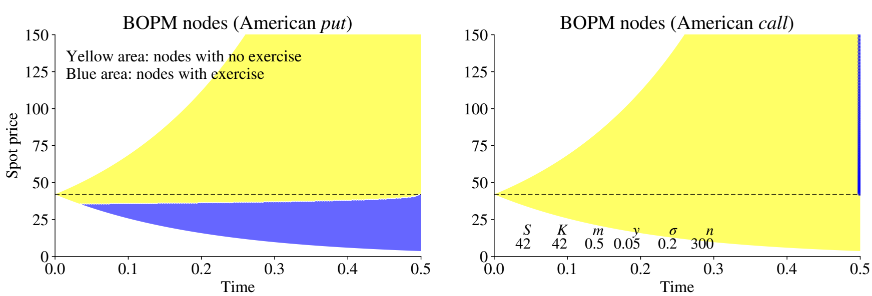


Figure 19.24: Numerical solution of American put and prices

Chapter 20

The Binomial Option Pricing Model

This chapter uses a binomial process (up/down) for the price of the underlying asset price as a basis for pricing options. By chaining many up/down moves in such a way that the model generates price movements similar to those in data, a realistic process is created. Option prices can then be calculated using a recursive approach. This can handle both European and American options.

20.1 Overview of Option Pricing

There are basically two ways to model option prices: a factor model (such as CAPM) or a no-arbitrage argument. This focus on the latter, in particular, the binomial model (BOPM).

20.2 The Basic Binomial Model

In the binomial model option pricing model (see Cox, Ross, and Rubinstein (1979) and Rendleman and Bartter (1979)), the price of the underlying asset can change in only two ways. This is very stylized, but it provides a foundation for a more realistic model (by cumulating many short subperiods). When applied to a European-style option, the binomial model converges to the well-known Black-Scholes model, as the subperiods become very many and short. However, the binomial model can also be applied to an American-style option (Black-Scholes cannot).

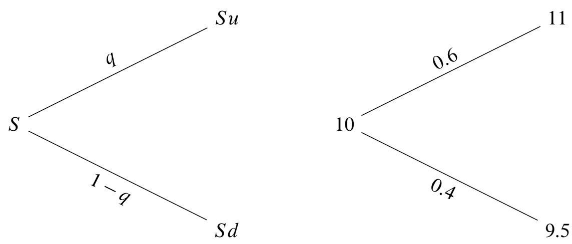


Figure 20.1: Binomial process for S and a numerical example

20.2.1 A Binomial Process for the Price of the Underlying Asset

The binomial tree for the underlying asset starts at the current price S and has probability q of moving to Su ($S \cdot u$, where typically $u > 1$) in the next period and a probability of $1 - q$ of moving to Sd (where $d < u$). This is illustrated in Figure 20.1. These probabilities are the true ("natural") probabilities.

Example 20.1 *(Binomial process) Suppose* $S = 10, u = 1.1, d = 0.95$*, and* $q = 0.6$*. Then, the process has a 60% probability of increasing from 10 to 11 and a 40% probability of decreasing to 9.5. See Figure 20.1.*

We take it for granted that

$$u > e^{yh} > d, \tag{20.1}$$

where h is time length (measured in years, since the interest y is) between this and the next period. If this condition is not satisfied, then trivial arbitrage opportunities arise. For instance, if $e^{yh} > u$, then we could shorten the underlying asset and buy bonds: this would *guarantee* a positive payoff for a zero investment (an arbitrage possibility).

20.2.2 No-Arbitrage Pricing of a Derivative

Basic Setup

Consider a derivative asset that will be worth f_u in case the underlying asset ends up at Su and f_d if it ends up at Sd. Notice that f_u is just the notation for the value (price) of the derivative in the up state (it should *not* be read as f times u).

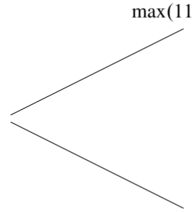


Figure 20.2: Numerical example of call option payoff, $K = 10$

As an example, when the derivative is a call option with strike price K and that the next period is the expiration date, then,

$$f_u = \max(Su - K, 0) \text{ and } f_d = \max(Sd - K, 0). \tag{20.2}$$

Example 20.2 *(European call option) With the parameters in Example 20.1, equation (20.2) shows that a European call option with strike price of 10 has* $f_u = \max(11-10, 0) = 1$ *and* $f_d = \max(9.5-10, 0) = 0$*. See Figure 20.2. In contrast, a strike price of 9 gives* $f_u = 2$ *and* $f_d = 0.5$*.*

Step 1: Construct a risk-free Portfolio

We now use a no-arbitrage argument to determine the present price of the derivative, denoted f. Consider the following portfolio

$$\begin{aligned} &\Delta \text{ of the underlying asset, and} \\ &-1 \text{ of the derivative,} \end{aligned} \tag{20.3}$$

where Δ is yet to be decided. (Note that Δ here denotes a quantity, *not* a difference.)

For a given value of Δ, the payoff of the portfolio in the next period is $\Delta Su - f_u$ in the "up" state and $\Delta Sd - f_d$ in the "down" state. To make the portfolio *risk-free*, Δ must be such that the payoff is the same in both states

$$\begin{aligned} \Delta Su - f_u &= \Delta Sd - f_d, \text{ so} \\ \Delta &= \frac{f_u - f_d}{S(u-d)}. \end{aligned} \tag{20.4}$$

With this choice of Δ (also called the "delta hedge") the portfolio is risk-free. For future reference, we can also notice that Δ looks like a derivative, $\partial f/\partial S$.

Example 20.3 *(European call option) Continuing from Example 20.2 we get*

$$\Delta = \frac{1-0}{10(1.1-0.95)} = \frac{2}{3} \text{ for } K = 10.$$

The payoff of this portfolio is indeed safe. For instance, for the $K = 10$ *option, the value in the up state is* $(2/3)\cdot 11 - 1 = 19/3$ *and in the down state* $(2/3)\cdot 9.5 - 0 = 19/3$. *For a* $K = 9$ *call option,* $\Delta = 1$.

Step 2: Make the Return of the Portfolio Equal to the Risk-free Rate

Since the choice of Δ in (20.4) makes the portfolio safe, it must have *same return as the risk-free asset* (otherwise, arbitrage opportunities would arise). This is the same as requiring that the present value of the portfolio payoff (left hand side in the equation below) equals the cost of the portfolio today (right hand side)

$$e^{-yh}(\Delta S u - f_u) = \Delta S - f, \tag{20.5}$$

where Δ is determined as in (20.4). We could equally well have used the payoff in the down state, $\Delta S d - f_d$, since it is the same. This equation defines the (current) arbitrage-free price f of the derivative.

Solve (20.5) for f and then use the value of Δ from (20.4) that ensures that the portfolio is risk-free

$$\begin{aligned} f &= \Delta S(1 - e^{-yh}u) + e^{-yh} f_u && (20.6)\\ &= \frac{f_u - f_d}{u-d}(1 - e^{-yh}u) + e^{-yh} f_u && (20.7)\\ &= e^{-yh}[p f_u + (1-p) f_d] \text{ with } p = \frac{e^{yh} - d}{u-d} && (20.8)\\ &= e^{-yh}\,\mathrm{E}^*(\text{future payoff of derivative}). && (20.9) \end{aligned}$$

These are alternative ways to express the price of the derivative, f.

Equation (20.7) shows what the price of the derivative must be, and is written in terms of the possible outcomes and the interest rate. Notice that neither probabilities, nor risk preferences enter this expression, since we have used a no-arbitrage argument to price this derivative. This works because there are as many relevant assets, (risk-free and underlying asset) as there are possible outcomes (up or down), meaning that it is possible to construct a risk-free portfolio).

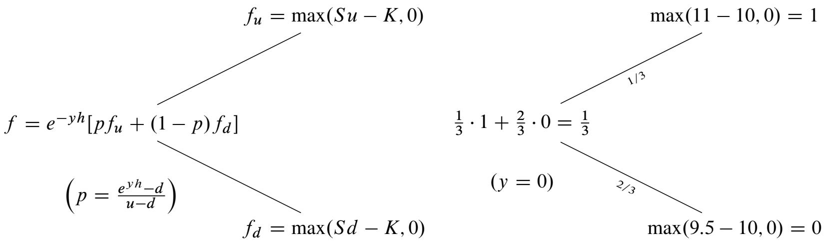


Figure 20.3: Solving for a call option price and numerical example. The numerical example is based on Example 20.4 and assumes $y = 0$.

Equation (20.8) shows that the current price of the derivative is the present value of what *looks like* an expectation of the payoff of the derivative ($pf_u + (1-p)\, f_d$). This expression is quite useful since we can think of p as a "risk neutral probability" although it is not a probability in the usual sense: it is just a convenient construction. Note, though, that under the restrictions in (20.1), $0 < p < 1$, as any "probability" should be. This interpretation is highlighted in (20.9), where E^* stands for the expectations according to the *risk neutral distribution* (more about that later). The computation in (20.8) is illustrated in Figure 20.3.

The risk neutral probability p does not depend on the specific derivative considered, as long as it has the same underlying asset; rather, p depends only on the underlying asset and the interest rate. See Hull (2022) 13 and McDonald (2014) 13–14 for further details.

Example 20.4 *(European call option) Continuing from Example 20.2 and assuming that $y = 0$, equation (20.8) gives the price of a call option with strike price 10 as*

$$
\begin{aligned}
f &= e^{-0}\,[p1 + (1-p)\,0] \text{ with } p = \frac{1 - 0.95}{1.1 - 0.95} = 1/3 \\
&= 1/3.
\end{aligned}
$$

See Figure 20.3. For the call option with a strike price of 9, we get $f = 1$.

20.2.3 Applying the No-Arbitrage Pricing on Different Derivatives

This section discusses how the pricing formula (20.8) can be applied to specific derivatives.

A *forward contract* has a zero current price (nothing is paid until expiry), and the payoff at expiry is $f_u = Su - F$ in the up state (the value of the underlying asset minus the forward price) and $f_d = Sd - F$ in the down state. Using this in (20.8) gives

$$0 = e^{-yh} [p (Su - F) + (1 - p) (Sd - F)], \text{ so} \tag{20.10}$$

$$F = pSu + (1 - p) Sd. \tag{20.11}$$

This shows that the mean of the risk neutral distribution equals the forward price.

Example 20.5 *(A forward contract) Continuing from Example 20.4, we get*

$$F = (1/3) \cdot 11 + (2/3) \cdot 9.5 = 10,$$

which is the same as $S = 10$ (since the interest rate is zero).

An "*Arrow-Debreu asset*" (a sort of theoretical derivative often used in asset pricing models) pays off one unit in the up state and zero otherwise ($f_u = 1$ and $f_d = 0$). This is also a so-called "cash-or-nothing" call option provided the up state means that the option is in the money ($Su > K$). From (20.8) we have

$$f = e^{-yh} p. \tag{20.12}$$

20.2.4 Change of Measure

The basic principle used in the previous analysis is summarized in the following proposition.

Proposition 20.6 *(Risk-neutral valuation) Today's price of the underlying asset (here S_t) equals the present value of the risk-neutral expected future asset price (S_{t+h})*

$$S_t = e^{-yh} \mathrm{E}_t^* \, S_{t+h},$$

which holds with p from (20.8).

This means that, according the risk-neutral distribution, all assets have the same expected gross return as the risk-free rate, so $\mathrm{E}_t^* \, S_{t+h}/S_t = pu + (1 - p) \, d = e^{yh}$. The same principle applies to all asset, in particular, derivatives of the underlying asset. This switch from the physical probabilities q to the risk-neutral probabilities p is sometimes called a "change of measure," which essentially implies adjusting the mean of the distribution.

Applying the same principle on a derivative $f(S)$ gives

$$f(S_t) = e^{-yh} \mathrm{E}_t^* f(S_{t+h}), \tag{20.13}$$

which is the same as (20.8). This expectation uses the same risk-neutral probabilities. (Recall that, in general, for a function $g(x)$, we have $\mathrm{E}\, g(x) = \Sigma \pi_i g(x_i)$ where π_i is the probability of x_i.)

Proof (of Proposition 20.6) Use $\mathrm{E}_t^* S_{t+h} = pS_t u + (1-p) S_t d$ (as in (20.11)) in the proposition to get $S_t = e^{-yh}[pS_t u + (1-p) S_t d]$. Simplify and notice that this gives the p value in (20.8). □

20.2.5 Replicating (and Hedging) a Derivative

The no-arbitrage argument in (20.4) was based on the fact that a portfolio with Δ of the underlying asset and -1 of the derivative replicates a safe asset.

This argument can be turned around to replicate the derivative by holding the following portfolio (these are values of the positions)

$$\begin{aligned} &\Delta S \text{ in the underlying asset, and} \\ &f - \Delta S \text{ in a safe asset.} \end{aligned} \tag{20.14}$$

This means that we hold Δ underlying assets (same Δ as in (20.4)) and hold the amount $f - \Delta S$ on a money market account (typically negative, meaning borrowing). This replicates the derivative's payoff. We can therefore hedge a short position in the derivative by portfolio (20.14).

Proof (of that (20.14) replicates the derivative) The payoff of this portfolio is $\Delta S u + e^{yh}(f - \Delta S)$ in the up state and $\Delta S d + e^{yh}(f - \Delta S)$ in the down state. For the up state, notice from (20.5) that $e^{yh}(f - \Delta S) = f_u - \Delta S u$, so the payoff in the up state simplifies to f_u. Also, Δ is such that $f_d - \Delta S d = e^{yh}(f - \Delta S)$, so the payoff in down state simplifies to f_d. The portfolio replicates the derivative. □

Example 20.7 *(Replicating a call option) For the call option with a strike price of 10 and with a zero interest rate, we have (see Examples 20.3 and 20.4)* $\Delta = 2/3$, $f = 1/3$ *and*

$$(f - \Delta S) = \frac{1}{3} - \frac{2}{3} \cdot 10 = -6\frac{1}{3},$$

so we borrow. The value of this portfolio in the up node is $\frac{2}{3} \cdot 11 - 6\frac{1}{3} = 1$ *and in the down node* $\frac{2}{3} \cdot 9.5 - 6\frac{1}{3} = 0$ *which are the same as the call option.*

20.2.6 Where is the Risk Premium?

We have used a no-arbitrage method to price the derivative. It works since the derivative is a redundant asset: it can be replicated by a portfolio of the underlying asset and a risk-free asset (see (20.14)) and therefore must have the same price as this portfolio. Clearly, this portfolio will incorporate a risk premium and so must the derivative.

It may seem as if the pricing formula (20.8) is free from the preference parameters that would determine the risk premium. However, that is not correct. The pricing formula contains the current asset price (through f_u and f_d) which is indeed affected by preference parameters.

20.3 Multi-Period Trees I: Basic Setup

20.3.1 The Binomial Tree for the Underlying Asset

In numerical applications, we chain a large number of up/down movement to get more realistic model properties of the underlying asset. This means that the (fixed) time to expiration is divided into more small time steps and that we can rebalance the portfolio at each of them.

Figure 20.4 is an illustration of a binomial tree with two subintervals. This tree has only three final nodes since $Sud = Sdu$: it is "recombining," which is very useful to keep the number of nodes manageable. This would not be the case if the up and down moves differed across subperiods (a non-iid price process). See Remark 20.9 for details.

Let m be the time to expiration of the derivative, typically measured in years. With n short time intervals, the length of each interval is $h = m/n$. Clearly, if we use more time steps, then each of them is shorter. The size of the up and down movements, as well as the discounting, must therefore be scaled by the number of time steps: compare Figures 20.3 and 20.4. Otherwise we cannot preserve/control the general properties of the underlying asset. Later sections will discuss this recalibration in detail.

Remark 20.8 *(Building the tree*) One way of building the tree is to calculate the value of the underlying asset in a node as* $Su^{N_u}d^{N_d}$ *where* N_u *is the number of up steps and* N_d *the number of down steps since the beginning of the tree. For time step* i, $(N_u, N_d) = (i, 0)$ *for the top node,* $(i-1, 1)$ *for the second node and so forth*

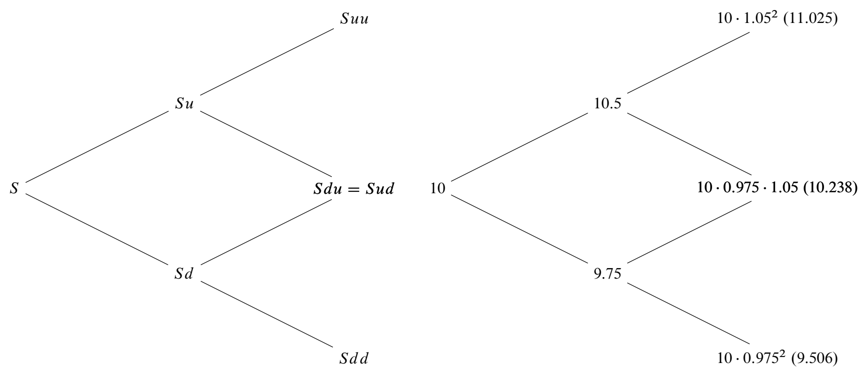


Figure 20.4: Binomial process ($n = 2$) for S and a numerical example

until $(0, i)$ *for the bottom node. In short, at time step* i*,* $N_u = [i, i-1, ..., 0]$ *and* $N_d = i - N_u$.

Remark 20.9 *(Size of the binomial tree) With n time steps, there are* $n+1$ *different prices at the end nodes. Also, there are a total of* $(n+1)(n+2)/2$ *nodes. There are* $n!/[(n-s)!s!]$ *different ways to reach the sth node below the top node (where* $x! = x \cdot (x-1) \cdot \ldots \cdot 1$*). Summing across the nodes shows that the tree contains* 2^n *different paths. In particular, our recombining tree with n time steps has*

n	*no. end nodes*	*no. total nodes*	*no. paths*
2	3	6	4
25	26	351	33,554,432
200	201	20,301	1.6×10^{60}

In contrast, a non-recombining tree has 2^n *end nodes, that is, as many as there are paths in the recombining tree. This can easily cause numerical issues.*

20.3.2 Using a Binomial Tree for Pricing European Options

We can now apply the pricing formula (20.8) to each "subtree," *starting at the end* of the tree (time step n) and *working backwards* towards the start of the tree (time step 0). Figure 20.5 illustrates the computations for a European call option with strike price K and two steps ($n = 2$). To simplify the figure, we assume $y = 0$. Figure 20.6 gives a numerical example.

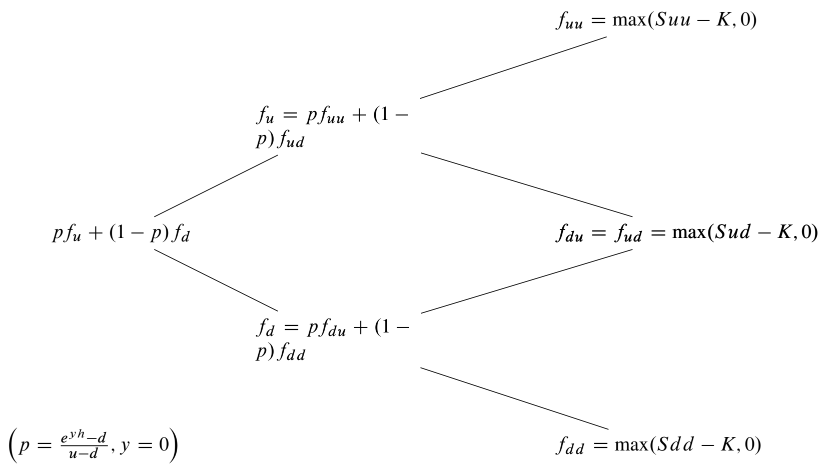


Figure 20.5: Binomial tree for European call option ($n = 2$), zero interest rate

The structure of the tree for a European put option is the same as for a European call option, except that the payoff at the end nodes differ ($\max(0, S_m - K)$ for the call and $\max(0, K - S_m)$ for the put), see Figure 20.7.

Example 20.10 *(Tree for a European put) For a put option with strike price $K = 10$, the values in Figure 20.6 would change to $f = 0.219$, $(f_u, f_d) = (0, 0.329)$ and $(f_{uu}, f_{ud}, f_{dd}) = (0, 0, 0.494)$.*

This recursive calculation (using a tree with $n = 2$ as in Figure 20.5) gives a European option price of the following general form

$$\begin{aligned} f &= e^{-yh}[pf_u + (1-p)f_d] \\ &= e^{-ym}[p^2 f_{uu} + 2p(1-p)f_{ud} + (1-p)^2 f_{dd}], \end{aligned} \tag{20.15}$$

since $e^{-y2h} = e^{-ym}$. The end node values (f_{uu}, f_{ud}, f_{dd}) depend on the type of derivative (call or put).

Notice that p^2 is the risk-neutral probability of the payoff f_{uu}, $2p(1-p)$ of the payoff f_{ud} and $(1-p)^2$ of the payoff f_{dd}, as illustrated in Figure 20.8. Therefore, (20.15) is a generalisation of (20.9):

$$f = e^{-ym}\,\mathrm{E}^*(\text{payoff of derivative at expiration}), \tag{20.16}$$

which says that the (European-style) derivative is the present value of the risk-neutral

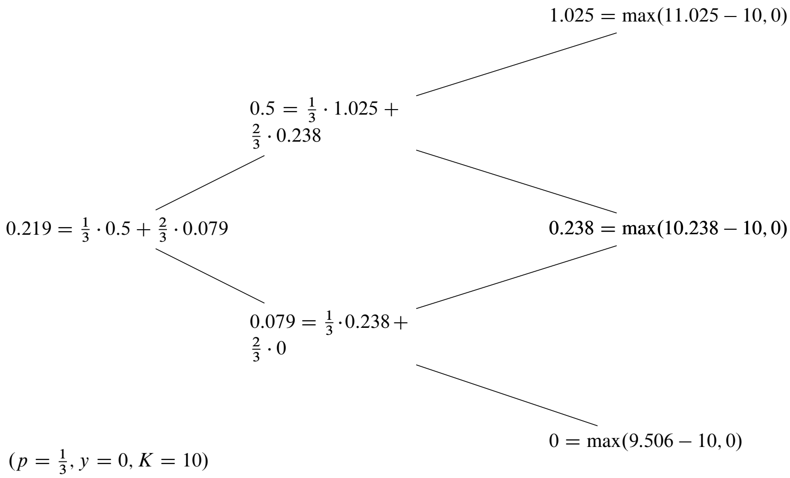


Figure 20.6: Numerical example of binomial tree for European call option ($n = 2$), zero interest rate. The underlying is described in Figure 20.4 .

expected payoff at expiration. The distribution behind the risk neutral expectation is clearly more involved than before, but the same logic applies. This will later be extended from 2 to n time steps.

Remark 20.11 *(The binomial distribution*) After n independent draws, the number of up moves (k) has the binomial pdf,* $n!/[k!(n-k)!]\,p^k(1-p)^{n-k}$ *for* $k = 0, 1, ..., n$*. For instance, with* $n = 2$*, we have* p^2 *for* $k = 2$*,* $2p(1-p)$ *for* $k = 1$*, and* $(1-p)^2$ *for* $k = 0$*.*

20.3.3 Using a Binomial Tree for Pricing American Options

The binomial tree we have used so far assumes that the derivative is "alive" until expiration. This is not necessarily the case for American options, so the approach needs to be modified to handle the possibility of early exercise.

Whenever you can exercise, the option value is the maximum of the exercise value and the value of keeping the option alive

$$\max(\text{value if exercised now, value of keeping an unexercised option}). \quad (20.17)$$

The value of an unexercised option is calculated as for a European option (see (20.8)): the present value of the risk neutral expected value in the next time step.

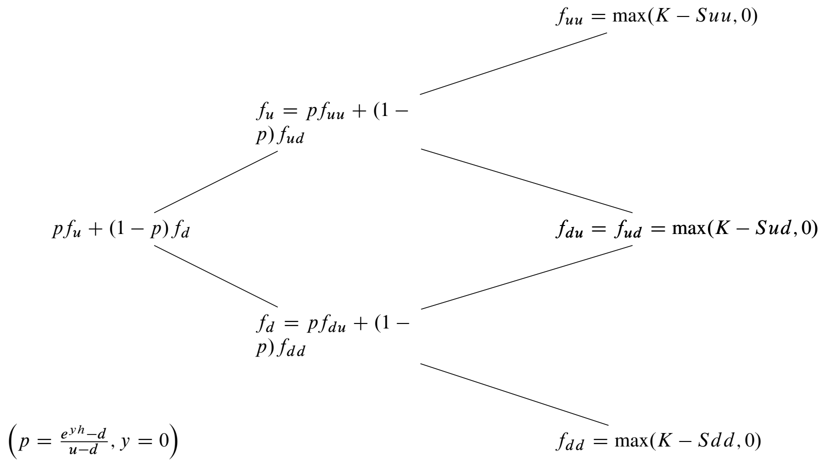


Figure 20.7: Binomial tree for a European put option ($n = 2$), zero interest rate

This means that we solve this problem starting from the expiration date (just like for the European options), and calculate the value at each node, assuming, perhaps counter factually, that the option has not already been exercised at an earlier time step. See Figure 20.9 for an illustration. Also, see Figure 20.10 for a numerical example (the nodes where exercise is optimal are indicated by bold).

Figure 20.11 illustrates the solution for an American put option on an asset without dividends (the details of the calculations will be discussed later). Notice that the American put price exceeds the European put price, and more so at low asset prices and high interest rates. The lower and upper limits on the put price are from the put-call "parity" (two inequalities) for American options. The call price C used in the figure is the same for European and American options (since there is no early exercise in this case).

20.4 Multi-Period Trees II: Calibrating the Tree

We now discuss how to construct a binomial tree (how to choose u and d) with *many* small time steps, so that it mimics the statistical properties of the underlying asset. The model has simple structure, with few degrees of freedom. However, we can calibrate it to fulfill its key role: price derivatives. To do that, we need the risk-neutral distribution to give an expected return equal to risk-free rate and a

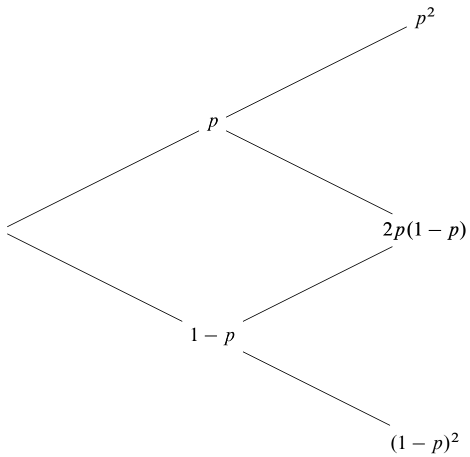


Figure 20.8: Probabilities of different nodes in a binomial tree

variance similar to the that of the underlying asset.

20.4.1 Variance of (data on) Returns

Suppose you have a sample of log returns (r_τ for $\tau = 1$ to T) of the underlying asset, and that you are willing to assume that they are *iid*. Calculate the sample variance and *annualize*

$$\hat{\sigma}^2 = \widehat{\mathrm{Var}}(r_\tau)/k, \tag{20.18}$$

where $1/k$ is the number of periods needed to get a full year. For instance, with daily return data $1/k = 252$ (only counting the trading days). Annualizing helps relating to the binomial model below. When needed, this estimate is adjusted to account for the current state of the market.

Example 20.12 *(Variance for daily return) If the data is daily ($k = 1/252$) and the standard deviation is estimated to be* 0.0315*, then the annualised variance is* $\hat{\sigma}^2 = 0.0315^2 \cdot 252 \approx 0.5^2$ *and the annualized standard deviation is* $\hat{\sigma} \approx 0.5$*.*

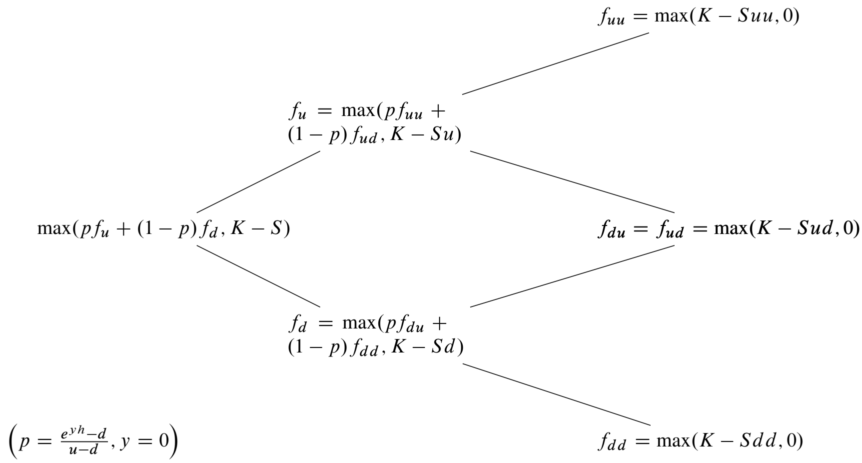


Figure 20.9: Binomial tree for an American put option ($n = 2$), zero interest rate

20.4.2 Risk-Neutral Mean and Variance according to the Binomial Model

Recall the binomial process (for instance, in Figure 20.1) implies that the log returns, $r_{t+h} = \ln(S_{t+h}/S_t)$, follow

$$r_{t+h} = \begin{cases} \ln u & \text{with probability } p \\ \ln d & \text{with probability } 1-p. \end{cases} \tag{20.19}$$

Remark 20.13 *(Mean and variance of a binomial process) The mean of a (shifted) binomial process like (20.19) is* $p \ln u + (1-p)\ln d$ *and the variance is* $p(1-p)(\ln u - \ln d)^2$*. If needed, we divide both by h to get annualized values.*

Example 20.14 *(Binomial process) Using the numbers in Example 20.4 gives a variance of* $(1/3) \cdot (2/3) \cdot (\ln 1.1 - \ln 0.95)^2$*. If the periods in the model are weeks (*$h = 1/52$*), then the annualized variance is approximately* 0.248 *and the annualized standard deviation approximately* 0.5.

20.4.3 Comparing Data and Model: The CRR Approach

As mention in Proposition 20.6, the choice of p according to (20.8) guarantees that the risk-neutral tree gives an *expected gross return* equal to e^{yh}. (This is not the exactly same as $\mathrm{E}\, r = y$, due to the logarithm, to be be discussed later.)

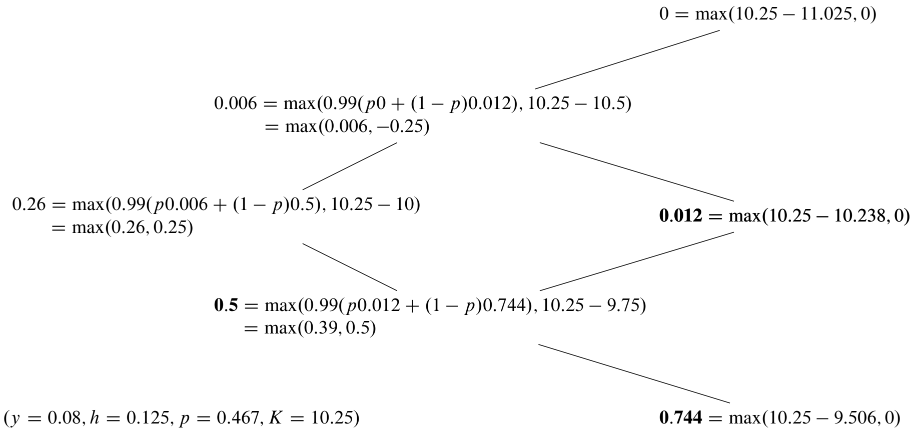


Figure 20.10: Numerical example of a binomial tree for an American put option ($n = 2$). Exercise is indicated by bold.

However, this leaves the choice of the (u, d) movements. Since option prices heavily depend on the volatility of the underlying asset and we want to make the model realistic, we typically calibrate (u, d) so the implied variance in the model is similar to the variance of actual returns, $\hat{\sigma}^2$ from (20.18).

The most common approach is that of Cox, Ross, and Rubinstein (1979) where

$$u = e^{\hat{\sigma}\sqrt{h}} \text{ and } d = 1/u. \tag{20.20}$$

Recall that p needs to change when (u, d) change, since p follows (20.8), that is, $p = (e^{yh} - d)/(u - d)$. (There are other approaches, for instance, that of Jarrow and Rudd (1983), but they have similar properties.) This gives the annualized risk-neutral variance as

$$\text{(annualized) } \operatorname{Var}(r) = \hat{\sigma}^2 \text{ as } n \to \infty. \tag{20.21}$$

Figure 20.12 shows an numerical example. A proof is given towards the end of the section.

Example 20.15 *(Parameters to binomial tree) With $h = 1/52$ and $\hat{\sigma} = 0.2$, (20.20) gives $u \approx 1.028$ and $d \approx 0.973$.*

However, we must ensure that (20.1) holds, that is,

$$u > e^{yh} > d, \tag{20.22}$$

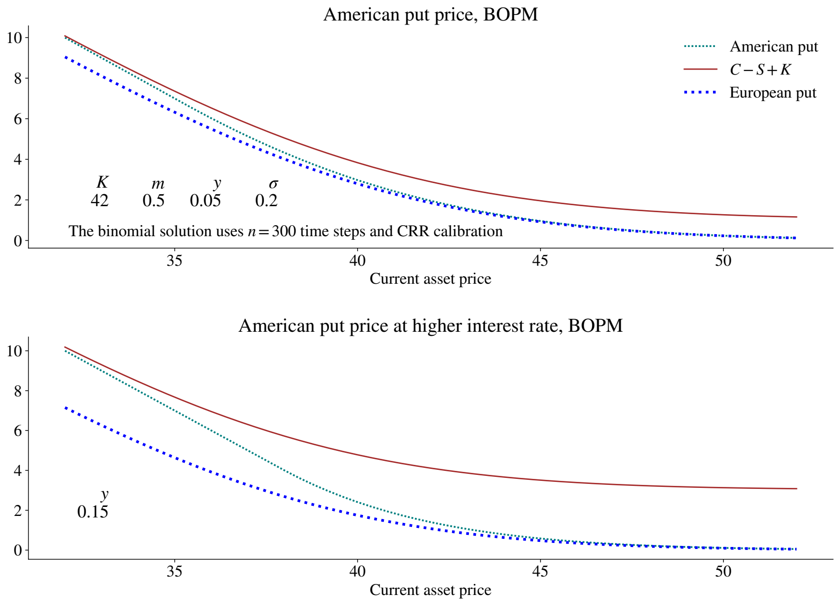


Figure 20.11: Numerical solution of an American put price

which requires $\hat{\sigma} > y\sqrt{h} > -\hat{\sigma}$. In practice, this means that h must be small (the number of steps, n, large). Always check that this condition is satisfied. Otherwise, the results of the calculations might be nonsense.

Example 20.16 *(Checking parameters of binomial tree) With the parameters in Example 20.15 and assuming $y = 0.05$, we notice that $e^{yh} = e^{0.05/52} \approx 1.001$, so the requirement is fulfilled*

$$1.028 > 1.001 > 0.973.$$

See Figure 20.13 for an illustration of how the parameters (p, u, d) converge as the number of time steps increases. The figure also shows that (20.22) holds.

Remark 20.17 *(The physical tree*) If we instead use a physical probability $q = (e^{\mu h} - d)/(u - d)$, then the tree will generate an expected gross return of $e^{\mu h}$. The physical variance converges to $\hat{\sigma}^2$ as $n \to \infty$, just like (20.21). See Figure 20.12 for an illustration where μ is higher than y. (To prove the result, replace p by q in the proof of (20.21).)*

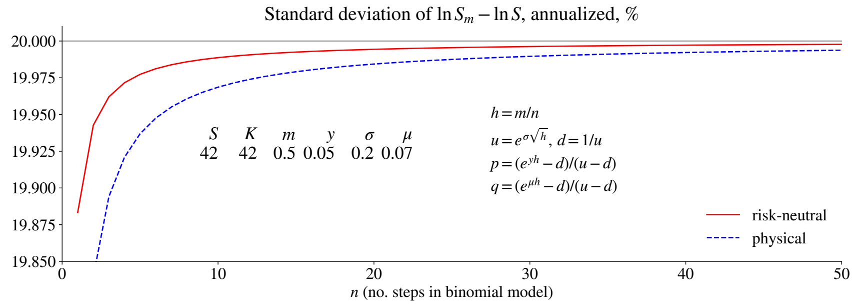


Figure 20.12: Convergence of the standard deviation of r in the binomial model

See Figure 20.14 for an illustration of how the resulting option price converges as the number of time steps increases. (The result from the Black-Scholes model will discussed in detail in another chapter.) The zig-zag pattern suggests that some kind of average price, across $n-1$ and n steps, may improve the performance (see Figure 20.14).

Figure 20.15 illustrates the calculations of the American put price for a single current value of the underlying asset (S). The coloured area shows the location of the nodes (possible prices of the underlying asset in the future) that are used in the calculation—and at which nodes that early exercise will happen. For comparison, the figure also shows that the numerical calculations verify the theoretical result that an American call option is not exercised early.

Proof (of (20.21)*) Recall that $\ln u = \hat{\sigma}\sqrt{h}$ and $\ln d = -\hat{\sigma}\sqrt{h}$. Calculate the mean as $\mathrm{E}\, r = (2p-1)\hat{\sigma}\sqrt{h}$. Similarly the 2nd moment is $\mathrm{E}\, r^2 = p(\ln u)^2 + (1-p)(\ln d)^2 = \hat{\sigma}^2 h$. Recall that $\mathrm{Var}(r) = \mathrm{E}\, r^2 - (\mathrm{E}\, r)^2$. Combining (and dividing by h) gives the annualized variance $\mathrm{Var}(r)/h = [1-(2p-1)^2]\hat{\sigma}^2$. Notice that $(2p-1)^2$ goes (quickly) to 0 as p goes to $1/2$ as h decreases. (See, for instance, Vrins (2025), Appendix E for an alternative proof.) □

20.5 Appendix – Continuous Dividends*

It is straightforward to construct another tree that allows for continuous dividends, provided they are proportional to the asset price.

Suppose dividends are paid at the known continuous *rate* δ and let the up and down movements in the asset price reflect the ex-dividend price (S in the initial

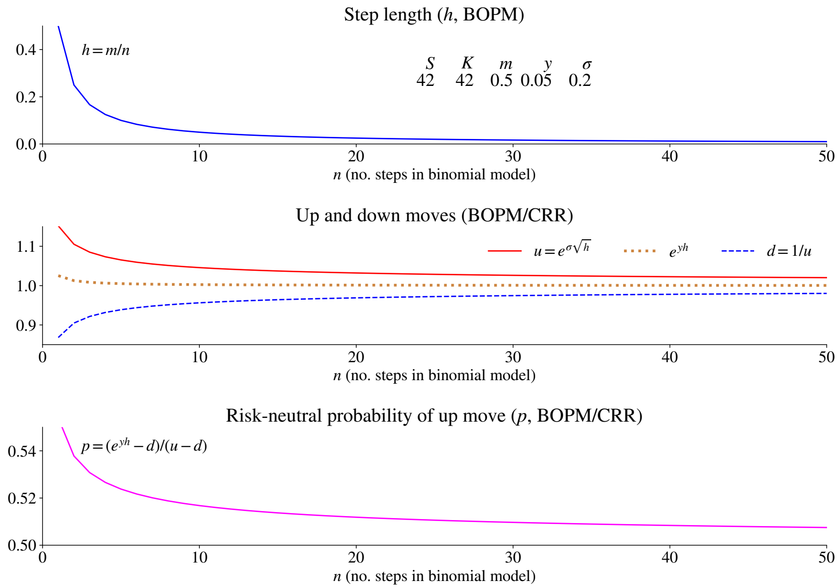


Figure 20.13: Convergence of the parameters in a binomial model

period). Buying one unit of the underlying asset in the initial period costs S. If we move to the "up state" in the next period (h), then we first get the dividend $Su(e^{\delta h} - 1)$ and can then sell the asset for the (ex-dividend) price Su: the total value is $Sue^{\delta h}$. Notice that the dividend is proportional to price in the same period. The "down state" is similar: just replace u by d.

We now construct a risk-free portfolio to find out how a derivative is priced in the initial period. First, to construct a risk-free portfolio, hold Δ of the underlying asset and -1 of the derivative. The payoff of the portfolio at expiry is $\Delta Sue^{\delta h} - f_u$ in the "up" state and $\Delta Sde^{\delta h} - f_d$ in the "down" state. To make the portfolio risk-free the delta must be

$$\Delta = \frac{f_u - f_d}{Se^{\delta h}(u-d)}. \tag{20.23}$$

Second, to make the return of the portfolio equal to the risk-free rate, we set the present value of our risk-free portfolio equal to the cost of the portfolio

$$e^{-yh}(\Delta Se^{\delta h}u - f_u) = \Delta S - f. \tag{20.24}$$

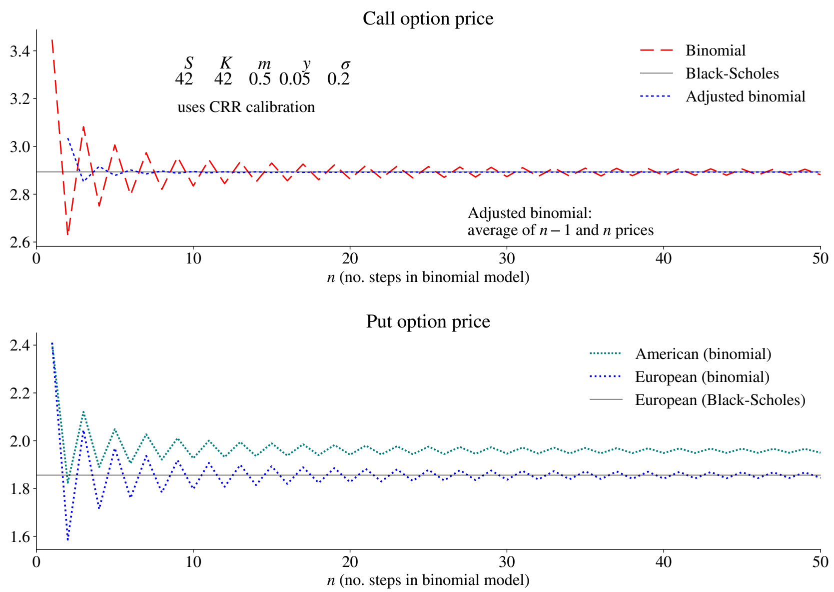


Figure 20.14: Convergence of the binomial option price

Use (20.23) and rearrange as

$$f = \Delta S(1 - e^{(\delta - y)h}u) + e^{-yh} f_u \tag{20.25}$$

$$= e^{-yh}[p f_u + (1-p) f_d] \text{ with } p = \frac{e^{(y-\delta)h} - d}{u - d}. \tag{20.26}$$

With this new definition of p, the rest of the computations are as in the case without dividends. In particular, the drift of the asset price does not matter, so u and d can be chosen as before, for instance, as in (20.20).

Remark 20.18 *(Risk neutral drift with continuous dividends) With continuous dividends, the risk neutral expected value is* $\mathrm{E}_t^* S_{t+h}/S_t = e^{(y-\delta)h}$, *so the drift is* $(y-\delta)h$ *over the short time interval* h.

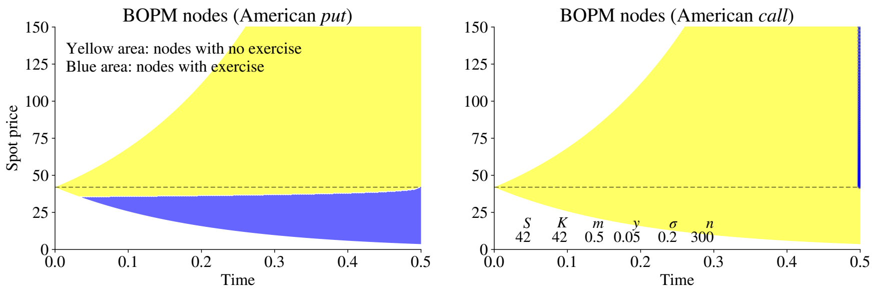


Figure 20.15: Numerical solution of American put and prices

Chapter 21

The Black-Scholes Model

This chapter discusses the Black-Scholes model . It applies to European options, *not* American options. The basic model assumption is that the log price of the underlying asset is normally distributed—which turns out to be the limiting case for the binomial model as the number of time steps increases. The last section performs empirical tests of some of the key implications of the model, with mixed results.

21.1 The Black-Scholes Model

21.1.1 The Basic Black-Scholes Model (No Dividends)

The Black-Scholes (B-S) formula (also known as the Black-Scholes-Merton formula, see Black and Scholes (1973) and Merton (1973)) for the price of a *European call option* on an underlying asset *without dividends* is

$$C = S\Phi\left(d_1\right) - e^{-ym} K\Phi\left(d_2\right), \text{ where} \tag{21.1}$$

$$d_1 = \frac{\ln(S/K) + (y + \sigma^2/2)m}{\sigma\sqrt{m}} \text{ and } d_2 = d_1 - \sigma\sqrt{m}, \tag{21.2}$$

where m is the time to expiration, y the interest rate, S the current asset price, K the strike price and σ^2 the annualised variance of the return on the asset (to be discussed in detail further on). Also, $\Phi(d)$ denotes the probability of $x \leq d$ when x has an $N(0, 1)$ distribution, that is, the value of the standard normal distribution function at d. See an appendix for numerical values. The background (derivation) of the model is discussed below.

Figure 21.1 shows that the Black-Scholes call option price at expiration coincides with the payoff functions (also demonstrated algebraically in an appendix), and how

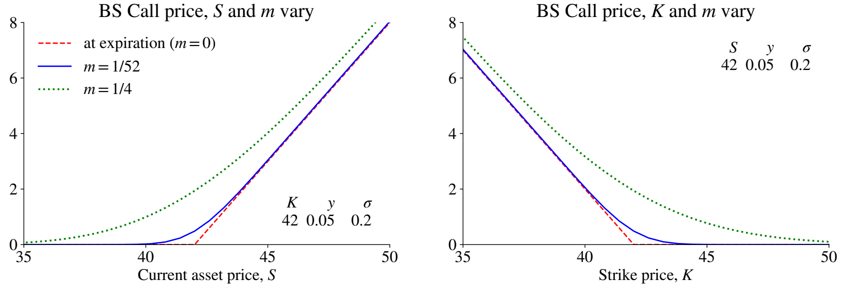


Figure 21.1: Call option price, Black-Scholes model

it moves away as the time to expiration increases. Figure 21.2 provides further details on how the option price is increasing in the current asset price, volatility (σ), time to maturity and the interest rate, but decreasing in the strike price.

Example 21.1 *(Call option price) With* $(S, K, y, m, \sigma) = (42, 42, 0.05, 0.5, 0.2)$*, (21.1)–(21.2) give* $C = 2.893$.

Together, the put-call parity and (21.1) give the pricing formula for a put option

$$P = e^{-ym} K\Phi\left(-d_2\right) - S\Phi\left(-d_1\right), \tag{21.3}$$

where d_1 and d_2 are defined in (21.2).

Proof (of (21.3)) Recall that the put-call parity for an asset without dividends is $C - P = S - e^{-my}K$. Use in (21.1) to get

$$P = S[\Phi\left(d_1\right) - 1] - e^{-ym}K[\Phi\left(d_2\right) - 1].$$

Since $\Phi\left(d\right) + \Phi\left(-d\right) = 1$, this can be written as (21.3). □

21.1.2 The Black-Scholes Model with Dividends

Consider a European option for an underlying asset that pays dividends before expiration. The Black-Scholes formula is then no longer valid. The basic reason is that the current price of the underlying embeds all future dividends, but the option will miss out on those dividends that are paid before the expiration.

To handle this, we could apply the B-S formula to a forward contract on the underlying, expiring on the same day as the option. The point is that the forward also misses out on the dividends. Consider a prepaid forward contract, $e^{-ym}F$,

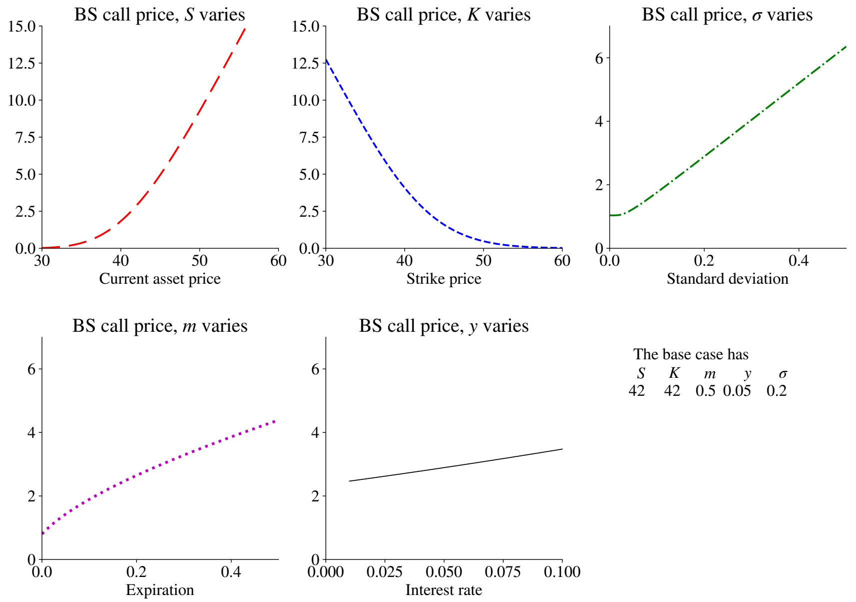


Figure 21.2: Call option price, Black-Scholes model

that is, the amount needed on a money market account today to guarantee that the forward price can be paid at expiration. Let it substitute for the underlying asset price S in the B-S formula (21.1)–(21.2)

$$C = e^{-ym} F \Phi(d_1) - e^{-ym} K \Phi(d_2), \tag{21.4}$$

$$P = e^{-ym} K \Phi(-d_2) - e^{-ym} F \Phi(-d_1), \text{ where} \tag{21.5}$$

$$d_1 = \frac{\ln(F/K) + (\sigma^2/2)m}{\sigma\sqrt{m}} \text{ and } d_2 = d_1 - \sigma\sqrt{m}. \tag{21.6}$$

This is *Black's model* (see Black (1976)) which has many applications. For instance, for an asset with a continuous dividend rate of δ, the forward-spot parity says $F = Se^{(y-\delta)m}$, which can be used to substitute for F. (See Appendix 21.5 for the equations.)

When the asset is a currency (read: foreign money market account) and δ is the foreign interest rate, then this is the Garman and Kohlhagen (1983) model. See McDonald (2014) 15–16 and Hull (2022) 15–17 for more detailed discussions.

Example 21.2 *(Put price) Using the same parameters as in Example 21.1 and*

$\delta = 0$, *we get* $P = 1.856$. *Instead, with* $\delta = 0.05$, *we get* $P = 2.309$.

21.2 Deriving B-S I: Risk Neutral Pricing

21.2.1 Physical and Risk-Neutral Distribution of the Asset Price

The log asset price at expiration ($\ln S_m$) is the current log asset price ($\ln S$) plus all log returns (r_i) between now and then

$$\ln S_m = \ln S + \textstyle\sum_{i=1}^{n} r_i, \text{ where } r_i = \ln S_{ih} - \ln S_{(i-1)h}. \tag{21.7}$$

If the process for the log returns is iid

$$r_i = h(\mu - \sigma^2/2) + \sqrt{h}\sigma\varepsilon_i, \text{ with } \varepsilon_i \sim \text{iid } N(0,1), \tag{21.8}$$

then $\ln S_t$ follows a random walk with drift and the (physical) distribution of $\ln S_m$ is

$$\ln S_m \sim N(\ln S + m\mu - m\sigma^2/2, m\sigma^2). \tag{21.9}$$

This means that the distribution of the asset price (the level, not log), S_m, is lognormal (see Remark 21.3) and that the expected gross return is

$$\mathrm{E}\, S_m/S = e^{\mu m}. \tag{21.10}$$

The difference between the means of $\ln S_m$ and S_m is due to the non-linear transformation.

Remark 21.3 *(Lognormal distribution) If* $Z = e^x$ *and* $x \sim N(\gamma, \omega^2)$, *then the pdf of* Z *is* $\phi[(\ln Z - \gamma)/\omega]/(\omega Z)$, *where* ϕ *is the pdf of an* $N(0,1)$ *variable. This distribution is positively skewed: see the pdf in Figure 21.3 for an example.*

Proof (of (21.10)) It is known that if $x \sim N(\gamma, \omega^2)$, then $\mathrm{E}\, e^x = e^{\gamma + \omega^2/2}$. With γ and ω^2 from (21.9), we get $\mathrm{E}\, S_m = e^{\ln S + \mu m}$. $\square$

To create a risk-neutral distribution, we adjust the physical distribution (21.9), that is, make a "change of measure", so that the risk-neutral expected gross return equals the gross interest rate

$$\mathrm{E}^*\, S_m/S = e^{ym}. \tag{21.11}$$

This is the same requirement for the binomial model (see an earlier chapter). A direct comparison with (21.9)–(21.10) shows that this is done by substituting the

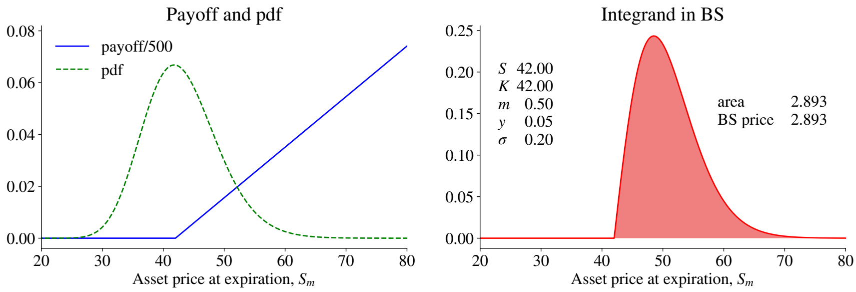


Figure 21.3: Numerical integration to the B-S call price. The payoff is scaled to fit the same figure as the pdf.

interest rate y for μ to get the *risk-neutral distribution* of $\ln S_m$

$$\ln S_m \sim^* N(\ln S + my - m\sigma^2/2, m\sigma^2). \tag{21.12}$$

Notice that the mean is shifted, but the variance is not.

21.2.2 Integrating to Get the Black-Scholes Price

We know that the risk neutral pricing of a European call option is

$$C = e^{-ym} \mathrm{E}^* \max(0, S_m - K), \tag{21.13}$$

which is the present value of the risk-neutral expectation. We can express this as

$$C = e^{-ym} \int_K^\infty \max(0, S_m - K) f^*(S_m) dS_m, \tag{21.14}$$

and where $f^*(S_m)$ is the risk neutral density function of the asset price at expiration. For convenience, the integration starts at K, since integrand is zero at $S_m < K$. Clearly, starting at $-\infty$ would not change the result.

The risk-neutral probability density function $f^*(S_m)$ is the lognormal distribution of the asset price (again, the level, not log) implied by (21.12), as detailed in Remark 21.3. Alternatively, we can do a change-of-variable to $\ln S_m$ and use (21.12). Anyhow, the solution of (21.14) is the Black-Scholes price (21.1)–(21.2). (A proof is in an appendix.) We can also calculate the integral by numerical methods to verify that we get the same value as from the Black-Scholes formula—see Figure 21.3.

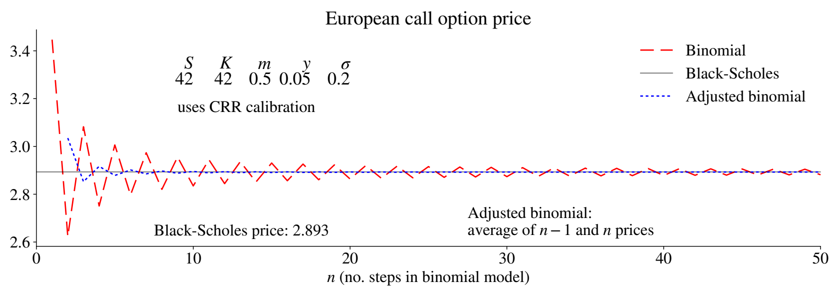


Figure 21.4: Convergence of the binomial price to the Black-Scholes price

Remark 21.4 *(B-S from a stochastic discount factor*) Let M be a stochastic discount factor that satisfies $P = \mathrm{E}\, M x$ for every asset, where P is the asset price and x the payoff of the asset. Then, $C = \mathrm{E}\, M \max(0, S_m - K)$ gives the Black-Scholes formula if $(\ln M, \ln S_m)$ has a joint normal distribution. (See, for instance, Söderlind and Svensson (1997) for a proof.)*

21.3 Deriving B-S II: Convergence of the BOPM

21.3.1 The Main Result

This section demonstrates that the option price from the binomial option pricing model (BOPM) converges to the price from the Black-Scholes model as we take more (but shorter) time steps to reach a fixed time to expiration m. See Figure 21.4 for an illustration of how the option price (from a CRR calibration) converges.

In the binomial option pricing model (BOPM), the risk neutral binomial process for the asset price gives the following binomial process for the *log returns* (changes of the log asset price)

$$r_{t+h} = \ln(S_{t+h}/S_t) = \begin{cases} \ln u & \text{with probability } p \\ \ln d & \text{with probability } 1 - p. \end{cases} \tag{21.15}$$

The parameters u, d and p all depend on the time step length h. With the CRR approach

$$u = e^{\hat{\sigma}\sqrt{h}},\ d = 1/u, \text{ and } p = (e^{yh} - d)/(u - d). \tag{21.16}$$

We demonstrate the convergence of this (as n increases) to the Black-Scholes risk-neutral distribution (21.12) in two steps. First, that the mean and variance

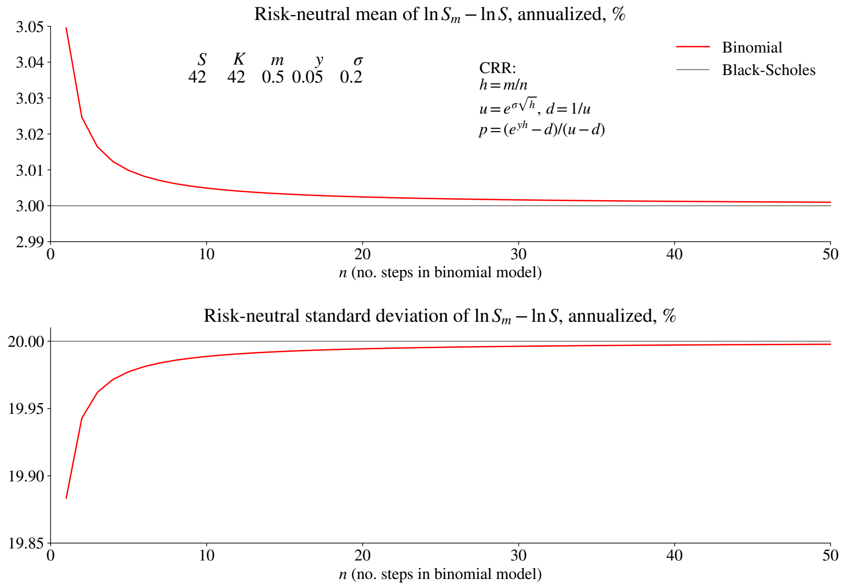


Figure 21.5: Convergence of the binomial mean and variance

converge to the values in (21.12). Second, that the the distribution converges to a normal distribution.

21.3.2 Convergence of the Mean and Variance

This section demonstrates that the mean and variance of the binomial distribution converge to the same values as in the risk neutral distribution of the Black-Scholes model (21.12). See Figure 21.5 for an illustration. The proposition below formalises this. (A proof is in an appendix.)

Proposition 21.5 *(Moments of CRR steps) In the Cox, Ross, and Rubinstein (1979) tree, as $n \to \infty$, but $h = m/n$ we have (since the price changes are independent)*

$$\mathrm{E}\sum_{i=1}^{n} r_i = m(y - \sigma^2/2) \text{ and } \mathrm{Var}(\textstyle\sum_{i=1}^{n} r_i) = m\sigma^2.$$

This is the same as in the risk neutral distribution of the Black-Scholes model.

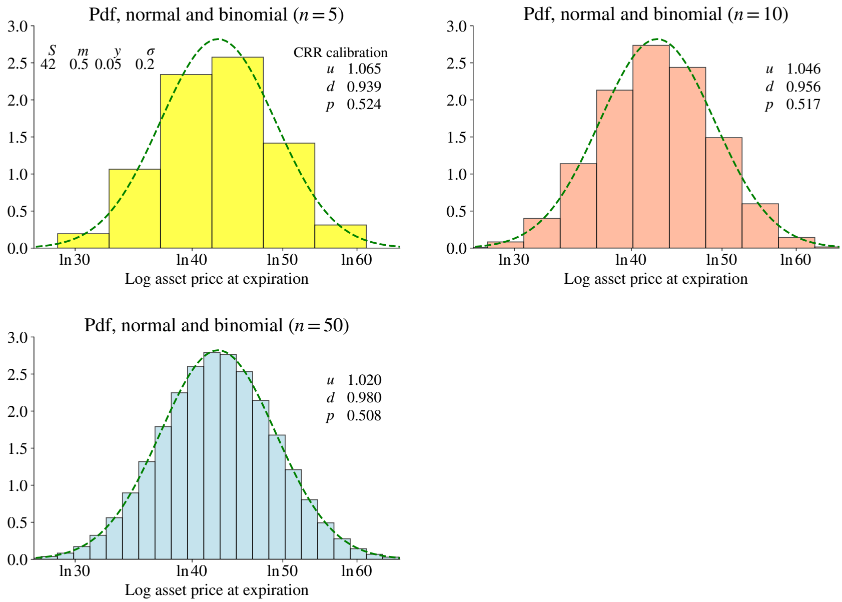


Figure 21.6: Convergence of the binomial model to the Black-Scholes model. The figure shows results for the log asset price. The (risk neutral) distribution from the binomial distribution is scaled so the area of the bars equals one.

21.3.3 The Central Limit Theorem at Work

The Black-Scholes model is based on normally distributed changes of log prices. In the binomial model, the log price changes can only take two values, but the sum of many such changes will converge to a normally distributed variable as the number of time steps increases. This may seem counter intuitive since central limit theorems apply to sample averages times the square root of the sample size, not to sums. However, the rescaling of (u, d, p) as the number of time steps increases, implies that the sum is effectively a (scaled) sample average, so a CLT indeed applies.

See Figure 21.6 for an example of how the distribution converges. Notice that the figure shows the density functions for the *log* asset price (at expiration). Also, the discrete distribution from the binomial model is illustrated by bars centered on the outcome, normalised to have an area of one. The next proposition formalises this, and it applies in the limit to the CRR approach. (A proof is in an appendix.)

Proposition 21.6 *If u, d and p in the binomial process (21.15) are such that the*

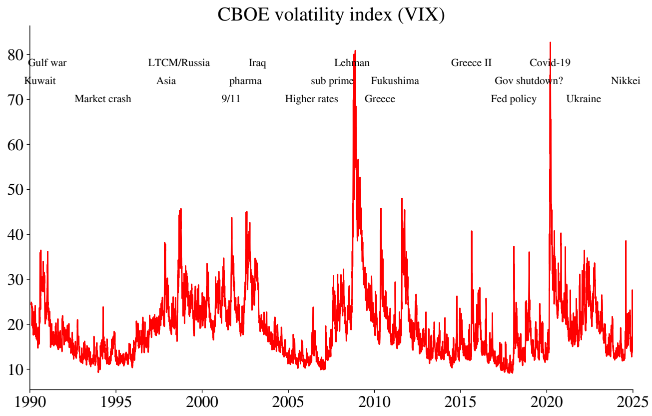


Figure 21.7: CBOE VIX, summary measure of implied volatilities (30 days) on US stock markets

mean and variance of $\ln S_{t+h} - \ln S_t$ *are proportional to* h *(which holds in the limit), then the distribution of* $\Sigma_{i=i}^{n} r_i$ *converges to a normal distribution as the number of time steps* n *increases, keeping the maturity* m *constant (so* $h = m/n$*).*

21.4 Testing the B-S Model

The Black-Scholes formula (21.1)–(21.2) for a European call option contains only one unknown parameter: the standard deviation σ of the distribution of $\ln S_m$. With data on the option price, spot and forward prices, the interest rate, and the strike price, we can solve for σ (see from Figure 21.2 that the option price and the volatility have a monotonic relation).

The σ calculated in this way is called the *implied volatility* and it is often used as an indicator of market uncertainty about the future asset price. It can be thought of as an annualized standard deviation. You can also calculate the implied volatility from a put option, since the put-call parity shows that a call and a put with the same strike price have the same implied volatility.

Empirical Example 21.7 *Figure 21.7 shows how the VIX has changed since it was first introduced. It is an average of implied volatilities of 30-day S&P 500 (close to) at-the-money options. The figure shows considerable variation over time.*

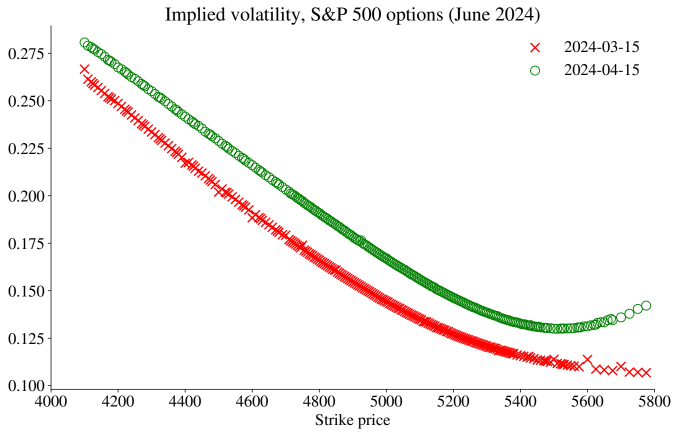


Figure 21.8: Implied volatilities of S&P 500 options, selected dates

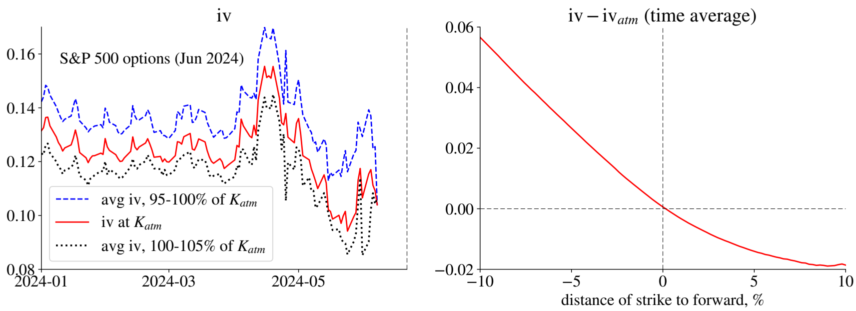


Figure 21.9: Implied volatilities over time

Note that we can solve for one implied volatility for each available strike price. If the Black-Scholes formula is correct, then these volatilities should be the same across strike prices, as the risk-neutral distribution refers to the underlying asset, not the derivative. The volatilities should also be constant over time, at least until the expiration of the option contract, since this is the basic assumption in B-S. Time-varying volatitilities require a more complicated option pricing model.

However, on equity markets, we often find a volatility "smirk," where the volatility is very high for very low strike prices. This is often interpreted as meaning that investors are willing to pay a premium for put options that protect them from a dramatic fall in the stock price. One possible explanation is thus that the distribution has more probability mass than a normal distribution at very low stock prices (negative skewness). Also, there are considerable movements over time.

In contrast, on currency markets, we often find a volatility "smile" (volatility is a U-shaped function of the strike price). One possible explanation is that the (perceived) distribution of the future asset price has relatively more probability mass in the tails ("fat tails") than a normal distribution has. (Recall, Black-Scholes is built on the assumption of a normal distribution.)

Together, this evidence suggests that the B-S model is best seen as an approximation, or perhaps, as a starting point for more advanced option pricing.

Empirical Example 21.8 *See Figures 21.8–21.9 show implied volatility and strike prices for S&P 500 options. The figures show considerable variation across strike prices and also across time.*

21.5 Appendix – Details on the B-S Model*

Remark 21.9 *(Black's model with a continuous dividend rate of δ) Using $F = Se^{(y-\delta)m}$ to substitute for F in (21.4)–(21.6) gives*

$$
\begin{aligned}
C &= e^{-\delta m} S\Phi(d_1) - e^{-ym} K\Phi(d_2) \\
P &= e^{-ym} K\Phi(-d_2) - e^{-\delta m} S\Phi(-d_1), \text{ where} \\
d_1 &= \frac{\ln(S/K) + (y - \delta + \sigma^2/2)m}{\sigma\sqrt{m}} \text{ and } d_2 = d_1 - \sigma\sqrt{m}.
\end{aligned}
$$

Remark 21.10 *(Black-Scholes formula when $\sigma = 0$*) From (21.2)* $\lim_{\sigma\to 0} d_1 = \infty$ *and also* $\lim_{\sigma\to 0} d_2 = \infty$ *if $e^{ym}S \geq K$ and $-\infty$ otherwise. Therefore,* $\lim_{\sigma\to 0}\Phi(d_1) = \lim_{\sigma\to 0}\Phi(d_2) = 1$ *if $e^{ym}S \geq K$ and 0 otherwise. The Black-Scholes call option price at $\sigma = 0$ is therefore* $\max(S - e^{-ym}K, 0)$.

Remark 21.11 *(Black-Scholes formula when $m = 0$*) From (21.2)* $\lim_{m\to 0} d_1 = \lim_{m\to 0} d_2 = \infty$ *if $S > K$ and $-\infty$ otherwise. Therefore,* $\lim_{m\to 0}\Phi(d_1) = \lim_{m\to 0}\Phi(d_2) = 1$ *if $S > K$ and 0 otherwise. (The $S = K$ borderline case gives 0 as both d_1 and d_2 go to $1/2$.) The Black-Scholes call option price at $m = 0$ is therefore* $\max(S - K, 0)$.

Remark 21.12 *(Practical hint: finding the dividend rate*) If you don't know what the dividend rate is, use the forward-spot parity, $F = Se^{(y-\delta)m}$, to calculate it as $\delta = y - \ln(F/S)/m$.*

21.6 Appendix: Proofs*

21.6.1 Integrating the B-S Model

By a change of variable from S_m to $x = \ln S_m$, the integral (21.14) can be written

$$C = e^{-ym}\int_{\ln K}^{\infty} e^{x} g(x)dx - e^{-ym}K\int_{\ln K}^{\infty} g(x)dx, \text{ where } x = \ln S_m, \quad (21.17)$$

and where $g(x)$ is the normal pdf (21.12).

Remark 21.13 *(Properties of normal and lognormal distributions) Let* $x \sim N(\gamma, \omega^2)$ *and define* $a_0 = (a-\gamma)/\omega$. *First, it is well known that* $\Pr(x > a) = \Phi(-a_0)$. *Second, it is also a standard result that* $\mathrm{E}(e^x|x > a) = e^{\gamma+\omega^2/2}\Phi(\omega-a_0)/\Phi(-a_0)$.

To apply this remark, we notice that in (21.14)–(21.12) we have $\gamma = \ln S + ym - \sigma^2 m/2$, $\omega^2 = \sigma^2 m$, and $a = \ln K$, so

$$\gamma + \omega^2/2 = \ln S + ym, \; -a_0 = d_2, \text{ and } \omega - a_0 = d_1. \quad (21.18)$$

Proof ((21.14) gives the Black-Scholes model (21.1)–(21.2)) For the second term in (21.17), (21.18) and Remark 21.13) show that the integral is $\Pr(x > \ln K) = \Phi(d_2)$. The whole second term is thus $e^{-ym}K\Phi(d_2)$. For the first term in (21.17), notice that the integral is $\mathrm{E}(e^x|x > \ln K)$ times $\Pr(x > \ln K)$, so (21.18) and Remark 21.13) show that the integral is $Se^{ym}\Phi(d_1)$. The whole first term is thus $S\Phi(d_1)$. □

21.6.2 Convergence of BOPM to B-S

Proof (*of Proposition 21.5) The convergence of the variance was proved in an earlier chapter. The same chapter showed that $\mathrm{E}\, r_i = (2p-1)\sigma\sqrt{h}$. Let $x = \sqrt{h}$. Then, second order Taylor expansions (around $x = 0$) give $e^{yx^2} \approx 1 + yx^2$, $e^{-\sigma x} \approx 1 - \sigma x + \sigma^2 x^2/2$ and $e^{\sigma x} \approx 1 + \sigma x + \sigma^2 x^2/2$, where the residuals involve x^3 or higher order ("$O(x^3)$"). This gives $e^{yh} - d \approx \sigma x + (y - \sigma^2/2)x^2$ and $u - d \approx 2\sigma x$. In the limit, higher order terms vanish so we can approximate the ratio $p = (e^{yh} - d)/u - d$ as $p \approx [1 + (y - \sigma^2/2)x/\sigma]/2$. This gives $\mathrm{E}\, r_i \approx (y - \sigma^2/2)x^2$. Substitute h for x^2 and multiply the mean and variance by $n = m/h$ (since the returns are iid). □

Proof (*of Proposition 21.6) The binomial model (21.15)–(21.7) means that we, in the limit, can write the return $r_i = h\alpha + \sqrt{h}\sigma\varepsilon_i$, where ε_i is an iid (zero mean, unit variance) random variable. In this formulation, $\mathrm{E}\, r_i = h\alpha$ and $\mathrm{Var}(r_i) = h\sigma^2$, so both moments are proportional to h. Write (21.7) as $\Sigma_{i=i}^{n} r_i = nh\alpha + \sqrt{h}\sigma\Sigma_{i=i}^{n}\varepsilon_i +$. Since $h = m/n$, this can be written $\Sigma_{i=i}^{n} r_i = \alpha m + \sqrt{m}\sigma(\sqrt{n}\frac{1}{n}\Sigma_{i=i}^{n}\varepsilon_i)$. The term in parenthesis is $\sqrt{n}$ times the sample average of an iid random variable (ε_i) with $\mathrm{E}\,\varepsilon_i = 0$ and $\mathrm{Var}(\varepsilon_i) = 1 < \infty$. We can therefore apply the (Lindeberg-Lévy) central limit theorem to show that () $\overset{d}{\to} N(0,1)$. The first term (αm) is just a constant. Together, we get that $\Sigma_{i=1}^{n} r_i \overset{d}{\to} N(\alpha m, \sigma^2 m)$. $\square$

21.7 Appendix – Statistical Tables

	0.0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.0	0.001	0.001	0.001	0.001	0.002	0.002	0.002	0.002	0.002	0.002
-2.9	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
-2.8	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
-2.7	0.003	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.005
-2.6	0.005	0.005	0.005	0.005	0.005	0.005	0.006	0.006	0.006	0.006
-2.5	0.006	0.006	0.007	0.007	0.007	0.007	0.007	0.008	0.008	0.008
-2.4	0.008	0.008	0.009	0.009	0.009	0.009	0.010	0.010	0.010	0.010
-2.3	0.011	0.011	0.011	0.012	0.012	0.012	0.013	0.013	0.013	0.014
-2.2	0.014	0.014	0.015	0.015	0.015	0.016	0.016	0.017	0.017	0.017
-2.1	0.018	0.018	0.019	0.019	0.020	0.020	0.021	0.021	0.022	0.022
-2.0	0.023	0.023	0.024	0.024	0.025	0.026	0.026	0.027	0.027	0.028
-1.9	0.029	0.029	0.030	0.031	0.031	0.032	0.033	0.034	0.034	0.035
-1.8	0.036	0.037	0.038	0.038	0.039	0.040	0.041	0.042	0.043	0.044
-1.7	0.045	0.046	0.046	0.047	0.048	0.049	0.051	0.052	0.053	0.054
-1.6	0.055	0.056	0.057	0.058	0.059	0.061	0.062	0.063	0.064	0.066
-1.5	0.067	0.068	0.069	0.071	0.072	0.074	0.075	0.076	0.078	0.079
-1.4	0.081	0.082	0.084	0.085	0.087	0.089	0.090	0.092	0.093	0.095
-1.3	0.097	0.099	0.100	0.102	0.104	0.106	0.107	0.109	0.111	0.113
-1.2	0.115	0.117	0.119	0.121	0.123	0.125	0.127	0.129	0.131	0.133
-1.1	0.136	0.138	0.140	0.142	0.145	0.147	0.149	0.152	0.154	0.156
-1.0	0.159	0.161	0.164	0.166	0.169	0.171	0.174	0.176	0.179	0.181
-0.9	0.184	0.187	0.189	0.192	0.195	0.198	0.200	0.203	0.206	0.209
-0.8	0.212	0.215	0.218	0.221	0.224	0.227	0.230	0.233	0.236	0.239
-0.7	0.242	0.245	0.248	0.251	0.255	0.258	0.261	0.264	0.268	0.271
-0.6	0.274	0.278	0.281	0.284	0.288	0.291	0.295	0.298	0.302	0.305
-0.5	0.309	0.312	0.316	0.319	0.323	0.326	0.330	0.334	0.337	0.341
-0.4	0.345	0.348	0.352	0.356	0.359	0.363	0.367	0.371	0.374	0.378
-0.3	0.382	0.386	0.390	0.394	0.397	0.401	0.405	0.409	0.413	0.417
-0.2	0.421	0.425	0.429	0.433	0.436	0.440	0.444	0.448	0.452	0.456
-0.1	0.460	0.464	0.468	0.472	0.476	0.480	0.484	0.488	0.492	0.496

Table 21.1: Values of the standard normal cumulative distribution function at x where x is the sum of the values in the first column and the first row.

	0.0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500	0.504	0.508	0.512	0.516	0.520	0.524	0.528	0.532	0.536
0.1	0.540	0.544	0.548	0.552	0.556	0.560	0.564	0.567	0.571	0.575
0.2	0.579	0.583	0.587	0.591	0.595	0.599	0.603	0.606	0.610	0.614
0.3	0.618	0.622	0.626	0.629	0.633	0.637	0.641	0.644	0.648	0.652
0.4	0.655	0.659	0.663	0.666	0.670	0.674	0.677	0.681	0.684	0.688
0.5	0.691	0.695	0.698	0.702	0.705	0.709	0.712	0.716	0.719	0.722
0.6	0.726	0.729	0.732	0.736	0.739	0.742	0.745	0.749	0.752	0.755
0.7	0.758	0.761	0.764	0.767	0.770	0.773	0.776	0.779	0.782	0.785
0.8	0.788	0.791	0.794	0.797	0.800	0.802	0.805	0.808	0.811	0.813
0.9	0.816	0.819	0.821	0.824	0.826	0.829	0.831	0.834	0.836	0.839
1.0	0.841	0.844	0.846	0.848	0.851	0.853	0.855	0.858	0.860	0.862
1.1	0.864	0.867	0.869	0.871	0.873	0.875	0.877	0.879	0.881	0.883
1.2	0.885	0.887	0.889	0.891	0.893	0.894	0.896	0.898	0.900	0.901
1.3	0.903	0.905	0.907	0.908	0.910	0.911	0.913	0.915	0.916	0.918
1.4	0.919	0.921	0.922	0.924	0.925	0.926	0.928	0.929	0.931	0.932
1.5	0.933	0.934	0.936	0.937	0.938	0.939	0.941	0.942	0.943	0.944
1.6	0.945	0.946	0.947	0.948	0.949	0.951	0.952	0.953	0.954	0.954
1.7	0.955	0.956	0.957	0.958	0.959	0.960	0.961	0.962	0.962	0.963
1.8	0.964	0.965	0.966	0.966	0.967	0.968	0.969	0.969	0.970	0.971
1.9	0.971	0.972	0.973	0.973	0.974	0.974	0.975	0.976	0.976	0.977
2.0	0.977	0.978	0.978	0.979	0.979	0.980	0.980	0.981	0.981	0.982
2.1	0.982	0.983	0.983	0.983	0.984	0.984	0.985	0.985	0.985	0.986
2.2	0.986	0.986	0.987	0.987	0.987	0.988	0.988	0.988	0.989	0.989
2.3	0.989	0.990	0.990	0.990	0.990	0.991	0.991	0.991	0.991	0.992
2.4	0.992	0.992	0.992	0.992	0.993	0.993	0.993	0.993	0.993	0.994
2.5	0.994	0.994	0.994	0.994	0.994	0.995	0.995	0.995	0.995	0.995
2.6	0.995	0.995	0.996	0.996	0.996	0.996	0.996	0.996	0.996	0.996
2.7	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.997
2.8	0.997	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998
2.9	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.999	0.999	0.999

Table 21.2: Values of the standard normal cumulative distribution function at x where x is the sum of the values in the first column and the first row.

Chapter 22

Hedging Options

This chapter shows how an option portfolio can be hedged. The setting is that we have written (sold, issued) an option portfolio, but we do not want to carry the risk. A first-order Taylor approximation (of the option price in term of the underlying price) leads to the delta hedging approach, which is illustred with an empirical example. A later section discusses how the hedging can be extended to capture higher-order moments and volatility changes.

22.1 Hedging an Option

Remark 22.1 *(dX notation) Warning: this section uses dX to indicate a change in variable X, mostly since Δ has another, and well established, interpretation in the option literature.*

A first order approximation suggests that the change (here indicated by d) in the option portfolio value (denoted L) due to a change in the underlying price is

$$dL \approx \Delta dS, \tag{22.1}$$

where we use Δ to denote the derivative $\partial L/\partial S$ (as is standard in the option literature). In particular, we use Δ_c for call options and Δ_p for put options.

When the option portfolio consists of a call option only, then $L = C$ (where C is the call option price) and the derivative is positive, see Figure 22.1.

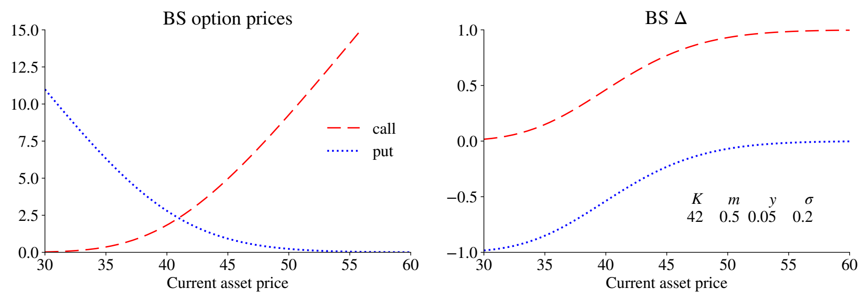


Figure 22.1: Option prices and deltas from the Black-Scholes model

22.2 An Approximate Hedge

22.2.1 Basic Setup

Consider a portfolio which is long v units of the underlying asset (the hedging portfolio) and short one option portfolio (with value L). The value of the overall position is

$$V = vS + M - L, \tag{22.2}$$

where M is a money market account. The idea is to find v so that vS and L are equally sensitive to changes in S. For instance, if $L = 3C$, then we have issued three call options, that is, are short three calls options. A long option position can be handled by $L < 0$, for instance, $L = -2P$.

For now, we focus on movements in the price of the underlying, disregarding, for instance, movements in volatility. Use (22.1) to approximate the change (indicated by d) of the value of the overall portfolio as

$$\begin{aligned} dV &\approx vdS - \Delta dS \\ &\approx 0 \text{ if } v = \Delta. \end{aligned} \tag{22.3}$$

This approach makes the overall portfolio *delta neutral*, $\partial V/\partial S = 0$, and is therefore called a *delta hedge*.

See Figure 22.1 for how the Black-Scholes option price and its derivative depend on the underlying asset price. Note that the derivative is positive for a call option and negative for a put option.

Example 22.2 *(Delta hedging a call or a put) Suppose $\Delta_c = 0.6$ and $\Delta_p = -0.4$.*

If $L = C$, then $v = 0.6$, but with $L = P$ we get $v = -0.4$. For $L = 3C - 2P$, we have $v = 3 \cdot 0.6 + 2 \cdot 0.4 = 2.6$.

The Δ will change over time, necessitating portfolio rebalancing. In practice, the overall portfolio includes a position in a short-term money market account to make the initial portfolio value zero. This money market position is typically negative for a call option, which means that we finance the purchase of the underlying asset with the proceeds from selling the option and from borrowing.

22.2.2 Deltas from the Black-Scholes Model

The following remark gives details of the Δ in the Black-Scholes model. (The other derivatives are presented in an appendix.)

Remark 22.3 *(Deltas in Black-Scholes) The Black-Scholes formula for a European call option on an asset paying continuous dividends (δ) is*

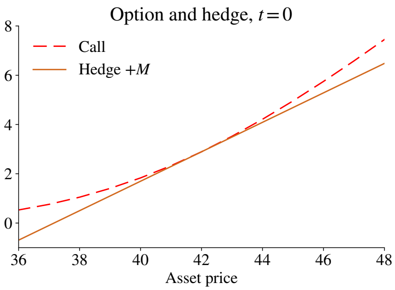
$$C = e^{-\delta m} S \Phi(d_1) - e^{-ym} K \Phi(d_2), \text{ where}$$
$$d_1 = \frac{\ln(S/K) + (y - \delta + \sigma^2/2)m}{\sigma\sqrt{m}} \text{ and } d_2 = d_1 - \sigma\sqrt{m}.$$

(Warning: d_1 and d_2 indicate the usual terms in the Black-Scholes formula. Do not confuse with the d used to indicate a change.) The derivatives of the call and put prices are

$$\Delta_c = \frac{\partial C}{\partial S} = e^{-\delta m}\Phi(d_1)$$
$$\Delta_p = \frac{\partial P}{\partial S} = \Delta_c - e^{-\delta m}.$$

The result for the put follows from the put-call parity which says $P = C - Se^{-\delta m} + e^{-my}K$. It is also useful to notice that the derivative with respect to a forward price ($F = Se^{(y-\delta)m}$) is $\partial C/\partial F = e^{-ym}\Phi(d_1)$, where d_1 is as above (or, equivalently, $d_1 = [\ln(F/K) + (\sigma^2/2)m]/(\sigma\sqrt{m})$).

See Figure 22.1 for an illustration of how the Black-Scholes Δ depends on the underlying price . In particular, notice that $0 \leq \Delta_c \leq e^{-\delta m}$ for a call and $-e^{-\delta m} \leq \Delta_p \leq 0$ for a put. In both cases, Δ is increasing with the price of the underlying asset. Intuitively, an option that is deep out of the money will not be very sensitive to the asset price—since the chance of exercising is low. Conversely,



Portfolio value:
$V = \Delta S + M - C$

Call option: from BS with

K	m	y	σ
42	0.5	0.05	0.2

Underlying: 42 on day 0,
but 43 on day 1

	S	Δ	ΔS	M	$-C$	V
Day 0	42.00	0.60	25.10	-22.21	-2.89	0.00
Day 1 (m-to-m)	43.00	0.60	25.70	-22.22	-3.51	-0.02
Day 1 (rebal)	43.00	0.66	28.40	-24.91	-3.51	-0.02

Figure 22.2: Delta hedging over time

a option that is deep in the money moves almost one for one in same direction if it is a call option and in the opposite direction if it is a put option. See See Hull (2022) 13 and McDonald (2014) 15–16 for more detailed treatments.

Example 22.4 *(Δ_c and Δ_p) With* $(S, K, m, y, \sigma) = (42, 42, 0.5, 0.05, 0.2)$ *and* $\delta = 0$*, we have* $\Delta_c \approx 0.60$ *and* $\Delta_p \approx -0.40$*. The difference is equal to one (since* $\delta = 0$*).*

Example 22.5 *(Delta hedging of a call option) Using the same parameters as in Example 22.4 and* $\delta = 0$*, Figure 22.2 illustrates the initial positions (day 0), and two snap shots of the day after (day 1: after marking to market, day 1: after rebalancing). On day 0, the overall portfolio includes* $\Delta = 0.6$ *of the underlying asset (at a value of* $0.6 \times 42 = 25.10$*),* -1 *of the call option (at the value* -2.89*) and the balance on a money market account (*$-25.10 + 2.89 = -22.21$*) so the total portfolio is worth zero. This clearly means that the investor has borrowed. As S changes on day 1, the overall portfolio is almost stable. However, the* Δ *has changed to 0.66 so the portfolio needs to be rebalanced.*

Empirical Example 22.6 *Figure 22.3 illustrates the hedging of a particular S&P 500 option over 6 months. The overall portfolio is much more stable than the option itself, but there are still some movements left. This suggests that the hedging strategy is largely effective, but remains imperfect.*

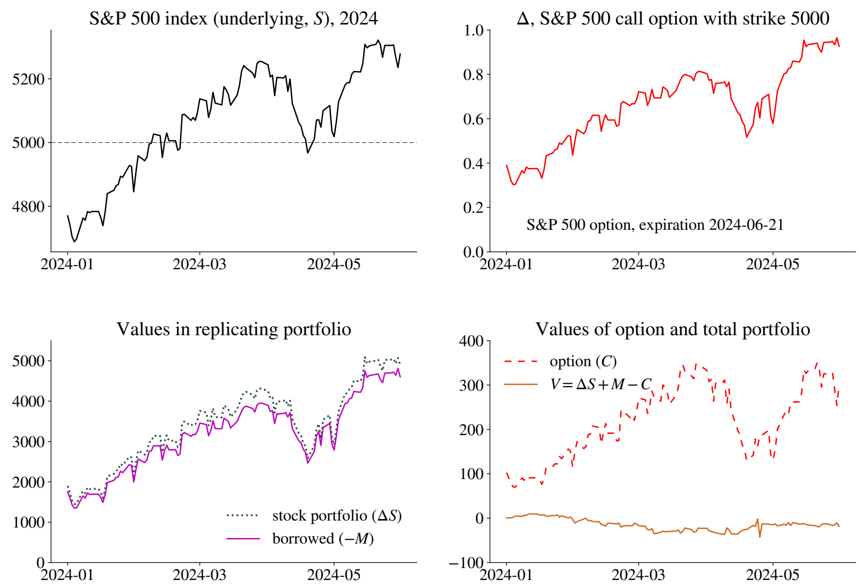


Figure 22.3: Delta hedging an S&P 500 call option

Remark 22.7 *(Hedging with a forward contract*) Consider using a forward contract as hedging instrument. Recall that* $W_t = e^{-ym}(F_t - F_\tau)$ *is the value of an old forward contract (written in* $\tau < t$*). The hedge portfolio is* $V = vW + M - C$*. This portfolio is almost stable if* $v = e^{ym}\partial C/\partial F$ *(see Remark 22.3 for the derivative). To see this, notice that* $dV = v\,dW - dC \approx ve^{-ym}dF - \frac{\partial C_t}{\partial F}dF$*.*

22.2.3 Deltas from Other Models

The Δ (the derivative in (22.3)) could also be computed from other option pricing models, for instance, the binomial model.

The basic approach is straightforward: consider two different values of the underlying asset (S_a and S_b), use the model to compute the option price at each of them (get $L(S_a)$ and $L(S_b)$) and approximate the derivative with a finite difference ratio: $[L(S_a) - L(S_b)]/(S_a - S_b)$. Clearly, this crude approach can be improved by using other numerical methods for approximating derivatives.

In particular, the binomial model has the advantage that it allows us to handle also American-style options. See Figure 22.4 and notice that the delta of an American put tends to be more negative than for a European put, especially at low prices of

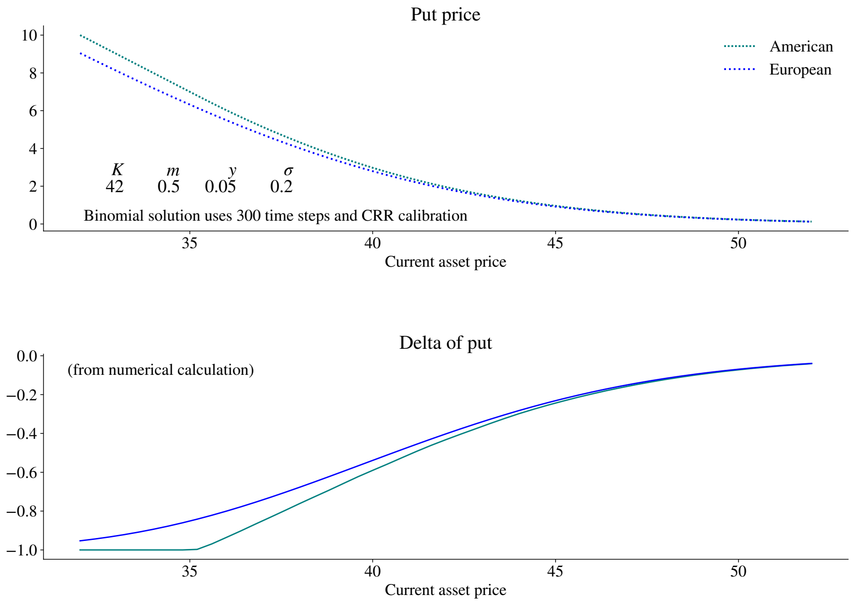


Figure 22.4: The deltas of American and European puts

the underlying.

22.3 Higher-Order Hedging*

Delta hedging can be imprecise for several reasons, for instance, for option portfolios with payoffs that are very non-linear in the underlying, or if the volatility changes markedly. For that reason, this section briefly discusses more sophisticated hedging strategies.

22.3.1 Delta-Gamma Hedging*

A delta hedge is likely to work poorly when there are large prices changes of the underlying, or when the option portfolio is a highly non-linear function of the underlying asset. As an example of the latter, Figure 22.5 illustrates the price of a straddle (according to Black-Scholes). If the current price of the underlying is close to the strike price, then the (first-order) derivative is zero, but the straddle gains value as soon as the underlying price moves in either direction.

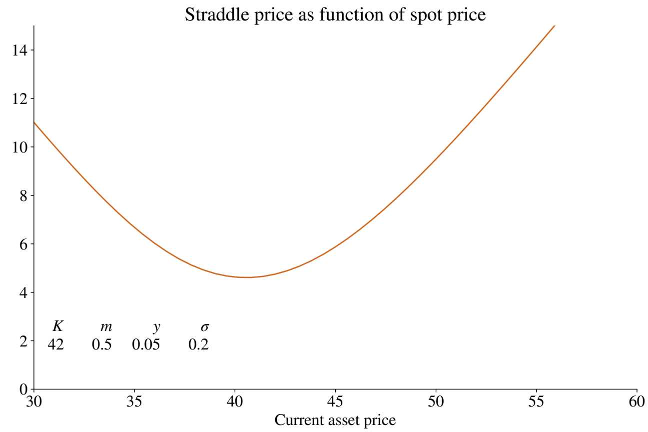


Figure 22.5: Price of a straddle (Black-Scholes)

We can improve the precision by using a second-order Taylor approximation of the option portfolio value

$$dL \approx \Delta dS + \frac{1}{2}\Gamma (dS)^2 \text{, where } \Gamma = \frac{\partial^2 L}{\partial S^2}. \tag{22.4}$$

The Γ (upper case gamma) of the Black-Scholes model is presented in an appendix.

To hedge, consider a portfolio with v of the underlying asset, w of another option (or another asset) with a price denoted L^* and short one option portfolio (with price L)

$$V = vS + wL^* - L. \tag{22.5}$$

We get $dV \approx 0$ by setting

$$w = \Gamma/\Gamma^*, \text{ and} \tag{22.6}$$

$$v = \Delta - w\Delta^*, \tag{22.7}$$

where Δ^* and Γ^* are the delta and gamma of L^*.

Proof (of (22.6)–(22.7)) A second-order Taylor approximation like (22.4) of the value of portfolio V gives

$$\begin{aligned} dV &\approx vdS + w[\Delta^* dS + \Gamma^* (dS)^2/2] - [\Delta dS + \frac{1}{2}\Gamma (dS)^2] \\ &\approx (v + w\Delta^* - \Delta)dS + (w\Gamma^* - \Gamma)(dS)^2/2. \end{aligned}$$

Using the values (w, v) in (22.6)–(22.7) makes this zero. □

Example 22.8 *(Delta-gamma hedging) Suppose* $(\Delta, \Gamma) = (0.5, 0.07)$ *and* $(\Delta^*, \Gamma^*) = (0.3, 0.03)$*, which requires* $w = 2.33$ *and* $v = -0.2$*. Clearly, this is quite different from a delta hedge (which has* $v = 0.5$ *and* $w = 0$*). Here, the lower sensitivity (gamma) of the second option to the quadratic term means that the hedge portfolio includes a lot of the second option. As a consequence, it becomes overexposed to the linear term, which is compensated for by a short position in the underlying asset.*

22.3.2 Delta-Vega Hedging*

The volatility of financial markets fluctuates over time. To account for this, a first-order Taylor approximation of the call option price in terms of *both* the underlying and volatility is

$$dL \approx \Delta dS + \frac{\partial L}{\partial \sigma} d\sigma, \tag{22.8}$$

where $\partial L/\partial \sigma$ is the "vega" of the option portfolio (presented in an appendix). Notice that the Black-Scholes model is inconsistent with time-variation in volatility—so it can only be used as an approximation.

Consider hedging by holding the following portfolio

$$V = vS + wL^* - L, \tag{22.9}$$

where L^* is the price of some other option (or asset). We get $dV \approx 0$ by setting

$$w = \frac{\partial L}{\partial \sigma} / \frac{\partial L^*}{\partial \sigma}, \text{ and} \tag{22.10}$$

$$v = \Delta - w\Delta^*, \tag{22.11}$$

where Δ^* and $\partial L^*/\partial \sigma$ are the delta and vega of L^*. For instance, if the L^* asset is directly linked to VIX, then $\Delta^* = 0$ and $\partial L^*/\partial \sigma = 1$.

Proof (of (22.10)–(22.11)) A first-order Taylor approximation like (22.8) of the value of portfolio V gives

$$\begin{aligned} dV &= vdS + w(\Delta^* dS + \frac{\partial L^*}{\partial \sigma} d\sigma) - (\Delta dS + \frac{\partial L}{\partial \sigma} d\sigma) \\ &= (v + w\Delta^* - \Delta)dS + (w\frac{\partial L^*}{\partial \sigma} - \frac{\partial L}{\partial \sigma})d\sigma. \end{aligned}$$

Using the values (w, v) in (22.10)–(22.11) makes this zero. □

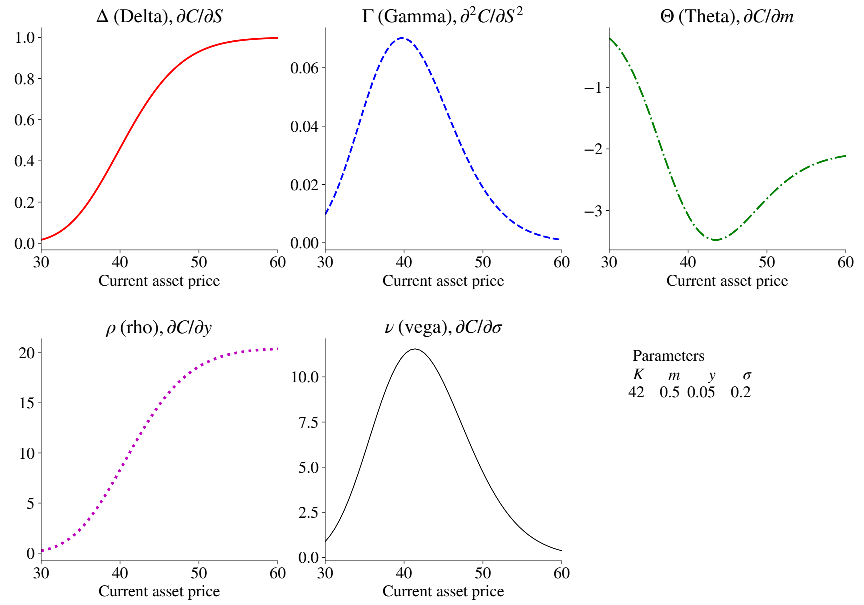


Figure 22.6: The Greeks in the Black-Scholes model as a function of the asset price

22.4 Appendix – More Greeks*

Remark 22.9 *(The "Greeks") In addition to the results in Remark 22.3, we have*

$$\Gamma = \frac{\partial^2 C}{\partial S^2} = \frac{e^{-\delta m}\phi\,(d_1)}{S\sigma\sqrt{m}}$$

$$\theta = \frac{\partial C}{\partial t} = -\frac{\partial C}{\partial m} = \delta S e^{-\delta m}\Phi(d_1) - yKe^{-ym}\Phi(d_2) - \frac{1}{2\sqrt{m}}e^{-\delta m}S\phi(d_1)\sigma$$

$$(vega) = \frac{\partial C}{\partial \sigma} = Se^{-\delta m}\phi\,(d_1)\sqrt{m}$$

$$\rho = \frac{\partial C}{\partial y} = mKe^{-ym}\Phi(d_2).$$

See Figure 22.6.

Chapter 23

Appendices

23.1 Appendix – Matrix Algebra*

This appendix introduces fundamental concepts of matrix algebra.

The discussion will use the vectors and matrices

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}, A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \text{ and } B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix},$$

where the elements (for instance, x_1 and A_{12}) are numbers. In most cases, we think of a vector as a matrix with one column (sometimes referred to as a column vector). In contrast, a row vector is a matrix with one row.

Example 23.1 *(Vector and matrix)*

$$x = \begin{bmatrix} 10 \\ 11 \end{bmatrix} \text{ and } A = \begin{bmatrix} 1 & 3 \\ 3 & 4 \end{bmatrix}.$$

23.1.1 Matrix and Scalar Addition and Multiplication

Multiplying a matrix by a scalar c means multiplying each element by the scalar

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} c = \begin{bmatrix} A_{11}c & A_{12}c \\ A_{21}c & A_{22}c \end{bmatrix}.$$

Example 23.2 *(Matrix × scalar)*

$$\begin{bmatrix} 1 & 3 \\ 3 & 4 \end{bmatrix} 10 = \begin{bmatrix} 10 & 30 \\ 30 & 40 \end{bmatrix}.$$

Adding/subtracting a scalar to each element of a matrix is done by

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} + cJ = \begin{bmatrix} A_{11}+c & A_{12}+c \\ A_{21}+c & A_{22}+c \end{bmatrix},$$

where J is a matrix (of the same size as A) filled with ones. This is sometimes written $A + c$, although that notation is not universally liked. In some applications, $\mathbf{1}_n$ (or just $\mathbf{1}$) is used to represent a vector of n ones.

Example 23.3 *(Matrix ± scalar)*

$$\begin{bmatrix} 10 \\ 11 \end{bmatrix} - 10 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$
$$\begin{bmatrix} 1 & 3 \\ 3 & 4 \end{bmatrix} + 10 \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 11 & 13 \\ 13 & 14 \end{bmatrix}.$$

23.1.2 Adding and Multiplying: Two Matrices

Matrix *addition* (or subtraction) of matrices of the same size is element by element

$$A + B = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} + \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} A_{11}+B_{11} & A_{12}+B_{12} \\ A_{21}+B_{21} & A_{22}+B_{22} \end{bmatrix}.$$

Example 23.4 *(Matrix addition and subtraction)*

$$\begin{bmatrix} 10 \\ 11 \end{bmatrix} - \begin{bmatrix} 2 \\ 5 \end{bmatrix} = \begin{bmatrix} 8 \\ 6 \end{bmatrix}$$
$$\begin{bmatrix} 1 & 3 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 2 \\ 3 & -2 \end{bmatrix} = \begin{bmatrix} 2 & 5 \\ 6 & 2 \end{bmatrix}$$

Matrix *multiplication* requires the two matrices to be conformable: with AB where A is $m \times n$, B must be $n \times p$. Element ij of the result (which is $m \times p$) is the multiplication of the ith row of the first matrix with the jth column of the second matrix

$$AB = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix}.$$

As a special case, multiplying a matrix A with a vector z gives a new vector

$$Az = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} A_{11}z_1 + A_{12}z_2 \\ A_{21}z_1 + A_{22}z_2 \end{bmatrix}.$$

Example 23.5 *(Matrix multiplication)*

$$\begin{bmatrix} 1 & 3 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & -2 \end{bmatrix} = \begin{bmatrix} 10 & -4 \\ 15 & -2 \end{bmatrix}$$
$$\begin{bmatrix} 1 & 3 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 5 \end{bmatrix} = \begin{bmatrix} 17 \\ 26 \end{bmatrix}$$

23.1.3 Transpose

Transposing a column vector gives a row vector. Similarly, transposing a matrix is like flipping it around the main diagonal so the former columns become the new rows

$$A' = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}' = \begin{bmatrix} A_{11} & A_{21} \\ A_{12} & A_{22} \end{bmatrix}.$$

Example 23.6 *(Matrix transpose)*

$$\begin{bmatrix} 10 \\ 11 \end{bmatrix}' = \begin{bmatrix} 10 & 11 \end{bmatrix}$$
$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}' = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

23.1.4 Inner and Outer Products, Quadratic Forms

For two vectors x and z, the product $x'z$ is called the *inner product* (a scalar)

$$x'z = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = x_1z_1 + x_2z_2,$$

and xz' the *outer product* (a matrix)

$$xz' = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \begin{bmatrix} z_1 & z_2 \end{bmatrix} = \begin{bmatrix} x_1z_1 & x_1z_2 \\ x_2z_1 & x_2z_2 \end{bmatrix}.$$

(Notice that xz does not work for two vectors.)

Example 23.7 *(Inner and outer products)*

$$\begin{bmatrix}10\\11\end{bmatrix}'\begin{bmatrix}2\\5\end{bmatrix}=\begin{bmatrix}10 & 11\end{bmatrix}\begin{bmatrix}2\\5\end{bmatrix}=75$$
$$\begin{bmatrix}10\\11\end{bmatrix}\begin{bmatrix}2\\5\end{bmatrix}'=\begin{bmatrix}10\\11\end{bmatrix}\begin{bmatrix}2 & 5\end{bmatrix}=\begin{bmatrix}20 & 50\\22 & 55\end{bmatrix}$$

If z is a vector and A a square matrix, then the product $z'Az$ is a quadratic form *(a scalar)*

$$z'Az=\begin{bmatrix}z_1\\z_2\end{bmatrix}'\begin{bmatrix}A_{11} & A_{12}\\A_{21} & A_{22}\end{bmatrix}\begin{bmatrix}z_1\\z_2\end{bmatrix}=z_1A_{11}z_1+z_1A_{12}z_2+z_2A_{21}z_1+z_2A_{22}z_2.$$

Example 23.8 *(Quadratic form)*

$$\begin{bmatrix}10\\11\end{bmatrix}'\begin{bmatrix}1 & 3\\3 & 4\end{bmatrix}\begin{bmatrix}10\\11\end{bmatrix}=1244$$

23.1.5 Kronecker Product

Let $\otimes$ represent the Kronecker product, that is, if A and B are matrices, then

$$A\otimes B=\begin{bmatrix}a_{11}B & \cdots & a_{1n}B\\ \vdots & & \vdots\\ a_{m1}B & \cdots & a_{mn}B\end{bmatrix}.$$

Example 23.9 *(Kronecker product)*

$$\begin{bmatrix}1 & 3\\2 & 4\end{bmatrix}\otimes\begin{bmatrix}10 & 11\end{bmatrix}=\begin{bmatrix}10 & 11 & 30 & 33\\20 & 22 & 40 & 44\end{bmatrix}.$$

23.1.6 Matrix Inverse

A matrix *inverse* is the closest we get to "dividing" by a square matrix. The inverse of a matrix A, denoted A^{-1}, is such that

$$AA^{-1}=I \text{ and } A^{-1}A=I,$$

where I is the *identity matrix* (ones along the diagonal, and zeros elsewhere). The matrix inverse is useful for solving systems of linear equations, $y=Ax$ as

$x = A^{-1}y$. Notice that not every square matrix is invertible, in particular not if some rows (or columns) are linear combinations of the other rows (columns).

For a 2×2 matrix we have

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \frac{1}{A_{11}A_{22} - A_{12}A_{21}} \begin{bmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{bmatrix}.$$

Example 23.10 *(Matrix inverse)*

$$\begin{bmatrix} -0.8 & 0.6 \\ 0.6 & -0.2 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \text{ so}$$

$$\begin{bmatrix} 1 & 3 \\ 3 & 4 \end{bmatrix}^{-1} = \begin{bmatrix} -0.8 & 0.6 \\ 0.6 & -0.2 \end{bmatrix}.$$

23.1.7 Solving Systems of Linear Equations

If A is $n \times n$ and invertible and b and y are $n \times 1$ vectors, then we can solve

$$Ab = y \text{ as } b = A^{-1}y.$$

This solution is unique. In numerical applications, this system can often be solved (faster and with better precision) without the explicit matrix inverse.

Example 23.11 *(Solving a system of linear equations)*

$$\begin{bmatrix} 1 & 3 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 10 \\ 11 \end{bmatrix}, \text{ gives}$$

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 3 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 10 \\ 11 \end{bmatrix} = \begin{bmatrix} -1.4 \\ 3.8 \end{bmatrix}.$$

Using this in the first equation verifies that we indeed get the right result.

23.1.8 Derivatives of Matrix Expressions

Let z and x be $n \times 1$ vectors. The *derivative of the inner product* is $\partial(x'z)/\partial x = z$.

Example 23.12 *(Derivative of an inner product) With $n = 2$*

$$x'z = x_1 z_1 + x_2 z_2, \text{ so } \partial(x'z)/\partial x = \begin{bmatrix} \partial(z_1 x_1 + z_2 x_2)/\partial x_1 \\ \partial(z_1 x_1 + z_2 x_2)/\partial x_2 \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}.$$

Let x be $n \times 1$ and A an $n \times n$ matrix. The *derivative of the quadratic form* is $\partial(x'Ax)/\partial x = (A + A')x$. (In case A is symmetric, the derivative is $2Ax$.)

Example 23.13 *(Derivative of a symmetric quadratic form) With $n = 2$, the symmetric quadratic form is*

$$x'Ax = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{12} & A_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1^2 A_{11} + x_2^2 A_{22} + 2x_1 x_2 A_{12}.$$

The derivatives with respect to x_1 and x_2 are

$$\partial(x'Ax)/\partial x_1 = 2x_1 A_{11} + 2x_2 A_{12} \text{ and } \partial(x'Ax)/\partial x_2 = 2x_2 A_{22} + 2x_1 A_{12}, \text{ or}$$
$$\partial(x'Ax)/\partial x = \begin{bmatrix} \partial(x'Ax)/\partial x_1 \\ \partial(x'Ax)/\partial x_2 \end{bmatrix} = 2 \begin{bmatrix} A_{11} & A_{12} \\ A_{12} & A_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

23.2 Appendix – Statistics*

This appendix first summarizes some mathematical statistics required for the financial models discussed in the text. Towards the end, it briefly addresses topics in estimation and testing.

23.2.1 The Distribution of a Random Variable

The distribution of a random variable x represents the probabilities of its possible values. See Figure 23.1 for illustrations of the (discrete) distribution of a binomial variable and of several different (continuous) normal distributions, often denoted $N(\mu, \sigma^2)$ to indicate the mean and variance.

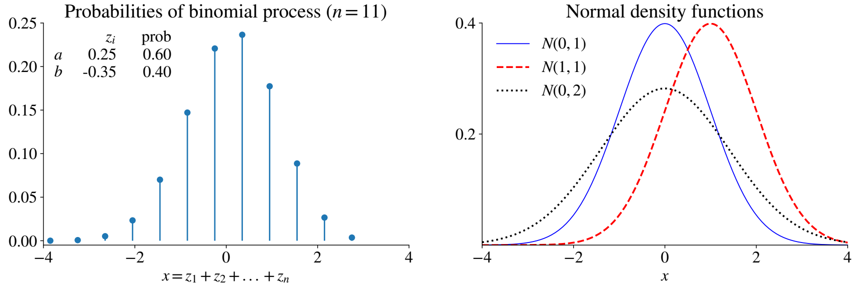


Figure 23.1: Density functions for a binomial and several normal distributions

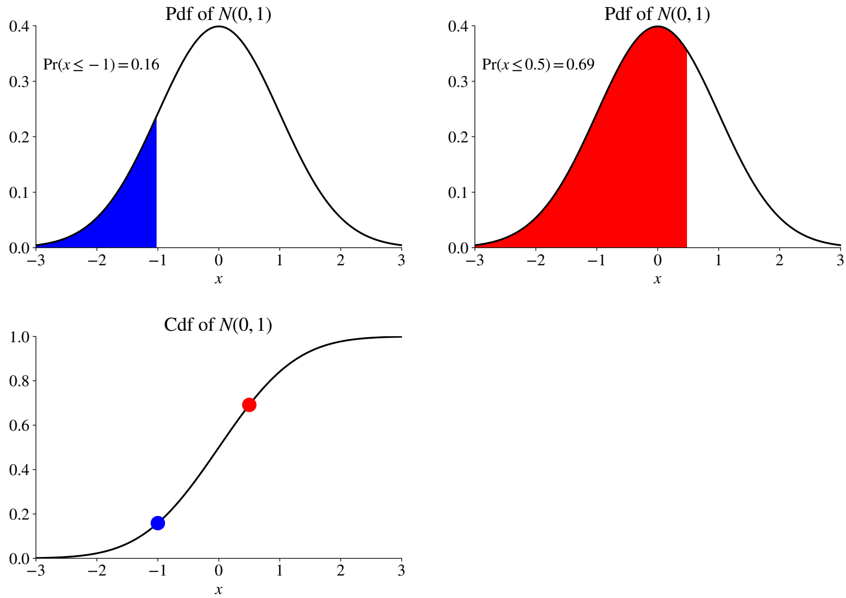


Figure 23.2: Pdf and cdf of $N(0,1)$

The probability that $x \leq B$ is given by the *cumulative distribution function*, $\text{cdf}(B)$. For instance, if x has a $N(0,1)$ distribution, then $\Pr(x \leq -1.645) = 0.05$ and $\Pr(x \leq 0) = 0.5$. See Figure 23.2 for an illustration.

If we invert the cdf, then we get the *quantiles* of the random variable. For instance, the 0.05th quantile of a $N(0,1)$ variable is -1.645, while the 0.5th quantile (also called the median) is 0.

23.2.2 Expected Value and Variance

The expected value (or mean) of a random variable x is defined as

$$\mathrm{E}\,x = \textstyle\sum_{s=1}^{S} \pi_s x_s \text{ or } \int f(x) x dx,$$

for a discrete and continuous random variable, respectively. For the former, π_s denotes the probability of outcome x_s, and for the latter $f(x)$ represents the probability density function (pdf). The probabilities must sum to unity; therefore $\Sigma_{s=1}^{S} \pi_s = 1$ and $\int f(x) dx = 1$. Again, see Figure 23.1. The expected value is sometimes denoted μ.

The expectation can be extended to a function $g(x)$ of the random variable as

$$\mathrm{E}\, g(x) = \textstyle\sum_{s=1}^{S} \pi_s g(x_s) \text{ or } \int f(x) g(x) dx.$$

A typical case is $g(x) = (x - \mu)^2$, which gives the variance

$$\mathrm{Var}(x) = \textstyle\sum_{s=1}^{S} \pi_s (x_s - \mu)^2 \text{ or } \int f(x)(x - \mu)^2 dx.$$

We often use σ^2 to denote the variance. The standard deviation σ is the square root of the variance, $\mathrm{Std}(x) = \mathrm{Var}(x)^{1/2}$, often denoted σ.

If a and b are two constants, then the previous expressions directly show that

$$\begin{aligned}
\mathrm{E}(a + bx) &= a + b\, \mathrm{E}\, x \\
\mathrm{Var}(a + bx) &= b^2\, \mathrm{Var}(x) \text{ or } \mathrm{Std}(a + bx) = |b|\, \mathrm{Std}(x).
\end{aligned}$$

Again, consider $\mathrm{E}\, g(x)$ and suppose x depends on a choice variable v, for instance, when x is the return of a portfolio of two assets, $vR_1 + (1 - v)R_2$. The derivative of $\mathrm{E}\, g(x)$ is then the expected value of the derivative, so we can interchange the order of E and the derivative

$$\frac{d\, \mathrm{E}\, g(x)}{dv} = \textstyle\sum_{s=1}^{S} \pi_s \frac{dg(x_s)}{dx} \frac{dx_s}{dv} = \mathrm{E}\, \frac{dg(x)}{dv}.$$

A similar expression holds for a continuous distribution.

23.2.3 Expected Value and the Variance of a Vector

There are straightforward extensions to vectors of random variables. For instance, if $x = [x_1, x_2]$ is a vector of the two random variables (returns?) x_1 and x_2 (the subscripts here indicate different variables, not time periods), then the mean of x is a vector of the means of the two individual returns

$$\mathrm{E}\, x = \begin{bmatrix} \mathrm{E}\, x_1 \\ \mathrm{E}\, x_2 \end{bmatrix}.$$

Also, the (2×2) variance-covariance matrix of x is

$$\mathrm{Var}(x) = \begin{bmatrix} \mathrm{Var}(x_1) & \mathrm{Cov}(x_1, x_2) \\ \mathrm{Cov}(x_2, x_1) & \mathrm{Var}(x_2) \end{bmatrix}.$$

Clearly, the variance-covariance matrix is symmetric (the two covariances are the same). The *correlation* of x_1 and x_2 is $\rho_{12} = \text{Cov}(x_1, x_2)/[\text{Std}(x_1)\,\text{Std}(x_2)]$. See Figure 23.3 for an example of a bivariate distribution.

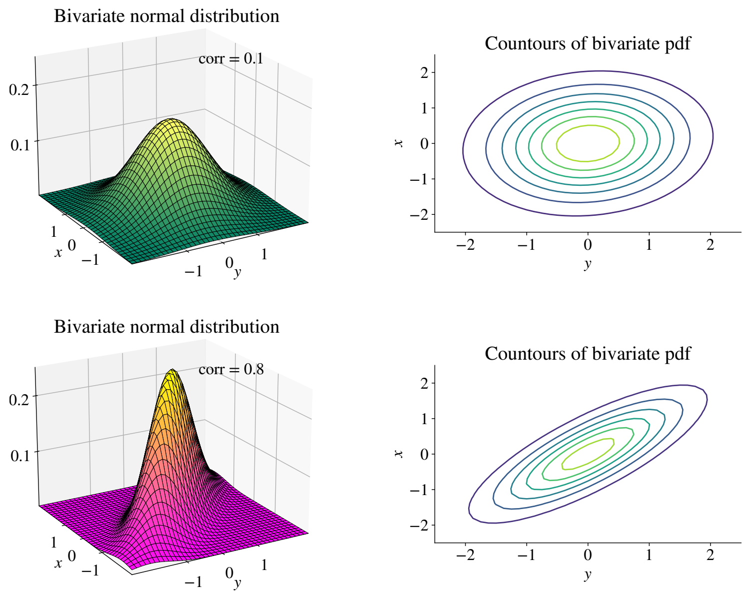


Figure 23.3: Density functions of bivariate normal distributions

23.2.4 Expected Value and the Variance of a Linear Combination

Consider a linear combination of the random variables $x_1, ..., x_n$

$$y = \textstyle\sum_{i=1}^{n} w_i x_i = w'x.$$

For instance, x could be a vector of portfolio returns and w a vector of portfolio weights.

The expected value and the variance are

$$\begin{aligned}\text{E}\, y &= w'\mu \\ \text{Var}(y) &= w'\Sigma w,\end{aligned}$$

where μ is a vector of average returns and Σ is the $n \times n$ variance-covariance matrix of x.

Also, consider another linear combination, $z = v'x$. Then, the covariance

$$\mathrm{Cov}(z, y) = v'\Sigma w.$$

This could, for instance, be two different portfolios.

Remark 23.14 *(Details on the matrix form) With two assets, we have the following:*

$$w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}, \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \text{ and } \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix},$$

where we use σ_{ii} to indicate σ_i^2 (this helps reading the matrices).

$$\begin{aligned} \mathrm{E}\, y &= w'\mu \\ &= \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \\ &= w_1\mu_1 + w_2\mu_2. \end{aligned}$$

$$\begin{aligned} \mathrm{Var}(y) &= w'\Sigma w \\ &= \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \\ &= \begin{bmatrix} w_1\sigma_{11} + w_2\sigma_{12} & w_1\sigma_{12} + w_2\sigma_{22} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \\ &= w_1^2\sigma_{11} + 2w_1w_2\sigma_{12} + w_2^2\sigma_{22}. \end{aligned}$$

23.2.5 Conditional Moments

Portfolio choice is based on expected future returns, variance and covariances. In general, these represent the beliefs of the investor at the time of investment. Clearly, this means that they may change over time and differ from the properties of historical data. Also, they are to be considered *conditional* in the sense that they refer to the current information/situation—and may therefore differ from *unconditional* moments.

As an example, suppose a random variable (return?) follows an AR(1) process

$$x_{t+1} = (1 - \rho)\mu + \rho x_t + u_{t+1},$$

where u_{t+1} is an iid term (innovation). In this case, the *conditional* expectation and variance are

$$\begin{aligned}
\mathrm{E}_t\, x_{t+1} &= (1-\rho)\mu + \rho x_t, \text{ and} \\
\mathrm{Var}_t(x_{t+1}) &= \mathrm{Var}(u_{t+1}).
\end{aligned}$$

This differ from the *unconditional*/long-run values which do not take into consideration the current state and are

$$\begin{aligned}
\mathrm{E}\, x_{t+1} &= \mu, \text{ and} \\
\mathrm{Var}(x_{t+1}) &= \mathrm{Var}(u_{t+1})/(1-\rho^2).
\end{aligned}$$

Note that there is no difference between conditional and unconditional moments when x is *iid* (independently and identically distributed), which here means $\rho = 0$. Notice that iid implies, among other things, that x is unpredictable and that the variance is constant over time.

23.2.6 Linear Regressions

Consider the linear regression model

$$\begin{aligned}
y_t &= x_{1t}\beta_1 + x_{2t}\beta_2 + \cdots + x_{kt}\beta_k + u_t \\
&= x_t'\beta + u_t,
\end{aligned}$$

where y_t and u_t are scalars, x_t a $k \times 1$ vector, and β is a $k \times 1$ vector of the true coefficients. In this expression, one of the elements of x_t is typically a constant equal to one (and the intercept is its coefficient).

Least squares minimizes the sum of the squared fitted residuals and gives

$$\hat{\beta} = S_{xx}^{-1}\textstyle\sum_{t=1}^{T} x_t y, \text{ where } S_{xx} = \textstyle\sum_{t=1}^{T} x_t x_t'.$$

Clearly, S_{xx} is an $k \times k$ matrix (and is often calculated as $X'X$ if row t of X contains x_t').

If the residuals are iid, then in large samples, we can approximate the distribution of $\hat{\beta}$ as

$$\hat{\beta} \sim N(\beta, S_{xx}^{-1}\sigma^2),$$

where β are the true values and $\sigma^2 = \mathrm{Var}(u_t)$ denotes the variance of the residuals. (In contrast, with autocorrelated residuals or time-varying variance of the residuals, then we have to apply Newey-West's or White's method for approximating the

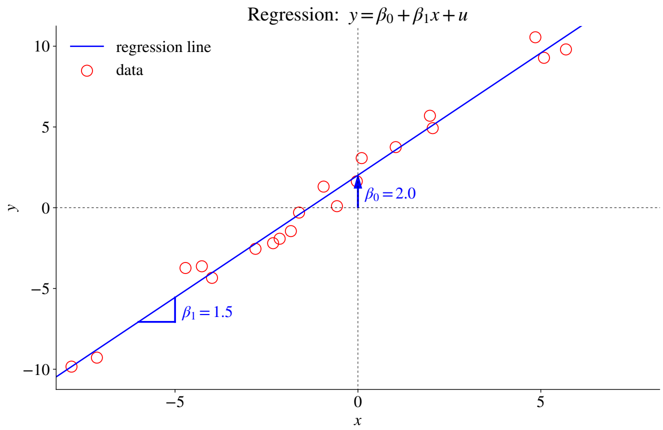


Figure 23.4: Example of OLS

variance-covariance matrix.) Based on this distribution, it is straightforward to test if a single coefficient equals a particular value by a t-test.

If $x_t = [1, z_t]$ we write the regression as $y_t = \alpha + z_t'\gamma + u_t$, then $\gamma = \Sigma_{zz}^{-1}\Sigma_{zy}$ where Σ_{zz} is the variance-covariance matrix of z_t and Σ_{zy} the vector of covariances of z_t and y_t. Also, $\alpha = \mathrm{E}\, y - (\mathrm{E}\, z_t')\gamma$. This holds both for the population and in a sample.

23.2.7 t-tests

Suppose the random variable x has a $N(\mu, \sigma^2)$ distribution. Then, the *standardized variable* $(x - \mu)/\sigma$ has a standard normal distribution

$$t = \frac{x - \mu}{\sigma} \sim N(0, 1).$$

To see this, notice that $x - \mu$ has a mean of zero and that x/σ has a standard deviation of unity. A t-distribution is sometimes used instead, since σ has to be estimated. However, with 30 or more data points, the t-distribution and the $N(0, 1)$ are almost indistinguishable.

23.3 Appendix – Calculus*

The following derivatives (with respect to x) are often used in this text

$$\frac{d}{dx}(ax^k + bx) = akx^{k-1} + b$$
$$\frac{d}{dx}\ln x = 1/x$$
$$\frac{d}{dx}e^x = e^x.$$

The first expression uses the *sum rule*: the derivative of a sum is the sum of the derivatives. Derivatives typically depend on at which x value we evaluate them at ($x = 1$ or $x = 2$, say), so the derivatives are themselves functions.

Example 23.15 *(Derivative of power function)* $3x^2 + 7x$ *has the derivative* $6x + 7$ *which is* -5 *at* $x = -2$ *and* 13 *at* $x = 1$.

The *chain rule* says that if $g()$ and $f()$ are two functions, then the derivative of the composite function $g(f(x))$ is

$$\frac{d}{dx}g(f(x)) = g'(u)f'(x), \text{ where } u = f(x),$$

and where $g'(u)$ is short hand (Lagrange's) notation for $\frac{d}{du}g(u)$, and similarly for $f'(x)$. The derivative $g'(u)$ is often referred to as the outer derivative and $f'(x)$ as the inner derivative.

Example 23.16 *(Chain rule) Let* $g(u) = u^2$ *and* $u = f(x) = 2 - 3x$*, so we are considering the composite function* $(2 - 3x)^2$*. We then get*

$$\frac{d}{dx}(2-3x)^2 = \underbrace{2(2-3x)}_{g'(u)}\underbrace{(-3)}_{f'(x)} = 18x - 12.$$

This derivative is -12 *at* $x = 0$ *and* 6 *at* $x = 1$.

Consider a function of two variables, $f(x, z)$. The *partial derivative* with respect to x is just a standard derivative, treating z as fixed. For instance,

$$\frac{\partial}{\partial x}ax^k bz = akx^{k-1}bz$$
$$\frac{\partial}{\partial z}ax^k bz = ax^k b.$$

Suppose the function $f(x)$ gives a scalar output, but x is a $n-$vector of inputs (with elements $x_1, x_2, ...x_n$). The *gradient* is then

$$\partial f(x)/\partial x = \begin{bmatrix} \partial f(x)/\partial x_1 \\ \vdots \\ \partial f(x)/\partial x_n . \end{bmatrix}$$

Similarly, $\partial f(x)/\partial x'$ is the transpose of this expression.

Example 23.17 *(Gradient) For the function $f(x) = (x_1 - 2)^2 + (4x_2 + 3)^2$, the gradient is*

$$\partial f(x)/\partial x = \begin{bmatrix} 2(x_1 - 2) \\ 8(4x_2 + 3) \end{bmatrix}.$$

The *Hessian* is the $n \times n$ matrix of second derivatives

$$\partial^2 f(x)/\partial x \partial x' = \begin{bmatrix} \partial^2 f(x)/\partial x_1^2 & \cdots & \partial^2 f(x)/\partial x_1 \partial x_n \\ & \ddots & \\ \partial^2 f(x)/\partial x_n \partial x_1 & & \partial^2 f(x)/\partial x_n^2 \end{bmatrix}.$$

(In case the derivatives are continuous, then this matrix is symmetric.)

Example 23.18 *(Hessian) Using the same function as in Example 23.17, we get*

$$\partial^2 f(x)/\partial x \partial x' = \begin{bmatrix} 2 & 0 \\ 0 & 32 \end{bmatrix}.$$

A *first-order Taylor approximation* of a differentiable function $f(x)$ is

$$f(b) \approx f(a) + \frac{\partial f(a)}{\partial x}(b - a),$$

where the derivative is evaluated as $x = a$. For highly non-linear functions, this only works well when a and b are close. For a vector of functions (which depend on a vector of variables x), we instead have

$$\begin{bmatrix} f_1(b) \\ \vdots \\ f_n(b) \end{bmatrix} \approx \begin{bmatrix} f_1(a) \\ \vdots \\ f_n(a) \end{bmatrix} + \begin{bmatrix} \frac{\partial f_1(a)}{\partial x_1} & \cdots & \frac{\partial f_1(a)}{\partial x_m} \\ \vdots & & \vdots \\ \frac{\partial f_n(a)}{\partial x_1} & \cdots & \frac{\partial f_n(a)}{\partial x_m} \end{bmatrix} \begin{bmatrix} b_1 - a_1 \\ \vdots \\ b_m - a_m \end{bmatrix} \text{ or}$$

$$f(b) \approx f(a) + \frac{\partial f(a)}{\partial x'}(b - a).$$

See Figure 23.5 for an illustration.

Example 23.19 *(Taylor approximation) Let* $f(x) = \ln x$ *and consider* $(a,b) = (1, 1.2)$*, so* $f(1.2) \approx 0 + 1 \times 0.2 = 0.2$*, when the true value is approximately* 0.18*. Instead, with* $b = 2$*, we get the approximation 1 and the true value around 0.69, so the error is considerable.*

A related concept is *the mean-value theorem* which says that for a differentiable function $f(x)$,

$$f(b) = f(a) + \frac{\partial f(c)}{\partial x}(b-a),$$

where the derivative is evaluated at value $x = c$ between a and b. A similar expression holds for a vector of functions

$$f(b) = f(a) + \frac{\partial f(c)}{\partial x'}(b-a),$$

but where the c vector might differ across the various functions. Again, see Figure 23.5 for an illustration.

Example 23.20 *(Mean-value theorem) Let* $f(x) = \ln x$ *and consider* $(a,b) = (1, 2)$*. With* $c \approx 1.443$*, we have* $0.693 \approx 0 + \frac{1}{1.443}(2-1)$*.*

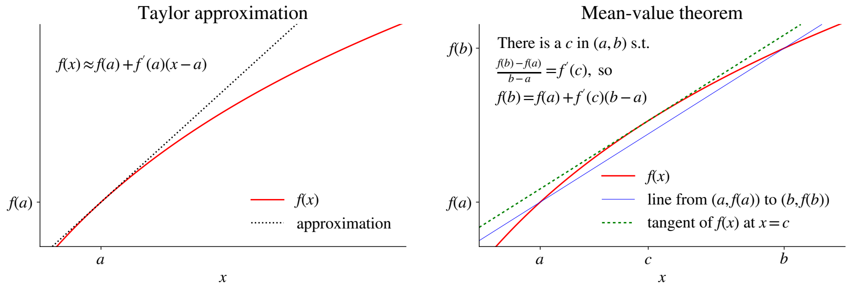


Figure 23.5: Illustration of a Taylor approximation and the mean-value theorem

Remark 23.21 *(K-th order Taylor approximation) A K-th order Taylor approximation is* $f(b) \approx f(a) + \Sigma_{n=1}^{K} \frac{1}{n!} \frac{\partial^n f(a)}{\partial x^n}(b-a)^n$*, where* $n! = n \times (n-1) \times \ldots \times 2 \times 1$*.*

Numerical derivatives are useful when the function is complicated. There are many software package that provide very fast and accurate calculations. The basic approach, however, is often to calculate a forward derivative as

$$f'(a) = [f(a+h) - f(a)]/h.$$

A careful choice of h is sometimes important. A central derivative is calculated as

$$f'(a) = [f(a+h) - f(a-h)]/(2h).$$

23.4 Appendix – Optimization*

Remark 23.22 *(First order condition for optimising a differentiable function). We want to find the value of b in the interval $b_{low} \leq b \leq b_{high}$, which makes the value of the differentiable function $f(b)$ as small as possible (a minimization problem). The answer is b_{low}, b_{high}, or a value of b where $df(b)/db = 0$. The latter is a necessary and sufficient condition for an unconstrained problem where $f(b)$ is convex. (If the function is twice differentiable, then convexity means that $f''(b) \geq 0$.) A maximization problem, except that we rather want $f(b)$ to be concave ($f''(b) \leq 0$).*

Suppose we want to *minimze* the loss function

$$L = (4y+3)^2$$

then we have to find the value of y that satisfy the *first order condition*

$$0 = dL/dy = 8(4y+3),$$

which requires $y = -3/4$. Notice that a *maximization problem* has the same type of first order conditions.

Instead, consider a loss function that depends on both x and y

$$L = (x-2)^2 + (4y+3)^2.$$

In this case, the first order conditions are

$$0 = \partial L/\partial x = 2(x-2)$$
$$0 = \partial L/\partial y = 8(4y+3),$$

which clearly requires $x = 2$ and $y = -3/4$. In this particular case, the first order condition with respect to x does not depend on y, but that is not a general property. See Figure 23.6 for the surface of the loss function and the contours. Also, in this

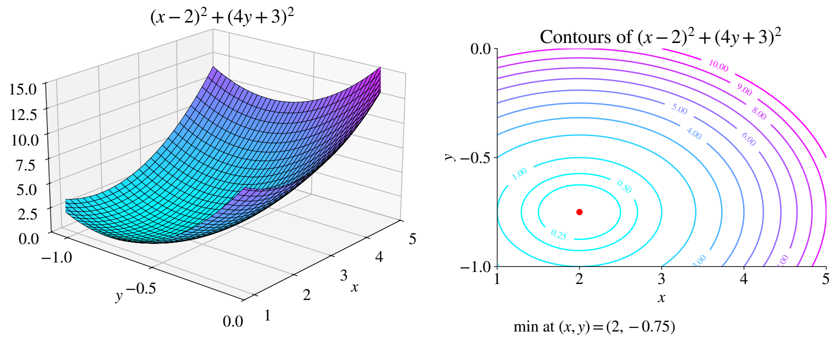


Figure 23.6: Minimization problem

case, there is a unique solution—but in more complicated problems, the first order conditions could be satisfied at different values of x and y.

If you want to add a *restriction* to the minimization problem, say

$$x + 2y = 3,$$

then we can proceed in two ways. The first is to simply substitute for $x = 3 - 2y$ in L to get

$$L = (1 - 2y)^2 + (4y + 3)^2,$$

with first order condition

$$0 = \partial L/\partial y = -4(1 - 2y) + 8(4y + 3) = 40y + 20,$$

which requires $y = -1/2$, which by implies $x = 4$. (We could equally well have substituted for y). This is also the unique solution. See Figure 23.7. This is an easy way to eliminate an equality restriction.

The second method is to use a *Lagrangian*. The problem is then to choose x, y, *and* λ to minimize

$$L = (x - 2)^2 + (4y + 3)^2 + \lambda\,(x + 2y - 3)\,.$$

The term multiplying λ is the restriction. (If you instead use $-\lambda()$ or write the restriction as $-x - 2y + 3$, you should get the same result. The interpretation of λ differs, though.)

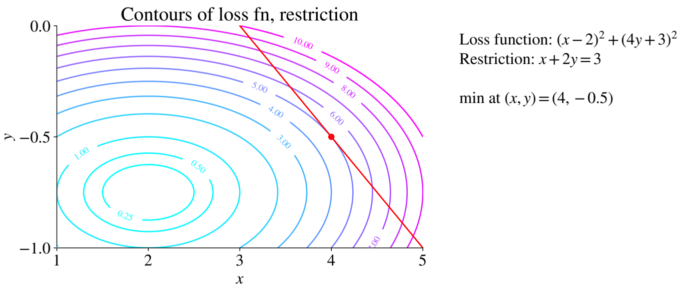


Figure 23.7: Minimization problem with restriction

The first order conditions are now

$$
\begin{aligned}
0 &= \partial L/\partial x = 2(x-2)+\lambda \\
0 &= \partial L/\partial y = 8(4y+3)+2\lambda \\
0 &= \partial L/\partial \lambda = x+2y-3.
\end{aligned}
$$

These are three equations in three unknowns (x, y, λ) which can be solved as $(x, y, \lambda) = (4, -1/2, -4)$.

Remark 23.23 *The three equations are linear, so we could rewrite them on matrix form as*

$$
\begin{bmatrix} 2 & 0 & 1 \\ 0 & 32 & 2 \\ 1 & 2 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ \lambda \end{bmatrix} = \begin{bmatrix} 4 \\ -24 \\ 3 \end{bmatrix},
$$

which is easily solved.

23.5 Appendix – Numerical Optimization Routines*

23.5.1 Unconstrained Minimization

Consider the loss function

$$
f(\theta) = (x-2)^2 + (4y+3)^2, \tag{23.1}
$$

where $\theta = (x, y)$ contains the two choice variables. Since this loss function is particularly simple—quadratic and also separable in x and y—the solution below

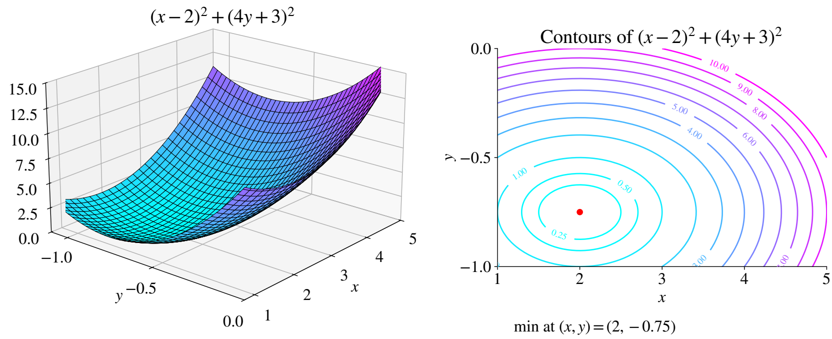


Figure 23.8: Numerical optimization, no restrictions

is straightforward (the minimum is at $(x, y) = (2, -3/4)$). However, the methods presented can also be used with more complicated loss functions.

A numerical minimization routine searches through different values of θ, typically starting from an initial guess, to find the values that makes $f(\theta)$ as small as possible. Convergence criteria, often set by the user, determine when the search will stop, for instance, when the improvement in $f(\theta)$ is smaller than a certain threshold or when the θ values stabilise. The starting guess is often important, so be sure to use reasonable values. See Figure 23.8 for an example.

Some algorithms use derivatives of the loss function (which may have to be coded by the user or are calculated numerically), while others do not ("derivative free"). The latter type is often slower, but sometimes more robust.

Most optimization algorithms are for minimizing a function value. In case you want to maximize, then just change the sign of the function and then minimize it. For instance, if you want to maximize $g(\theta)$, then you can do that by minimizing $-g(\theta)$.

23.5.2 Bounds on Variables

Many numerical optimization packages have options for setting bounds on the solution ("box minimization"). As an alternative, we could transform the variables and then apply an algorithm for unconstrained optimisation. The latter is briefly discussed below.

A simple way to handle a lower bound, such as $a \leq x$, is to let the routine optimize, without any restrictions, with respect to a transformed variable, $\tilde{x} =$

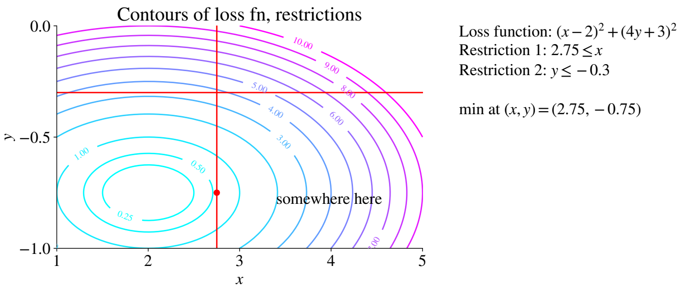


Figure 23.9: Numerical optimization with bounds on the solution

$\ln(x-a)$. Within the loss function—and also after having obtained the minimizer—the variable can be transformed back using $x = \exp(\tilde{x}) + a$.

Instead, with an upper bound, $x \leq b$, we optimize over $\tilde{x} = \ln(b-x)$ and transform back using $x = b - \exp(\tilde{x})$.

Suppose we use the same loss function (23.1) as before, but also impose the bounds

$$2.75 \leq x \text{ and } y \leq -0.3. \tag{23.2}$$

The solution is $(x, y) = (2.75, -3/4)$, so only one of the bounds is really binding. See Figure 23.9 for an illustration.

Remark 23.24 *With both lower and upper bounds $a \leq x \leq b$, we instead work with the (unbounded) $v = \operatorname{logit}(\frac{x-a}{b-a})$, where the logit function and its inverse are defined as $\operatorname{logit}(u) = \ln(\frac{u}{1-u})$ and $\operatorname{logit}^{-1}(v) = \frac{1}{1+\exp(-v)}$. (The inverse is also called the logistic function.) We can transform back using $x = a + (b-a)\operatorname{logit}^{-1}(v)$*

23.5.3 Equality Constraints

Suppose you want an *equality constraint* on the minimization problem, say

$$h_1(\theta) = x + 2y - 3 = 0. \tag{23.3}$$

One way to handle this is to use the constraint to rewrite the loss function (in this case, we could use $x = 3 - 2y$ to replace x in (23.1)). If this is tricky, then we try to find a routine that can handle equality constraints. The short discussion below outlines how these routines work (and also suggests how we could construct such a routine ourselves).

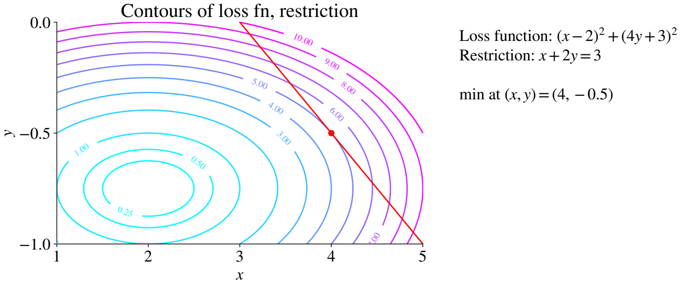


Figure 23.10: Numerical optimization with an equality restriction

A simple approach is to apply a penalty for deviations from the constraint, thereby modifying the overall loss function to

$$f(\theta) + \lambda \textstyle\sum_{i=1}^{p} h_i(\theta)^2, \tag{23.4}$$

where $h_i(\theta)$ is the ith equality constraint. In our example (23.3), there is only one restriction ($p = 1$). See Figure 23.10 for an example. (The solution should be $(x, y) = (4, -1/2)$.)

Start by setting $\lambda = 0$ and find the optimal value of θ, and call it θ_1. This is clearly the unconstrained solution. Then, increase λ and redo the optimization (using θ_1 as the starting guess) to get the optimal value θ_2. Now, increase λ further and redo the optimization (using θ_2 as the starting guess). Keep doing this (with higher and higher values of λ) until the solutions do not change much anymore. It is often worthwhile to experiment a bit with the sequence of λ values. In general, it seems as if initially making small increases and later larger ones works well in many cases. See Figure 23.11 for an example. (Clearly, there are more systematic ways to pick the sequence of penalties.)

23.5.4 Inequality Constraints

Instead, we now want to minimize (23.1) under the *inequality constraint* $y \leq -(x-4)^2$.

It is convenient to rewrite all inequality constraints on a common form, and we here choose to write them all on ≤ 0 form, which gives

$$g_1(\theta) = y + (x-4)^2 \leq 0. \tag{23.5}$$

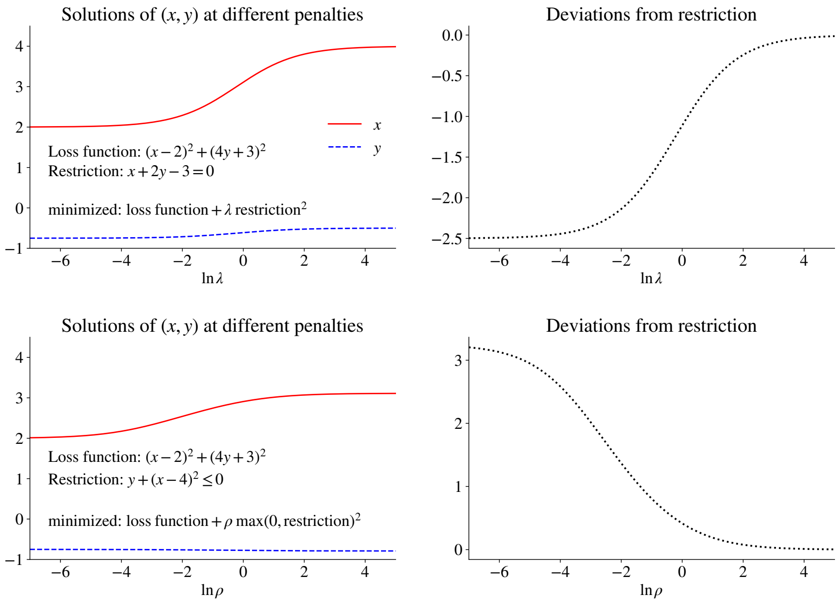


Figure 23.11: Numerical optimizations with penalty on the restriction

Now, we minimise the overall loss function

$$f(\theta) + \rho\textstyle\sum_{j=1}^{q} \max[0, g_j(\theta)]^2, \tag{23.6}$$

where $g_j(\theta)$ is the jth inequality constraint (there is only one in our example). Notice that ρ plays the same role as λ: start by solving for $\rho = 0$, then use that solution as a starting guess for the problem with a higher ρ, etc. See Figure 23.12 for an example. (The solution should be close to $(x, y) = (3.1, -0.79)$.) See Figure 23.11 for an iterative approach with a larger and larger penalty.

Finally, we can combine equality and inequality constraints as

$$f(\theta) + \lambda\textstyle\sum_{i=1}^{p} h_i(\theta)^2 + \rho\textstyle\sum_{j=1}^{q} \max[0, g_j(\theta)]^2. \tag{23.7}$$

Further Reading

See Brandimarte (2006), Stan manual (`http://mc-stan.org/users/documentation/`), Kochenderfer and Wheeler (2019)

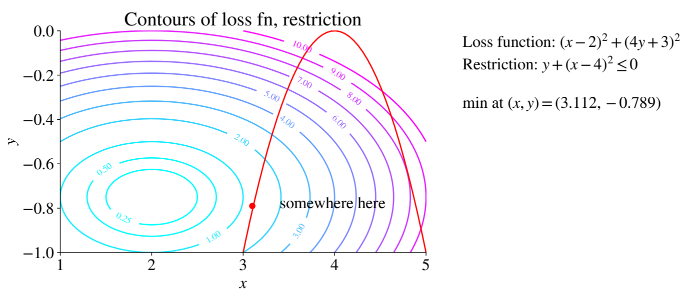


Figure 23.12: Numerical optimization with inequality restriction

Bibliography

Black, F., 1976, "The Pricing of Commodity Contracts," *Journal of Financial Economics*, 3, 167–179.

Black, F., and M. Scholes, 1973, "The pricing of options and corporate liabilities," *Journal of Political Economy*, 81, 637–659.

Brandimarte, P., 2006, *Numerical Methods in Finance and Economics*, Wiley, Hoboken, NJ.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay, 1997, *The econometrics of financial markets*, Princeton University Press, Princeton, New Jersey.

Cox, J. C., S. A. Ross, and M. Rubinstein, 1979, "Option pricing: a simplified approach," *Journal of Financial Economics*, 7, 229–263.

Elton, E. J., M. J. Gruber, S. J. Brown, and W. N. Goetzmann, 2014, *Modern portfolio theory and investment analysis*, John Wiley and Sons, 9th edn.

Fabozzi, F. J., 2004, *Bond markets, analysis, and strategies*, Pearson Prentice Hall, 5th edn.

Garman, M. B., and S. W. Kohlhagen, 1983, "Foreign currency option values," *Journal of International Money and Finance*, 2, 231–237.

Hull, J. C., 2022, *Options, futures, and other derivatives*, Pearson, 11th edn.

Jarrow, R. A., and A. Rudd, 1983, *Option Pricing*, Dow Jones-Irwin, Homewood, IL.

Kochenderfer, M. J., and T. A. Wheeler, 2019, *Algorithms for Optimization*, The MIT Press.

Lustig, H. N., N. L. Roussanov, and A. Verdelhan, 2011, "Common risk factors in currency markets," *Review of Financial Studies*, 24, 3731–3777.

McCulloch, J., 1975, "The tax-adjusted yield curve," *Journal of Finance*, 30, 811–830.

McDonald, R. L., 2014, *Derivatives markets*, Pearson, 3rd edn.

Merton, R. C., 1973, "Rational theory of option pricing," *Bell Journal of Economics and Management Science*, 4, 141–183.

Nelson, C., and A. Siegel, 1987, "Parsimonious modeling of yield curves," *Journal of Business*, 60, 473–489.

Ranaldo, A., and P. Söderlind, 2010, "Safe haven currencies," *Review of Finance*, 14, 385–407.

Rendleman, R. J., and B. J. Bartter, 1979, "Two-State Option Pricing," *The Journal of Finance*, 34, 1093–1110.

Sercu, P., 2009, *International Finance*, Princeton University Press.

Söderlind, P., and L. E. O. Svensson, 1997, "New techniques to extract market expectations from financial instruments," *Journal of Monetary Economics*, 40, 383–429.

Sundaram, R. K., and S. R. Das, 2015, *Derivatives: Principles and Practice*, McGraw-Hill Education, 2 edn.

Sundaresan, S., 2009, *Fixed Income Markets and Their Derivatives*, John Wiley & Sons, Hoboken, NJ, 3rd edn.

Svensson, L., 1995, "Estimating forward interest rates with the extended Nelson&Siegel method," *Quarterly Review, Sveriges Riksbank*, 1995:3, 13–26.

Vasicek, O. A., 1977, "An equilibrium characterization of the term structure," *Journal of Financial Economics*, 5, 177–188.

Vrins, F. D., 2025, *Derivatives Pricing*, Cambridge University Press, New York, NY.